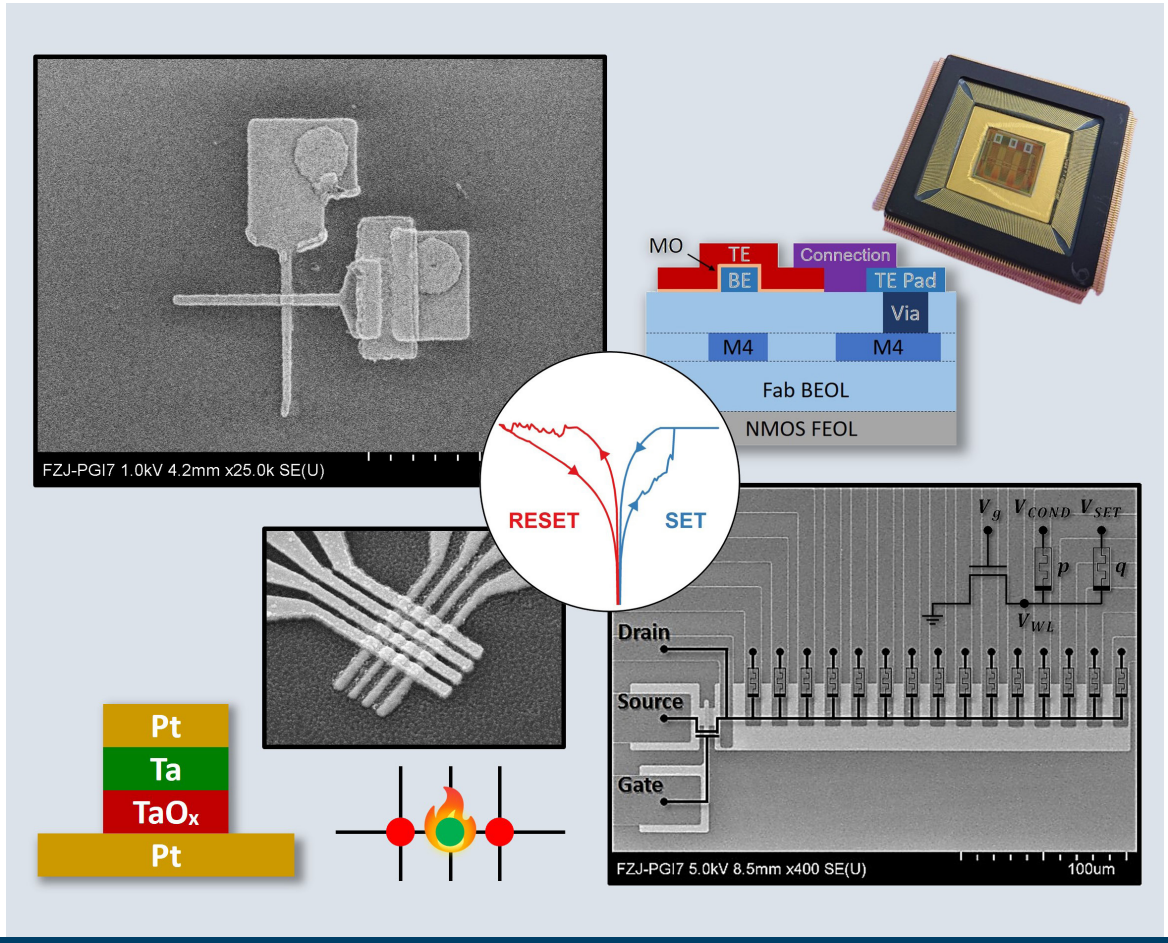

Redox-based Random Access Memory Arrays for Computing-In-Memory and Neuromorphic Computing

Hsin-Yu Chen

Information
Band / Volume 109
ISBN 978-3-95806-814-8

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum

Schriften des Forschungszentrums Jülich
Reihe Information / Information Band / Volume 109

Forschungszentrum Jülich GmbH
Peter Grünberg Institut (PGI)
JARA-Institut Energy-Efficient Information Technology (PGI-10)

Redox-based Random Access Memory Arrays for Computing-In-Memory and Neuromorphic Computing

Hsin-Yu Chen

Schriften des Forschungszentrums Jülich
Reihe Information / Information Band / Volume 109

ISSN 1866-1777 ISBN 978-3-95806-814-8

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über http://dnb.d-nb.de abrufbar.

Herausgeber und Vertrieb:	Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag 52425 Jülich Tel.: +49 2461 61-5368 Fax: +49 2461 61-6103 zb-publikation@fz-juelich.de www.fz-juelich.de/zb
Umschlaggestaltung:	Grafische Medien, Forschungszentrum Jülich GmbH
Druck:	Grafische Medien, Forschungszentrum Jülich GmbH
Copyright:	Forschungszentrum Jülich 2025

Schriften des Forschungszentrums Jülich
Reihe Information / Information, Band / Volume 109

D 82 (Diss. RWTH Aachen University, 2025)

ISSN 1866-1777
ISBN 978-3-95806-814-8

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER)
unter www.fz-juelich.de/zb/openaccess.

Redox-based Random Access Memory Arrays for Computing-In-Memory and Neuromorphic Computing

From the Faculty of Georesources and Materials Engineering of the
RWTH Aachen University

to obtain the academic degree of

Doctor of Natural Science

approved thesis

submitted by

Hsin-Yu Chen, M. Sc.

Advisors: **Univ.-Prof. Dr.-Ing. Rainer Waser**
Univ.-Prof. Dr. rer. nat. Joachim Mayer

Date of the oral examination: 29.01.2025

Abstract

The advancement in modern computing technology and applications strongly relies on the transistor downscaling that has been following Moore's law for almost 60 years. However, the device miniaturization is substantially approaching its physical limit. The further development of computation performance requires "more than Moore" innovations such as memory-centric computing architectures, which have been proposed to break the von Neumann bottleneck. Recently, computing-in-memory (CIM), combining the processor function into the memory and executing computation directly in the memory, and neuromorphic computing (NC), using artificial electronic synapses and neurons to form brain-inspired architectures, have attracted extensive research interests from academia and industry. Apart from conventional charge-based memory, redox-based random access memory (RRAM) has been acknowledged as a low-cost, high-speed, and non-volatile resistance-based memory for CIM and NC. Additionally, it has excellent compatibility to advanced complementary metal-oxide-semiconductor (CMOS) technology, and also exhibits ultra-low energy consumption, offering a great advantage to edge artificial intelligence (AI) applications.

This thesis work focuses on the back-end-of-line (BEOL) integration and electrical characterization of active RRAM arrays based on valance change memory. Adopting N-channel metal-oxide-semiconductor field-effect transistors (MOSFETs) as selecting components, microscale and nanoscale technology platforms of active RRAM arrays were developed at the Helmholtz Nano Facility in Research Center Jülich. On the one hand, in the microscale technology platform, plug-type TaO_x RRAMs were integrated on the NiSi drain contacts of planar high-k metal-gate MOSFETs, where the NiSi layer was not suggested to serve as the bottom electrode of RRAM directly. In the process of producing contact holes with areas of 2×2 μm^2 to expose the NiSi drain contacts, a light interference issue was identified in the contact lithography, and the microloading effect was found considerable in the reactive-ion-etching (RIE) using CHF_3. Accordingly, a direct writing approach was introduced by employing a maskless aligner, and the etching time was prolonged with additional wet etching in 1 % HF solution. On the other hand, the nanoscale technology platform is based on monolithic integration of RRAMs with CMOS circuitry taped out with TSMC 180 nm technology node. Configured with 64×64 1T-1R arrays, this platform is designed with on-chip signal amplifiers and driving/sensing circuitry to realize dot product engines, which serve as brain-inspired energy-efficient AI accelerators. Using e-beam lithography (EBL), the N-channel MOSFETs fabricated in the front-end-of-line were integrated with crossbar RRAM devices in the BEOL. In the fabrication of nanoscale RRAMs, the significantly low device yield was attributed to the redeposition during the Pt etching through Ar reactive-ion-beam-etching (RIBE), which is also known as fencing. Consequently, the fence removal was carried out with an additional CF_4 RIBE process at a tilted angle following after the Ar-based RIBE process. Besides, a fence-free RIE process with Cr hard masks using a gas mixture of Cl_2 and Ar was developed to avoid significant fencing during the Ar-based RIBE process. To drive the RRAM-integrated CMOS die, chip packaging was carried out to enable the connection to a customized operating hardware. Eventually, bipolar resistive switching was successfully

performed on the packaged chip, which verifies the functionality and paves the way to realizing NC applications.

From quasi-static electrical measurements of the TaO_x RRAMs integrated on the established technology platforms, the 1T-1R configuration was proven advantageous in improving the current overshoot control, which enables consistent and reliable switching characteristics, in comparison to the 1R configuration. In addition, multi-level resistive switching was demonstrated on 1T-1R unit cells through modulations of the gate voltage during SET process and the RESET-stop voltage respectively. In the former case, when the gate voltage for SET increases, the resistance of low resistance state (LRS) decreases, and the voltage required to trigger the subsequent RESET process exhibits a considerable increase. In the latter case, a higher RESET-stop voltage results in a higher resistance of high resistance state (HRS), and therefore a minor increase in the voltage required to trigger the subsequent SET process. Notably, a surge of variability in the HRS resistance was observed as the RESET-stop voltage increases, which appears to be the limitation for the multi-level resistive switching modulated by the RESET-stop voltage. Furthermore, the bipolar resistive switching property was found to be affected by the gate voltage during electro-forming process, which determines initial conditions of oxygen vacancy concentration and conductive filament geometry for subsequent RESET and SET processes. As the gate voltage for electro-forming increases, the resistances of LRS and HRS decrease, and the voltages required for SET and RESET are lowered simultaneously. It was also found that the electro-formed condition changed dramatically when the equivalent current compliance exceeded a critical value.

Using a 1T-nR line array configured with TaO_x RRAMs, CIM was experimentally demonstrated by a stateful logic gate for material implication (IMP), where the transistor not only provides flexible tuning of the series resistance by the gate voltage, but also improves current overshoot control during SET processes in the RRAMs when performing logic operations. With the focus on the SET process, impact of device-to-device (D2D) variability on the IMP stateful logic operations was investigated. By assigning the RRAM device with the lower voltage required for SET as the q bit, an absolute advantage in success rate enhancement was concluded from experiments, suggesting that the inherent D2D variability in the RRAM array can be exploited for the IMP stateful logic operations.

Lastly, passive nano-crossbar arrays of TaO_x RRAMs with the device spacing of 70 nm were fabricated to study the thermal crosstalk effect in high-density RRAM arrays. In the fabrication, the actual e-beam exposure area was shrunk to compensate the pattern enlargement effect caused by the non-ideal undercut of resist profiles in lift-off patterning and the proximity effect in EBL. By emulating the scenario of WRITE operations for SET under the V/2 biasing scheme, the half-selected RRAM cells adjacent to the fully-selected cell show a high retention failure rate of 72 % in average, when they were originally in the HRS. The average bit-flip probability of 46 % implies the bias polarity over the half-selected cell determines the tendency of retention failure leading to a lower resistance state.

Kurzfassung

Der Fortschritt der modernen Rechentechnik und -anwendungen hängt stark von der Verkleinerung der Transistoren ab, die seit fast 60 Jahren dem Mooreschen Gesetz folgt. Jedoch nähert sich die Miniaturisierung der Bauelemente zusehends ihrer physikalischen Grenze. Die weitere Entwicklung der Rechenleistung erfordert "More than Moore"-Innovationen wie speicherzentrierte Rechenarchitekturen, die vorgeschlagen wurden, um den von-Neumann-Engpass zu überwinden. In den letzten Jahren haben Computing-in-Memory (CIM), bei dem die Prozessorfunktion mit dem Speicher kombiniert und die Berechnungen direkt im Speicher ausgeführt werden, und Neuromorphic Computing (NC), bei dem künstliche elektronische Synapsen und Neuronen verwendet werden, um vom Gehirn inspirierte Architekturen zu bilden, großes Forschungsinteresse in Wissenschaft und Industrie geweckt. Neben dem herkömmlichen ladungsbasierten Speicher ist der redoxbasierte wahlzugriffsfreiern Speicher (RRAM) als kostengünstiger, schneller und nichtflüchtiger widerstandsbasierter Speicher für CIM und NC anerkannt. Darüber hinaus zeichnet er sich durch einen extrem niedrigen Stromverbrauch aus, was für Anwendungen mit Künstlicher Intelligenz (KI) einen großen Vorteil darstellt.

Diese Dissertation befasst sich mit der Integration in das Back-End-of-Line (BEOL) und der elektrischen Charakterisierung von aktiven RRAM-Arrays auf der Basis von Valance-Change-Memory. Unter Verwendung von Metall-Oxid-Halbleiter-Feldeffekttransistoren (MOSFETs) als selektierende Komponenten wurden an der Helmholtz Nano Facility im Forschungszentrum Jülich mikro- und nanoskalige Technologieplattformen für aktive RRAM-Arrays entwickelt. In der mikroskaligen Technologieplattform wurden einerseits TaO_x-RRAMs vom Plug-Typ auf den NiSi-Drain-Kontakten von planaren High-k-Metal-Gate-MOSFETs integriert, wobei die NiSi-Schicht nicht direkt als untere Elektrode des RRAMs dienen sollte. Bei der Herstellung von Kontaktlöchern mit einer Fläche von 2×2 μm^2 zur Freilegung der NiSi-Drain-Kontakte wurde bei der Kontaktlithografie ein Lichtinterferenzproblem, und beim reaktiven Ionenätzen (RIE) mit CHF_3 wurde ein erheblicher Mikrobelastungseffekt festgestellt. Dementsprechend wurde ein direkter Schreibansatz mit einem maskenlosen Aligner eingeführt, zuerdem wurde die Ätzzeit durch zusätzliches Nassätzen in einprozentiger HF-Lösung verlängert. Andererseits basiert die Technologieplattform im Nanomaßstab auf der monolithischen Integration von RRAMs mit CMOS-Schaltungen (Complementary Metal-Oxide-Semiconductor), die mit dem 180-nm-Technologieknoten von TSMC hergestellt werden. Diese mit 64×64 1T-1R-Arrays konfigurierte Plattform ist mit On-Chip-Signalverstärkern und Ansteuerungs-/Erfassungsschaltungen ausgestattet, um Punktprodukt-Engines zu realisieren, die als vom Gehirn inspirierte energieeffiziente KI-Beschleuniger dienen. Mit Hilfe der Elektronenstrahllithographie (EBL) wurden die im Front-End-of-Line hergestellten MOSFETs mit Crossbar-RRAM-Bauelementen im BEOL integriert. Bei der Herstellung von RRAMs im Nanomaßstab wurde die signifikant niedrige Ausbeute an Bauelementen auf die erneute Abscheidung von Pt während des Ätzens durch reaktives Ionenstrahlätzen mit Ar (RIBE) zurückgeführt, welches auch als "Fencing" bezeichnet wird. Daher wurde die Entfernung des Fencings mit einem zusätzlichen CF_4-RIBE-Prozess unter

einem geneigten Winkel im Anschluss an den Ar-RIBE-Prozess durchgeführt. Außerdem wurde ein Fencing-freier RIE-Prozess mit Cr-Hartmasken unter Verwendung eines Gasgemischs aus Cl_2 und Ar entwickelt, um eine signifikante Fencing-bildung während des Ar-RIBE-Prozesses zu vermeiden. Zur Ansteuerung des RRAM-integrierten CMOS-Chip wurde ein Packaging durchgeführt, um die Anbindung an einer spezifischen Betriebshardware zu ermöglichen. Schließlich wurde das bipolare Widerstandsschalten auf dem verpackten Chip erfolgreich durchgeführt, was die Funktionalität verifiziert und den Weg zur Realisierung von NC-Anwendungen ebnet.

Quasistatische elektrische Messungen der TaO_x-RRAMs, die auf den etablierten Technologieplattformen integriert sind, ergaben, dass die 1T-1R-Konfiguration im Vergleich zur 1R-Konfiguration Vorteile bei der Verbesserung der Stromüberschwingungskontrolle bietet, die konsistente und zuverlässige Schalteigenschaften ermöglicht. Darüber hinaus wurde auf 1T-1R-Einheitszellen durch Modulationen der Gate-Spannung während des SET-Prozesses bzw. der RESET-Stopp-Spannung ein mehrstufiges Widerstandsschalten demonstriert. Im ersten Fall sinkt der Widerstand des niederohmigen Zustands (LRS), wenn die Gatespannung für den SET-Prozess erhöht wird, und die Spannung, die zum Auslösen des anschließenden RESET-Prozesses erforderlich ist, steigt erheblich an. Im zweiten Fall führt eine höhere RESET-Stopp-Spannung zu einem höheren Widerstand des hochohmigen Zustands (HRS) und damit zu einem geringfügigen Anstieg der für die Auslösung des nachfolgenden SET-Vorgangs erforderlichen Spannung. Auffällig ist, dass der HRS-Widerstand mit zunehmender RESET-Stopp-Spannung stark schwankt. Darüber hinaus wurde festgestellt, dass die bipolare Widerstandsschalteigenschaft durch die Gate-Spannung während des Elektroformierungsprozesses beeinflusst wird, welche die Anfangsbedingungen der Sauerstoffvakanzkonzentration und der leitfähigen Filamentgeometrie für die nachfolgenden RESET- und SET-Prozesse bestimmt. Wenn die Gate-Spannung für die Elektroformierung steigt, sinken die Widerstände von LRS und HRS, und die für SET und RESET erforderlichen Spannungen werden gleichzeitig gesenkt.

Unter Verwendung eines 1T-nR-Line-Arrays, das mit TaO_x-RRAMs konfiguriert ist, wurde CIM experimentell durch ein zustandsabhängiges Logikgatter für Materialimplikation (IMP) demonstriert. Mit dem Schwerpunkt auf dem SET-Prozess wurden auch die Auswirkungen der Variabilität von Bauelement zu Bauelement (D2D) auf die zustandsabhängigen IMP-Logikoperationen untersucht. Durch die Zuweisung des RRAM-Bausteins mit der niedrigeren Spannung, die für SET erforderlich ist, als q-Bit wurde in Experimenten ein absoluter Vorteil bei der Verbesserung der Erfolgsrate festgestellt, was darauf hindeutet, dass die inhärente D2D-Variabilität im RRAM-Array für die zustandsabhängigen Logikoperationen von IMP ausgenutzt werden kann.

Schließlich wurden passive Nano-Crossbar-Arrays aus TaO_x-RRAMs mit einem Bauelementabstand von 70 nm hergestellt, um den Effekt des thermischen Übersprechens in RRAM-Arrays mit hoher Dichte zu untersuchen. Bei der Herstellung wurde die tatsächliche E-Beam-Belichtungsfläche verkleinert, um den Effekt der Mustervergrößerung zu kompensieren, der durch die nicht ideale Unterschneidung der Resistprofile bei der Lift-Off-Strukturierung und den Proximity-Effekt bei EBL verursacht wird. Durch die Emulation des Szenarios von WRITE-Operationen für SET unter dem V/2-Vorspannungsschema zeigt die halb-selektierten RRAM-Zellen neben der voll-selektierten Zelle eine hohe Retentionsfehlerrate von durchschnittlich 72 %, wenn sie sich ursprünglich im HRS befanden. Die durchschnittliche Bit-Flip-Wahrscheinlichkeit von 46 % deutet darauf hin, dass die Vorspannungspolarität über die halb-selektierten Zellen die Tendenz zum Retentionsausfall, was zu einem niedrigeren Widerstandszustand führt.

Acknowledgments

I should admit I could not finish this dissertation and obtain a doctor degree without lots of helping hands and great supports from countless individuals. Considering the length of writing, I am afraid that I cannot list them all. However, I would like to thank every one of them for their contributions in different forms.

First and foremost, I would like to express my sincere gratitude to Prof. Dr. Rainer Waser for being my first advisor and offering this precious opportunity allowing me to pursue my doctoral study oversea in Germany. I am also grateful to work on Redox-based Random Access Memory (RRAM) in PGI-7 *Electronic Materials*, which is the division led by Prof. R. Waser in Peter Grünberg Institute (PGI) at Research Center Jülich (Forschungszentrum Jülich). I was motivated every day simply by doing research in this international research institution with advanced resources and diverse talents. I am glad that I was able to work frequently in the high-class clean room at Helmholtz Nano facility (HNF), allowing me to quickly absorb the knowledge and accumulate the experience of semiconductor processes. In addition, by participating the NEUROTEC, a large-scale research project led by Prof. R. Waser, I was able to gain a bigger picture on how RRAM research can be integrated vertically from materials to applications. Fortunately, my interest in RRAM research has been growing consistently over time, eventually leading me to my first career as an RRAM device engineer at GlobalFoundries in Dresden.

Next, I would like to thank Prof. Dr. Joachim Mayer for being my second advisor, and I appreciate the fantastic research discussion brought up by his questions during the examination. I also thank Prof. Dr. Sebastian Münstermann, the chair of doctoral committee, for his kindness in general and his clear communication regarding the examination procedure and regulation.

Thanks to the technology supports on transistor fabrication from PGI-9 *Semiconductor Nanoelectronics*, I could integrate RRAM devices and transistors to develop RRAM active arrays, which eventually became a major part of my work. In this regard, I would like to thank Prof. Dr. Qing-Tai Zhao for enabling this close collaboration and providing valuable insights on layout designs, process improvements, and device characteristic optimizations.

I thank Dr. Vikas Rana for being my supervisor, introducing me into the research field of RRAM and facilitating important connections for collaborations and technical assistances. In addition, a special thank goes to Dr. Stephan Menzel, who initiated a collaboration idea on stateful logic gates incorporating simulations and experiments. I also appreciate the prompt assistances and the practical opinions given by Dr. Susanne Hoffmann-Eifert regarding physical analysis on materials and devices. Moreover, I am grateful to Dr. Stefan Wiefels for his useful insights on testing instruments and testing methods for electrical characterization.

I am profoundly thankful for all the helps from my colleagues at PGI-7. I thank Tim Kempen, who used to be my only colleague in the same group, for demonstrating a lot of device fabrications and electrical characterizations when I joined the group and first stepped into this research field. He helped me to build up LuA programming skills from scratch for semi-auto electrical testing, and we had many technical discussions that were mutually helpful for our works, when we were sitting in our office seats facing each other. Besides, I am glad Yongmin

Wang joined our group after me also as a doctoral researcher so that I could speak mandarin to share some of my thoughts at work with her, which helped me a lot when I was dealing with working pressure and obstacles. I am deeply grateful to Prof. Dr. Christian Pithan for bringing up some small talks about Taiwan at work and inviting me to his house for lovely BBQ parties, which made me feel some warmth when living abroad. I thank Dr. Felix Cüppers and Dr. Solomon Amsalu Chekol for the numerous constructive discussions when I encountered problems of either fabrication processes or electrical characterizations. From time to time when I was trapped in the haze of frustrations, their honest advice reminded me of the direction towards the destination of my PhD journey, even though they merely came from our random small talks in the institute kitchen or corridor most of the time. I am grateful to Daniel Schön for our collaboration on thermal crosstalk study, bridging experiments and simulations. The excitement I felt during our discussions was a pure enjoyment of doing research. I thank René Borowski, Grigory Potemkin, Clemens Wiedenhöft, and Jochen Friedrich for offering technical supports and sharing equipment knowledge regarding clean room facility, sputtering deposition, and scanning electron microscopes. I also thank Marcel Gerst for his IT supports and Maria Garcia for administrative assistances.

An extraordinary thank goes to Dr. Christopher Bengel from Institute of Materials in Electrical Engineering 2 (IWE2) at RWTH Aachen University. Under his lead, our smooth collaboration on stateful logic gates successfully results in a conference publication. Thanks to this collaboration, I could eventually implement a computing-in-memory application using the RRAM arrays that I developed from scratch, which is truly a meaningful achievement for me.

For the seamless collaboration with PGI-9, I would like to express my deepest gratitude to Dr. Fengben Xi for transferring process knowledge and discussing various process issues. I appreciate all the spontaneous training you gave in person in the clean room at HNF, which allows me to work independently in the shortest time frame. I also thank Dr. Yi Han for his advice on process issues. In addition, I thank Dr. Jürgen Schubert and Jin Hee Bae for their professional assistances on Rutherford backscattering spectrometry and analysis with transmission electron microscope respectively.

As for the time over 900 hours I have spent in the clean room at HNF, I want to thank equipment owners and technology advisors for their professional supports on training, troubleshooting, and technical consulting. I am grateful especially to Dr. Jhih-Sian Tu, who is one of the experienced technology advisors. We both come from Taiwan, so not only we can communicate in mandarin, but also his friendliness provided an intangible emotional support in the workplace. Many thanks go to Sven Clausen from the photolithography group, Bich Nguyen from the reactive ion etching group, Dr. Florian Lentz and Dr. Stefan Trellenkamp from the e-beam lithography group for their technical expertise.

With heartfelt appreciation, I thank my parents for their unwavering support all the way from Taiwan throughout my entire PhD journey oversea despite the ups and downs. I am delighted to share with them the joy of completing my PhD. 感謝我的父母不曾間斷的支持，即使遠在臺灣，他們還是我最可靠的後盾，我很高興能跟他們分享這份完成博士學位的喜悅與感動。

Last but not least, I would like thank myself for not letting unexpected difficulties stop me from moving forward. Everyone's PhD journey is unique, and only I can complete it by creating my own path and destination regardless of the amount of help I receive. Every choice comes with sacrifice and compromise, but I have no doubt that I will never regret this one.

Contents

Chapter 1
Introduction

1.1 State of the art

Computer is one of the biggest human inventions that have changed the world. As semiconductor technology improves in the past decades by countless efforts dedicated to fundamental research and engineering development, the advancement of miniaturization drives the computing performance to pursue faster computation and lower energy consumption. Starting from the era of vacuum tube in the 1940s, computer technology was then brought into a whole new page after a semiconductor-based electronic device, transistor, was introduced to the computer world. In the past decades, metal-oxide-semiconductor field effect transistors (MOSFETs) not only shrank in the size by development of semiconductor process technology, but also evolved from planar structures to three-dimensional (3D) fin-like structures, so called FinFETs. As shown in Figure 1.1, Moore's law depicts the fact and the projection of future that the number of transistors doubles in a dense integrated circuit (IC) about every two years since 1965. However, Moore's law is nowadays approaching the physical limit of atoms. The research efforts to fulfill Moore's law increases dramatically. Many research paths have been addressed for "more than Moore" including new device architectures, heterogeneous integration, efficient computing architectures, advanced packaging, etc.

Modern computing architecture based on von Neumann architecture mainly consists of two separate units namely processor and memory, as shown in Figure 1.2a. Essentially, it is a processor-centric approach, where data needs to be firstly fetched from and afterwards sent back to the memory in order to perform computation in the processor. Data movement in between processor and memory is therefore an inherent part of the von Neumann architecture, resulting in the costs of operation latency and energy consumption. When performances of processor and memory chips are pushed towards the extreme by state-of-the-art IC manufacturing technology, the drawback of data movement becomes more prominent, and it is nowadays the well-known von Neumann bottleneck that hinders the path to the next generation of computing. For example, accessing memory to compute can consume 100 to 1000 times of energy in comparison to a complex addition computation. Data movement can take up to 62.7% of total system energy [1]. In this regard, a few non-von Neumann architectures have been proposed.

Computing-in-memory (CIM) or in-memory-computing (IMC) is one of the promising approaches beyond the von Neumann architecture, as illustrated in Figure 1.2b. The CIM takes the advantage of doing computation directly in the memory unit to prevent any extra data movement. In addition to a pure CIM architecture system, the CIM concept can be combined with conventional processors as a hybrid architecture. The first level of computations can take place in the memory and only the necessary data as the result is then transferred to the processor for advanced computations.

Similar to the CIM, neuromorphic computing (NC) also directly exploits memory components to perform computation. As human brain is naturally an ultra-energy-efficient

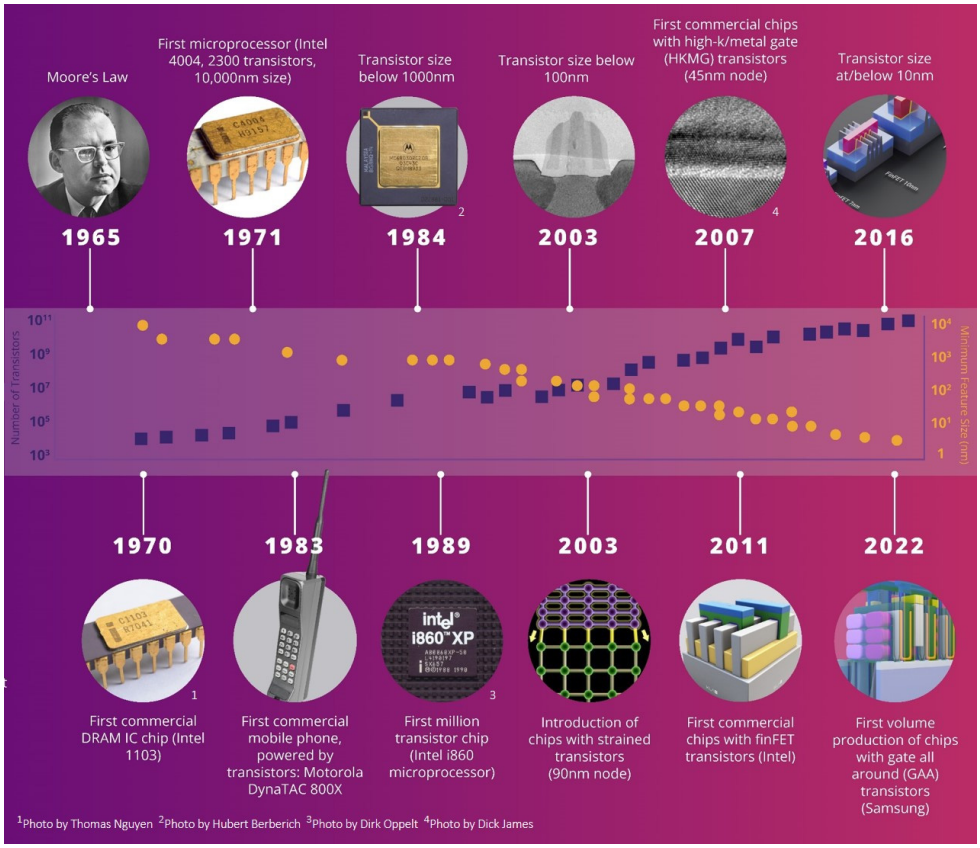

Figure 1.1: *Moore's law and transistor technology development. Moore's law predicts the number of transistors doubles in a dense integrated circuit (IC) about every two years since 1965. After technological development of almost 60 years, the transistor miniaturization driven by innovations on material property, process technology, and device structure is now facing the physical limit. Adapted from [2].*

computing system, the brain-inspired architecture shows great potentials in terms of energy consumption and especially artificial intelligence (AI) applications. The NC utilizes memory devices to work as artificial synapses and neurons [3-7]. Their special electrical features and responses can be exploited in an IC chip to emulate the computing process in a human brain.

The CIM and NC both place memory as the center role with computation functions added on top of it. The transition to memory-centric approach arises the importance of memory devices. In the current hierarchy of memory, all memory classes from fast and volatile memory, e.g. static random access memory (SRAM) and dynamic random access memory (DRAM), to slow and non-volatile memory, e.g. flash memory, are charge-based memory, of which electric charge is the memorizing medium. They are available technologies for the CIM and NC. Nonetheless, with comprehensive concerns of cost, speed, and energy consumption, emerging classes of resistance-based memory step in as promising players for next-generation computing.

Resistance-based memory notably connects to the concept of memristor, which was proposed by Leon Chua in 1971 and known as the fourth fundamental circuit element besides

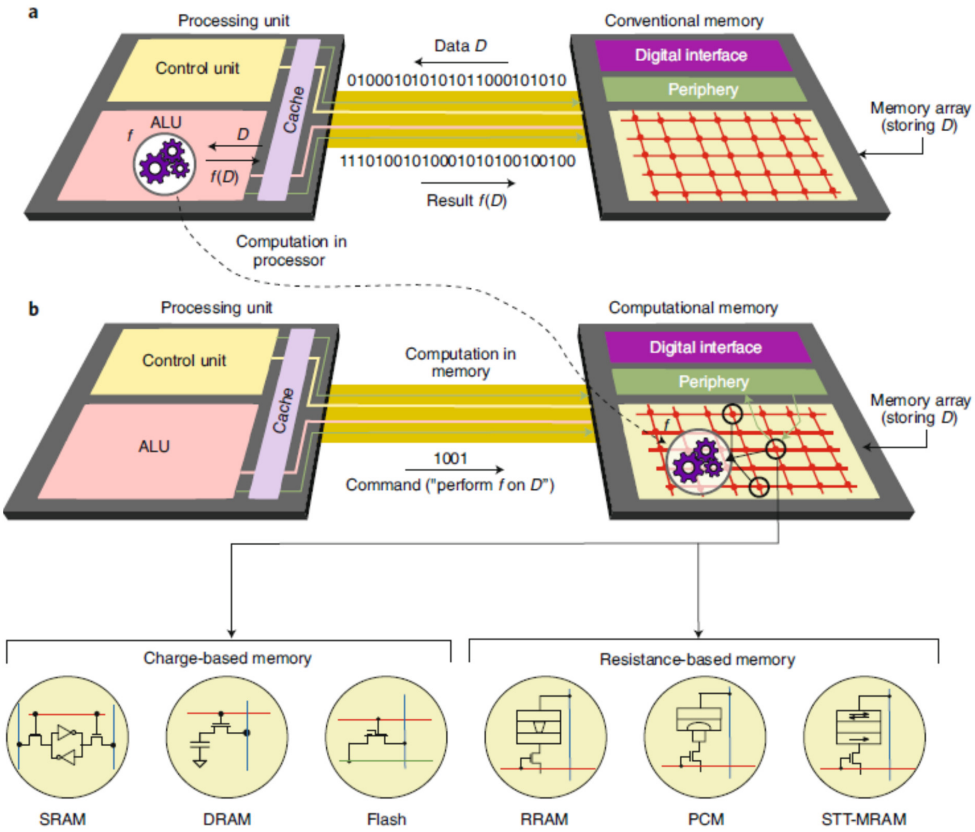

Figure 1.2: *Schematics of computing architectures [8]. (a) A conventional computing architecture requires data movements between the processing unit and the memory unit to perform an operation f on data D and store the result f(D). (b) A CIM architecture offers the possibility to perform f on D within the memory unit, thereby obviating the need of data movements to and from the processing unit. Both charge-based memory, such as SRAM, DRAM and flash memory, and resistance-based memory, such as RRAM, PCM and spin-transfer torque (STT) MRAM, can serve as elements of such a computational memory unit.*

resistor, capacitor, and inductor [9, 10]. The current development of resistance-based memory can be broadly categorized into redox-based random access memory (RRAM), phase change memory (PCM), and magnetoresistive random access memory (MRAM). RRAMs have appeared to be a competitive candidate of non-volatile memory (NVM) and attracted significant attention from academia and industry. The non-volatility and ultra-low energy consumption are appealing to on-chip training, deep learning, and other AI applications. With new computing architectures and algorithms, RRAMs have successfully demonstrated the CIM and NC in different approaches [11-21]. In addition to the practicality, RRAMs have excellent compatibility to advanced complementary metal-oxide-semiconductor (CMOS) technology. Besides easy adaptation in chip manufacturing, the CMOS compatibility also enables monolithic integration of memory and logic units for further performance boost. In terms of device structure, an RRAM has a two-terminal structure with stacked materials. The simplicity of device gives rise to the advantages of scalability and integration density.

Memristors, such as RRAMs, are normally incorporated in array configurations to achieve large data capacity for actual computing applications. In operation of such memristive array, the sneak path issue has been acknowledged as one of the largest challenges for revealing its applicable values in computations. The presence of sneak paths in the memristive array affects the controllable access to the memristors, which disables reliable array operations to write and read designated memristors. As a result, various types of selecting devices have been studied to be integrated with memristors. Both electrical characteristic matching and manufacturing process compatibility should be considered for device integration. Concerning integration simplicity in terms of fabrication and circuit design, a two-terminal selecting device is preferred such that it can connect a memristor, which also has two terminals, by direct device stacking. Chalcogenide selector devices with ovonic threshold switching (OTS) behaviors have a prevailing potential, but such novel devices are still under development for a convincing performance realized in an integrated memristive array. Owing to the technological matureness on device manufacturing and electrical characteristic tuning, transistors have been widely used as the selecting devices in memristive arrays. Depending on the CMOS technology platform and the memristor type, process integration remains a pivotal task for realizing functional memristive arrays. Furthermore, apart from the issue of electrical sneak paths, which can be diminished by the use of selecting devices, thermal effect can possibly affect array operation when device miniaturization enhances the device density of memristive array. Even though an experimental study is missing to demonstrate the potential thermal effect, a few simulation studies already revealed the possible impacts on array operations [22-26].

1.2 Scope of this work

To realize memristive arrays for the CIM and NC, this work is dedicated to the back-end-of-line (BEOL) integration, and electrical characterization of RRAM arrays with the major focus on valance-change memory (VCM) based on amorphous TaO_x. The research outcomes are organized into following three main chapters:

Chapter 4 presents the integration of TaO_x RRAMs and planar MOSFETs for realizing active RRAM arrays including one-transistor-one-RRAM (1T-1R) configuration and 1T-nR configuration, where one transistor is deployed for multiple RRAMs. While a TaO_x RRAM was used, a high-k metal-gate (HKMG) transistor was adopted as an active selecting device in the memristive arrays. Based on silicon-on-insulator (SOI) wafers, the entire design and fabrication were started from scratch using micro-scale semiconductor manufacturing technology. The fabrication development is discussed in detail including solutions to process issues. By using the developed active memristive array, a stateful logic gate was studied as CIM implementation.

Scaling down the RRAM integration technology, Chapter 5 presents the integration of nanoscale RRAMs on silicon circuits taped out with TSMC 180 nm technology node for the NC. Following the introduction of BEOL process integration, electrical characterization brings out the discussions on process issues and the yield enhancement of integrated RRAMs by process improvement. A verification test was also demonstrated on a customized operating hardware for the packaged chip.

To understand possible impacts from thermal effect when device spacing shrinks to pursue a higher device density in an RRAM array, Chapter 6 introduces a crosstalk study in passive crossbar arrays of nanoscale TaO_x RRAMs. After describing the design and fabrication flow of the RRAM arrays, constraints of available fabrication processes are discussed with possible improvements to achieve minimum device spacing. Investigations of retention and bit-flips on

unselected RRAM cells are summarized in the end according to programmed electrical measurements.

Chapter 2
Fundamentals

2.1 Filamentary switching mechanism of valence-change memory

Within the family of RRAM, the class of VCM shares the features of metal-insulator-metal (MIM) stacking and two-terminal device structure, and it has a resistive switching mechanism based on ion migrations in metal oxides accompanied by local valence changes [27]. The VCM stack consists of an n-type mixed ionic-electronic conducting (MIEC) oxide layer commonly sandwiched by an ohmic electrode, a metal with high oxygen affinity, and a high-work-function electrode, a chemically inert metal. As the critical switching layer, the n-type MIEC oxide mainly utilizes the oxygen cations as mobile ions, whose movements are commonly described by oxygen vacancy ($V_{\ddot{o}}$) migrations in oxygen-deficient oxides. Upon the presence of $V_{\ddot{o}}$, local valence changes in the cation sublattice occur as the result of nanoionic redox reactions, where the oxygen anions are oxidized and the metal cations are reduced. The consequent stoichiometry change alters the electron transport mechanism, where the $V_{\ddot{o}}$ are regarded as donor-type defects, and thus the electronic conductivity. Besides, the ohmic electrode serves as an oxygen exchange layer owing to its high oxygen affinity. Through the redox reaction occurring at the electrode interface, $V_{\ddot{o}}$ can be introduced into or eliminate from the MIEC oxide, which makes the ohmic electrode substantially the $V_{\ddot{o}}$ reservoir. The direction of the redox reaction can be manipulated by the electric bias polarity during the operation of the VCM device.

The VCM devices can be categorized into area-dependent/interfacial switching and filamentary switching according to the geometrical location of resistive switching events. The filamentary switching, as the focus of this work, relies on a conductive filament formed by the $V_{\ddot{o}}$ to control the resistance state. When the conductive filament connects the two electrodes, an electronic conducting path is created, resulting in a low resistance state (LRS). In contrast, when the conductive filament cannot provide a complete path from one electrode to the other, the device is then in a high resistance state (HRS).

For the class of filamentary-type VCM, a variety of switching materials has been studied, including binary and ternary metal oxides, such as TiO_2 [28, 29], HfO_2 [30, 31], Ta_2O_5 [32], ZrO_x [33], Al_2O_3 [34], and Nb_2O_5 [35], and $SrTiO_3$ [35, 36]. In addition to monolayer metal oxides, homogenous and heterogeneous bilayer metal oxides both have been investigated as material engineering solutions for the switching layer [20, 37-40]. Due to the fact that the majority of switching layers used in the filamentary-type VCM is highly insulating as fabricated, an electro-forming step is required to generate the conductive filament in the first place by electrically manipulating the $V_{\ddot{o}}$ in the system, except for some forming-free switching layers, which obtain high $V_{\ddot{o}}$ concentrations through material growth engineering and electrode selection [41-46]. Macroscopically, the electro-forming step can be regarded as a soft dielectric breakdown process that is reversible to some extent so that the device resistance can be subsequently modulated by electrical stimuli. Microscopically, when an electrical bias is

applied positively on the ohmic electrode over the high-work-function electrode of a pristine VCM device, the $V_{\ddot{O}}$ migration takes place in the direction towards the high-work-function electrode because the $V_{\ddot{O}}$ are positively charged. Besides, the electrical bias enables the $V_{\ddot{O}}$ formation at the ohmic electrode interface by redox reactions, which release oxygen ions in the metal oxide lattice and oxidize the metal used for the ohmic electrode. Owing to the introduction and migration of the $V_{\ddot{O}}$ inside the metal oxide, the $V_{\ddot{O}}$ are spatially arranged into a geometry that expands from the ohmic electrode to the high-work-function electrode, eventually forming a connecting path or so-called conductive filament between them. With this electronic conducting path, the device is then switched from the pristine state with a very high resistance to the LRS, which is the desirable result of electro-forming in the filamentary-type VCM device.

Following the formation of conductive filament by the electro-forming process, electrical biasing in different polarities can initiate the reversible resistive switching processes. The typical filamentary switching characteristic is illustrated in Figure 2.1 by taking the example of a ZrO_x-based VCM where a Zr layer and a Pt layer are deployed as the ohmic electrode and the high-work-function electrode respectively [47]. The explanatory illustration starts from status A, where the device is in the HRS. As depicted in the zoomed-in area of device cross section, the conductive filament formed by the $V_{\ddot{O}}$ does not connect the two electrodes. The existing conductive filament is considered as the plug region, while the gap between the plug region and the Pt electrode is regarded as the disc region. The disc region has a high resistance,

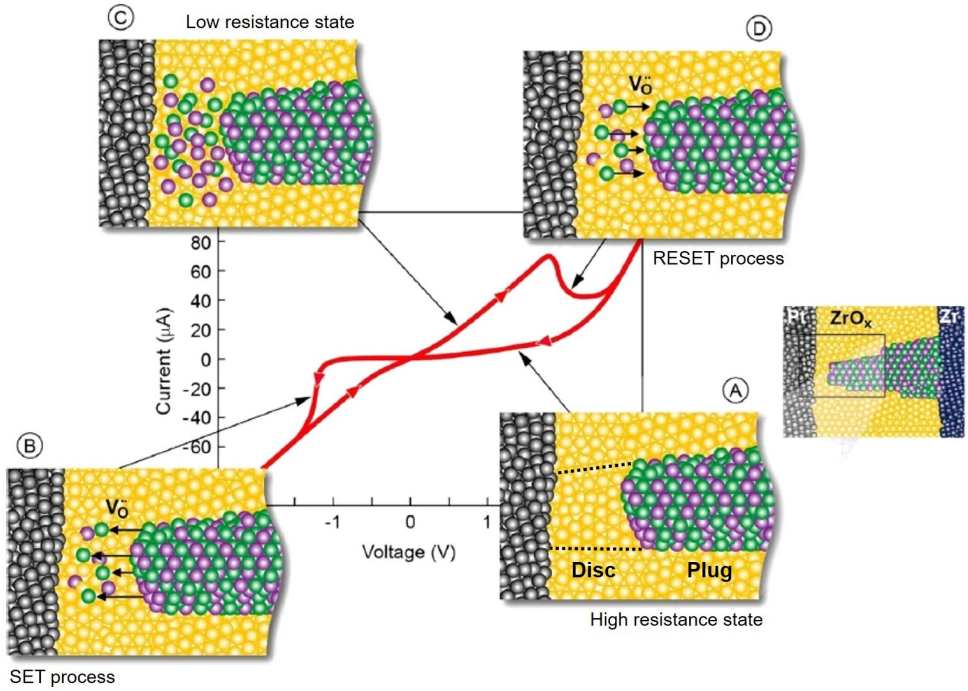

Figure 2.1: *Resistive switching of a typical filamentary-type VCM. A full switching cycle involves a SET and a RESET processes that enable transitions between the LRS and HRS. When the Zr ohmic electrode of a ZrO_x-based VCM connects to the ground, negative voltage biasing on the Pt high-work-function electrode can trigger the SET process. Oppositely, positive voltage biasing can trigger the RESET process. Adapted and reprinted with permission from [47]. © 2012 by Wiley-VCH.*

so the overall device resistance is high. When applying a negative voltage on the Pt electrode (or a positive voltage on the Zr electrode), the $V_{\ddot{O}}$ start to migrate at a critical voltage from the plug to the disc region according to the applied electric field, which is referred to status B and the so-called SET process. Once the $V_{\ddot{O}}$ are accumulated at the disc region to a critical extent that brings a substantial resistance decrease, the device is then switched to the LRS, which is depicted as status C. Here, the LRS generally depends on the current flowing through the device such that a higher current results in a lower resistance. Thus, it is possible to control the LRS by a current compliance mechanism during the SET process, which limits the loading current from exceeding a preset value by removing an appropriate portion of preset voltage from the device under test in real time. By applying a positive voltage on the Pt electrode (or a negative voltage on the Zr electrode) when the device is in the LRS, the $V_{\ddot{O}}$ at the disc region start to migrate at a critical voltage from the disc to the plug region according to the applied electric field, which is referred to status D and the so-called RESET process. Generally, the RESET process stops at the maximum applied voltage, which determines the eventual HRS such that a higher maximum applied voltage results in a higher resistance. Following the same sequence, the device can be switched between the LRS and HRS spontaneously by electrical biasing.

In this work, the I-V characteristic of bipolar resistive switching is described with the voltage defined by the electrical potential difference with respect to the high-work-function electrode. Figure 2.2 shows the typical I-V diagram on a logarithmic scale. The positive voltage represents the positive biasing on the ohmic electrode with respect to the high-work-function electrode, which is able to trigger the SET process. Hence, the positive voltage is regarded as the SET voltage (V_{SET}). Particularly, the voltage triggering the SET process is defined as the critical SET voltage ($V_{SET,C}$), which induces a significant current jump and a sudden decrease in resistance. Once the increasing current reaches a preset current compliance (I_{CC}) value, the SET process is stopped and the loading current is maintained at the I_{CC} by relieving the voltage applied on the device, even when the external applied voltage is increasing. On the other hand, the negative voltage implies the negative biasing on the ohmic electrode with respect to the high-work-function electrode, which is able to trigger the RESET process. The negative voltage, therefore, is regarded as the RESET voltage (V_{RESET}). Here, the current induced by the

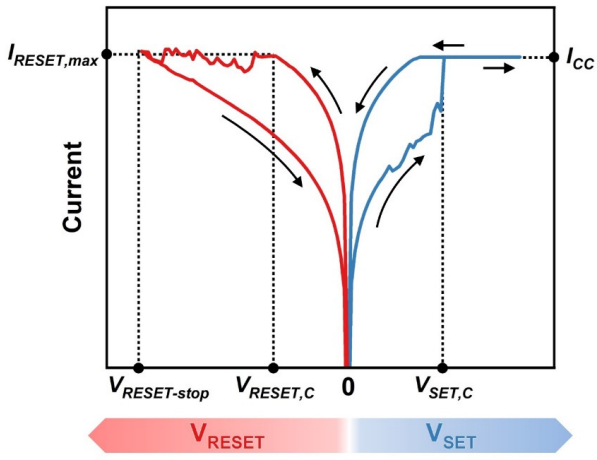

Figure 2.2: *I-V diagram of bipolar filamentary-type resistive switching on logarithmic scale. The definition of resistive switching parameters adopted in this work is illustrated such that the blue and the red curves indicate the SET and the RESET processes respectively.*

V_{RESET} (I_{RESET}) is plotted with its absolute value for illustrative purposes. As the I_{RESET} increases with rising V_{RESET}, the RESET process is triggered at a critical voltage ($V_{RESET,C}$), where the device resistance increases, and the I_{RESET} shows a lower increasing rate or even a decrease over a voltage increase. The highest current that can be sustained in the LRS before the actual RESET process takes place is regarded as the maximum RESET current ($I_{RESET,max}$). The RESET process ceases at the highest applied V_{RESET} or the so-called RESET-stop voltage ($V_{RESET-stop}$), from which the V_{RESET} starts to decrease and the device is able to stay in a stable HRS. Furthermore, the ON/OFF ratio is defined as the conductance ratio of LRS over HRS or the resistance ratio of HRS over LRS, which is technically obtained by the current ratio of LRS over HRS at a very low constant voltage, for instance between 0.1 V and 0.2 V.

2.2 Design of RRAM arrays

To scale up the memory capacity, multiple memory units incorporate in a memory array, where they are arranged in a repetitive manner. For RRAM arrays, one RRAM device can be used solely as a memory unit or unit cell without any other electronic component. Besides, an RRAM device can be combined with another electronic component to form a unit cell. Depending on whether there is an active switch to spontaneously control the access to the RRAM in a unit cell, the arrays are categorized in passive and active RRAM arrays. In this regard, the concepts and the relevant concerns of both array architectures will be introduced in this subchapter.

2.2.1 Passive RRAM arrays

Directly using one single RRAM device as an individual memory unit, the simplest array can be built solely by RRAMs. Owing to the two-terminal feature of RRAM devices, this array can be produced intuitively in a crossbar structure shown in Figure 2.3a, taking a 2×2 array as the example. To address unit cells in the array, there are two set of parallel metal lines perpendicular to each other, namely word lines and bit lines. One set connects RRAM devices on one terminal, while the other set connects on the other terminal. By selecting one of each set, a certain RRAM or unit cell can be addressed in the crossbar array for READ and WRITE operations. Since there is no extra electronic component actively controlling the access to the RRAM in each memory unit, it is categorized as the passive RRAM array. According to the crossbar structure of passive RRAM array, the physical size of the unit cell can be determined as $4F^2$, where F is the minimum feature size of manufacturing, as shown in Figure 2.3b. Favorably, the compact size is competitive compared to other memory arrays based on conventional devices and architectures, such as SRAM, DRAM, and flash memories.

Apart from the advantage of the compactness for realizing high-density memory arrays, this type of passive RRAM arrays is prone to a major drawback of sneak paths, which is illustrated in Figure 2.4. When addressing a unit cell by a selected bit line and a selected word line, electronic sneak paths can be formed through the adjacent unit cells, especially if they are in the LRS. One possible sneak path is shown in Figure 2.4. On the one hand, extra currents can flow via the sneak paths, causing changes in the current collected from the selected word line, i.e., the output of READ operation. On the other hand, unintended voltage drops can occur along the sneak paths across the unselected cells, on which resistive switching can be possibly triggered. The consequent change of resistance state affects the stored data in the memory unit, and thus it is regarded as an operating error due to an unintended WRITE operation.

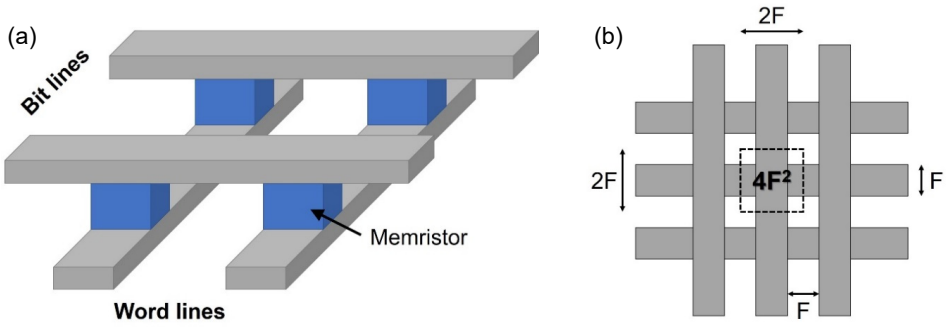

Figure 2.3: *Schematic diagram of passive RRAM arrays. (a) A 2×2 crossbar array of RRAMs. (b) Top view of a passive crossbar array. The unit cell theoretically only requires an area of $4F^2$, when F is the minimum feature size of manufacturing.*

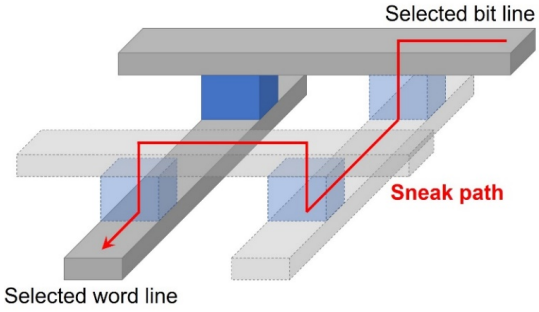

Figure 2.4: *Schematic diagram of a potential sneak path in a passive RRAM array.*

To suppress the effect of sneak paths, several operating schemes or biasing schemes have been proposed such that neither any bit line nor any word line is left floating during operations [48, 49]. The two most common operating schemes are the $V/2$ scheme and the $V/3$ scheme, both of which have an operating voltage (V) applied on the selected bit line, while the selected word line is grounded, as shown in Figure 2.5. The distinction lies in the voltage biasing on unselected bit lines and word lines. For the $V/2$ scheme, the unselected bit lines and word lines are biased by $V/2$. As a result, the cells half selected by either the selected bit line or the selected word line carry a forward bias of $V/2$, while the non-selected cells at the intersections of an unselected bit line and an unselected word line have a zero bias. For the $V/3$ scheme, the unselected bit lines are biased by $V/3$, whereas the unselected word lines are biased by $2V/3$. Except the selected cell forward biased by V, all other cells are averagely applied with an absolute bias of $V/3$ such that the half-selected cells are biased forward, while the non-selected cells are biased reversely. Compared to the $V/2$ scheme, the half-selected cells have a lower voltage bias at the cost of reverse biasing on the non-selected cells. In general, the use of operating schemes can limit the unintended voltage drops across the unselected memory cells to constant values, which improves the controllability during array operations and mitigates the issue of sneak paths.

Besides the operating schemes, various selecting electronic components have been studied and demonstrated in combination with RRAM devices to further diminish the issue of sneak paths [50-52]. In the framework of passive RRAM arrays, selecting electronic components with

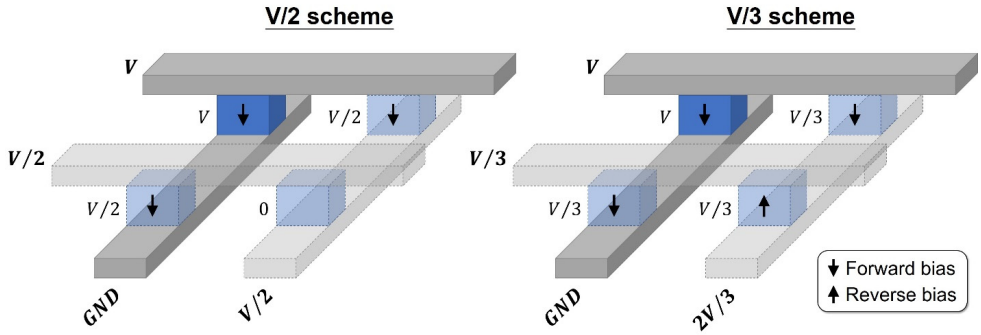

Figure 2.5: *Schematic diagram of the $V/2$ scheme and the $V/3$ scheme for passive RRAM array operation. V is the desired operating voltage on the selected cell. The theoretical voltage bias on each cell in the array during operation is marked aside with the direction indicated by an upward or a downward arrow, implying a forward or a negative bias respectively.*

two terminals, commonly called selectors, can be added to the arrays directly by device stacking without changes in the array architecture and additional metal connections, as shown in Figure 2.6. Each selector directly connects an RRAM in series, forming a memory unit addressed similarly by a bit line and a word line, thereby holding the cell size as $4F^2$. A selector does not require an active switch controlled by an extra voltage bias, and it works passively in a memory unit, leading to a passive RRAM array. Technically, selectors are non-linear resistors with volatility, whose resistances decrease as the applied voltage increases and return to the original values as the applied voltage is removed. Owing to a critical increase in the overall non-linearity of the unit cells, the presence of selectors can suppress the effect of sneak paths in terms of unintended WRITE operations on unselected cells and influences on the current output of READ operations. Except the RRAMs showing unipolar resistive switching, the diode-type selectors do not match the RRAMs showing bipolar resistive switching, such as VCM-type RRAMs. In this regard, a variety of selectors that function in both biasing polarities have been investigated, e.g. ovonic threshold switching (OTS) devices based on chalcogenides. The integration of one-selector-one-RRAM (1S-1R) array has also been demonstrated using different selectors and RRAMs [53-57].

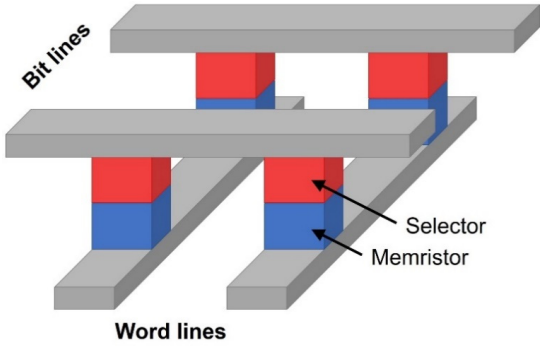

Figure 2.6: *Schematic diagram of a 1S-1R passive RRAM array.*

2.2.2 Active RRAM arrays

Based on the mature Si-based semiconductor technology, MOSFETs are the most common electronic switches used in integrated circuits, which can be ideal selecting components in RRAM arrays. The MOSFET is operated by three terminals such that a voltage bias on the gate terminal can control the current flowing from the drain terminal to the source terminal through the charge carrier channel underneath the gate dielectric. In other words, the MOSFET is an electronic switch with an active switch control by the gate voltage. By connecting an RRAM device to the drain of an MOSFET in series to form a memory unit, the RRAM can only be electrically accessed when the charge carrier channel is switched on by applying a certain gate voltage. An illustrative cross section of the one-transistor-one-RRAM (1T-1R) unit cell is displayed in Figure 2.7a with the equivalent electrical circuit depicted in Figure 2.7b. The 1T-1R unit cells can be arranged to form an active RRAM array, in which bit lines connect the open terminals of RRAMs, whereas word lines and source lines connect the gate and the source terminals of MOSFETs respectively, as shown in Figure 2.7c. When the access to an RRAM device is activated by applying an appropriate gate voltage via the word line, the selected RRAM device can be operated by applying the desired operating voltage via the bit line, meanwhile the output current can be measured from the source line correspondingly.

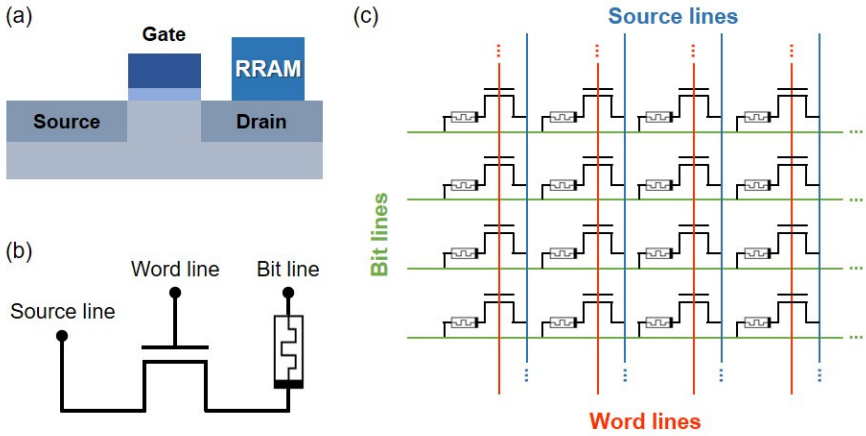

Figure 2.7: *Schematic diagrams of 1T-1R unit cell and array. (a) Structure cross section and (b) circuit diagram of 1T-1R unit cell. (c) Circuit diagram of a typical 1T-1R array.*

In comparison to the passive RRAM arrays, the array architecture is more complicated in the active RRAM arrays, with an extra set of gate voltage inputs required for manipulating MOSFETs. Due to a more complex device structure, MOSFETs not only require more fabrication steps than selectors, which rises the fabrication cost, but also need a larger fabrication area, implying a 1T-1R array with the same memory capacity consumes a larger area on chip. However, the technology maturity of MOSFET ensures superior electrical characteristics for the demanded selecting feature in RRAM arrays. Furthermore, miniaturization of MOSFET has been advancing for decades from breakthroughs in materials and innovations in device structure, which evolves from planar to vertical structures, e.g. FinFETs and more recently gate-all-around or nanosheet FETs. With the fabrication compatibility of RRAMs towards CMOS technology, the advanced and reliable manufacturing capability from semiconductor foundries on the market can foster the commercialization of 1T-1R and other types of active RRAM arrays.

Chapter 3
Methods

This research work encompasses the fabrication of micro/nanoelectronics, analysis of materials, and characterization of electrical properties. This chapter will introduce pertinent methods to provide a foundational grasp of the technology and instrument solutions employed.

3.1 Fabrication methods

A series of semiconductor manufacturing techniques are commonly used to fabricate microscale and nanoscale semiconductor devices, including not only transistors but also RRAM devices. According to their functions, they can be mainly categorized into three groups, deposition, etching, and lithography. This subchapter will explain the working principle of different techniques in each group that were employed in this work.

3.1.1 Deposition

Deposition is a manufacturing process of material growth especially for thin films with a thickness ranging from a few μm down to a few Å. As a basic building block to fabricate semiconductor devices, material stacking can be achieved by multiple deposition steps. The choice of deposition technique depends on the material specification, substrate condition, and process compatibility. According to the growth mechanism, deposition techniques can be generally classified into physical vapor deposition (PVD) and chemical vapor deposition (CVD). They both involve depositions of atoms from the vapor phase onto a solid substrate. The PVD is mainly based on physical effects, whereas the CVD is mainly based on chemical reactions. Two critical PVD techniques frequently used in this work, sputter deposition and thermal evaporation, will be presented in detail.

3.1.1.1 Sputter deposition

Within the PVD family, sputter deposition operates through sputtering of a source material, also called a target, by ion bombardments [58]. This process relies on the generating electrically induced plasma between a cathode and an anode. A typical sputter deposition system is illustrated in

Figure ***3.1***. Gas species are introduced via gas flow controllers into a vacuum chamber, where they are collided with high-energy electrons accelerated by a high voltage applied between the electrodes, and hence ionized into a plasma. Typically, Ar, a chemically inert gas, is solely used. In this case, the argon ions in the plasma, accelerated by the applied voltage, bombard the target, which is attached to the cathode. As a result, atoms or molecules are sputtered off the target and deposited on the substrate surface and the vacuum chamber wall through random walks in the vapor phase. The sputter deposition requires low pressure

normally from 10^{-6} to 10^{-3} bar, to ensure a sufficiently large mean free path for the particles involved in the process. This enables free electrons to gain enough kinetic energy for the plasma ignition and allows sputtered particles to reach the substrate with fewer collisions, promoting deposition while preventing a low deposition rate. The sputter deposition offers superior adhesion and uniform coverage due to its isotropic sputtering nature. However, in lift-off patterning, the isotropic deposition can hinder an effective lift-off, as the deposited thin film covers the sidewalls of the resist profile, posing a challenge for solvent-based lift-off.

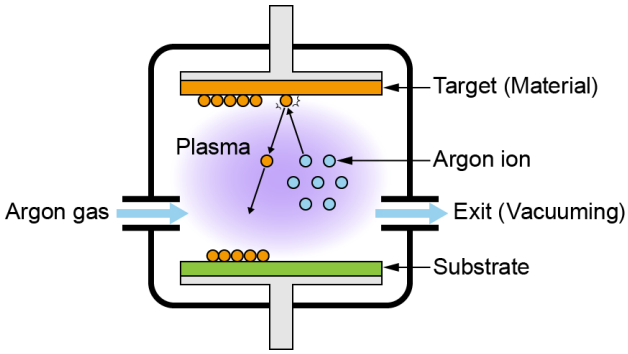

Figure 3.1: *Schematic diagram of a general sputter deposition system. Adapted from [59].*

According to the type of voltage source, the sputter deposition is mainly categorized into direct current (DC) and radio frequency (RF) sputtering. The DC sputtering is the most basic and cost-effective type of sputter deposition, mainly for electrically conductive materials. When an insulating material is used as the target, it can hold the charges both on the target and the deposited thin film during the deposition process, affecting the growth quality and even causing cessation of the deposition. In this regard, the RF sputtering is a technical alternative to overcome this limitation. By alternating the electrical potential at RF, the charges built up in a depositing cycle can be neutralized successively in a discharging cycle. Despite the lower deposition rate when applying the same voltage, the RF sputtering can ensure an appropriate growth of insulating materials compared to the DC sputtering.

Besides using compound targets, reactive gases, such as oxygen and nitrogen, can be added to an inert gas, such as argon, to deposit compound materials when single-element targets are utilized, i.e., reactive sputtering. For instance, a tantalum oxide can be deposited by using a Ta target and oxygen gas in addition to argon gas. The plasma of the gas mixture contains oxygen ions, which react with the sputtered Ta atoms. Consequently, the tantalum oxide is formed as the product deposited on the substrate.

In this work, the sputter deposition is mainly performed by the Univex 450C cluster tool from Oerikon-Leybold, which consists of six sputtering chambers, a loading chamber, and a transfer chamber with a robotic arm. Once a substrate is loaded into the tool, the robotic arm can transfer it between different sputtering chambers to achieve in-situ sputtering of different materials, which is ideal for producing material stacks in micro- and nanoelectronics. For the TaO_x-based RRAM in this work, the electrode materials, Pt and Ta, and the TaO_x switching layer were all deposited in this sputtering system. The Pt and Ta are both sputtered by a single-element target in a pure Ar gas atmosphere. While DC sputtering is used for the Pt, RF sputtering is used for the Ta. The Ta sputtering chamber is also utilized for reactive RF sputtering of the TaO_x by using the same Ta target and a gas mixture of argon and oxygen. The parameter details of the sputter depositions are listed in Table 3.1.

Table 3.1: *Process parameters of sputter depositions for the Pt, Ta, and TaO_x thin films involved in this work.*

Parameters	***Pt***	***Ta***	***TaO_x***
Power source	DC	RF	RF
Power (W)	500	116	236
Gas flow (sccm)	60 (Ar)	60 (Ar)	46/14 (Ar/O_2)
Pressure (mbar)	2.4×10^{-2}	2.4×10^{-2}	2.3×10^{-2}
Deposition rate (nm/min)	130	10	3.25

3.1.1.2 Thermal evaporation

Thermal evaporation stands out as one of the most popular PVD techniques due to its operational simplicity and high deposition rate. Essentially, it employs thermal energy to evaporate a target material, known as the evaporation source, into the vapor phase. The vaporous material subsequently condense on a relatively cold substrate surface, facilitating deposition. Distinguished by the heat source, resistive evaporation and electron beam (e-beam) evaporation are the most common methods, respectively illustrated in Figure 3.2. They both operate in a vacuum chamber with the evaporation source placed at the bottom and the substrate mounted at the top. A high vacuum environment with pressure typically below 10^{-5} bar is maintained to ensure a sufficient mean free path for the vaporized target material to reach the substrate without significant collision with ambient gas molecules. In the resistive evaporation, the target material evaporates due to resistive heating induced by electric current passing through a heater typically in the form of a filament, boat, or crucible. The heater is composed of refractory metals, such as W, Mo, or Ta, often with ceramic coatings. The choice of target material is limited to those with lower evaporation points than the heater material. E-beam evaporation, in contrast, utilizes an e-beam generated by thermionic emission, field electron emission or the anodic arc method to evaporate the target material housed in a water-cooled crucible, usually made of Cu, W, Mo, Ta, boron nitride, or graphite. This favorably enable deposition of refractory metals and other materials with high evaporating points. Moreover, the

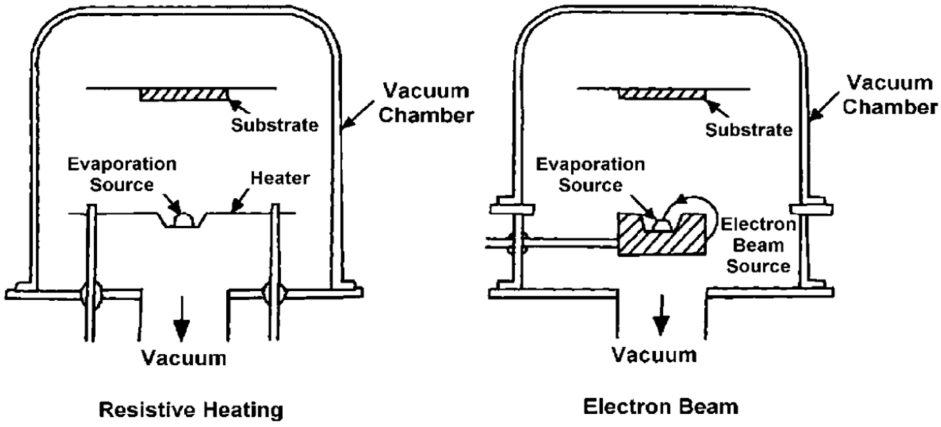

Figure 3.2: *Structure schematics of thermal deposition systems configured with resistive heating or an electron beam [60]. © 2004 by William Andrew Publishing.*

use of e-beam can offer uniform heating, as magnetic fields can direct the e-beam, while additional electric fields steer it over the target surface.

The thermal evaporation typically results in directional and anisotropic deposition via a line-of-sight process. While it could lead to poor coverage on complex morphology, the anisotropic nature proves advantageous in lift-off patterning, as it helps avoiding deposition on resist sidewalls. In this work, the thermal evaporation was employed for the deposition of thick metal layers, including Al metallization layers, Pt metal pads/connections, and Cr hard masks.

3.1.2 Etching

Etching in semiconductor manufacturing is a process of material removal from a substrate, which enables creation of intricate patterns on semiconductor wafers for micro- and nanoelectornics fabrication. It is commonly categorized into wet etching, which employs chemical liquid solutions, and dry etching, which utilizes chemical reactions and/or ion bombardment in plasma. Involving immersion in a chemical etchant, the wet etching tends to be less directional and could etch the material under the mask. In contrast, the dry etching poses concerns about damaging the mask layer and the preserved layers/structures due to plasma exposure or ion bombardment. Tailoring the etching method to a specific application is crucial for achieving the desired result. In this work, the focus lies on dry etching techniques, including reactive ion etching (RIE), reactive ion beam etching (RIBE), and barrel plasma etching. These methods will be discussed in detail, considering their respective working principles and applications in the fabrication processes [61, 62].

3.1.2.1 Reactive ion etching (RIE)

RIE uses ions in the plasma of reactive gases (e.g. Cl_2, SF_6, CF_4, CHF_3, O_2, H_2) with or without non-reactive gases (e.g. Ar) to chemically and physically remove materials from the substrate surface. Parallel plate RIE is the most basic RIE systems, whose configuration is similar to RF sputter deposition but operated by the opposite voltage biasing polarity. It is equipped with two parallel electrodes placed inside a vacuum chamber and biased by an RF power supply. The electrode attached to the substrate is supposed to be the cathode, whereas the other is the anode. The applied electric field accelerates positively charged ions in the electrically induced plasma towards the substrate, resulting in chemical reactions and physical sputtering at the surface simultaneously. The combination of these two mechanisms turn atoms or molecules on the substrate surface into vaporous etching byproducts, which are then pumped out from the vacuum chamber to the exhaust system.

The *PlasmaPro 100 ICP RIE* system from Oxford Instruments Plasma Technology employed in this work is an advanced RIE system configuring an alternative inductively coupled plasma (ICP) generator, whose system schematic is displayed in Figure 3.3. Besides RF voltage biasing on the two electrodes, plasma ignition can be alternatively achieved by an oscillating magnetic field induced from an RF coil wrapped around the vacuum chamber. In this case, a separate RF voltage biasing is meanwhile utilized to control the ion energy and directionality. Typically, gas is introduced from the top through a showerhead gas inlet owing to the uniformity, and exhausted from the bottom, where a vacuum pump connects to maintain the working pressure and remove the vaporous etching byproducts. Wafer clamping with He backside cooling is configured for optimum wafer temperature control.

In this work, the RIE was used solely with RF voltage biasing because a less aggressive and relatively slow etching was favorable for process control. The silicon oxides and silicon nitrides serving as passivation layers were etched by using the reactive gas of CHF_3 with or without

Ar. When a Cr hard mask is used, the Pt electrode layer was etched by using the reactive gas of Cl_2 with Ar. The RIE parameter details are listed in Table 3.2.

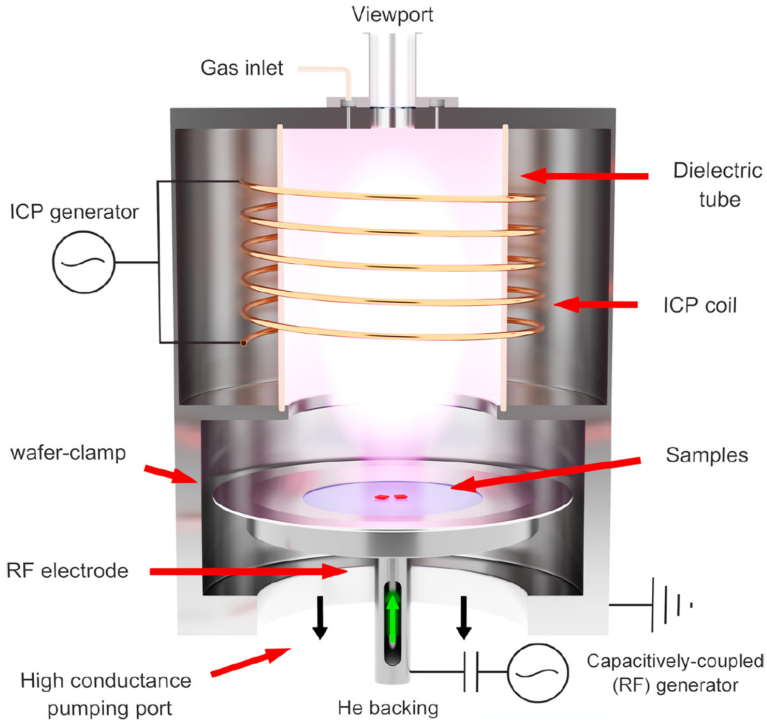

Figure 3.3: *Structure schematic of the PlasmaPro 100 ICP RIE system from Oxford Instruments Plasma Technology [63].*

Table 3.2: *Process parameters of RIE for different material stacks involved in this work.*

Parameters	***SiO_2***		***$SiO_2 + Si_3N_4$***	***Pt***
Temperature (°C)	20	0	0	20
Power (W)	200	200	140	150
Gas flow (sccm)	30 (CHF_3)	30 (CHF_3)	50/50 (Ar/ CHF_3)	16/40 (Ar/Cl_2)
Pressure (mbar)	2×10^{-2}	2.4×10^{-2}	4×10^{-2}	1×10^{-2}
Etching rate (nm/min)	35	27	-	9

3.1.2.2 Reactive ion beam etching (RIBE)

RIBE can be regarded as a variant of RIE that utilizes the plasma as an ion beam source instead of the direct etching agent. In conventional RIE, the plasma is in proximity to the substrate, and the ions are accelerated by the applied voltage all the way towards the substrate, which is directly attached to the cathode. In RIBE, on the other hand, the plasma is generated

externally in an ion beam source, where the ions are accelerated and eventually ejected in the form of parallel beams towards the substrate. Figure 3.4 depicts a simplified system configuration of the RIBE used in this work, the Ionfab Ion Beam system from Oxford Instruments Plasma Technology, with a clockwise rotation of 90 degrees. An ICP generator with an RF coil ionizes the introduced gas and creates a high-density plasma in the ion beam source unit. The ions are electrically accelerated towards the grids, where they exit as collimated ion beams without further acceleration in the direction of the substrate. When the ion beams impact on the substrate, the atoms or molecules are removed from the surface simultaneously by chemical reactions and physical sputtering. The vaporous etching byproducts are pumped out from the bottom to the exhaust system. Owing to the precise control over the ion energy and directionality attributed to the design of ion beam source, RIBE has a high anisotropy and a superior selectivity in comparison to conventional RIE. Especially for chemically inert materials, ion beam milling, the etching merely enabled by the physical sputtering of ion bombardment, is available in the same RIBE system by using an inert gas, such as Ar.

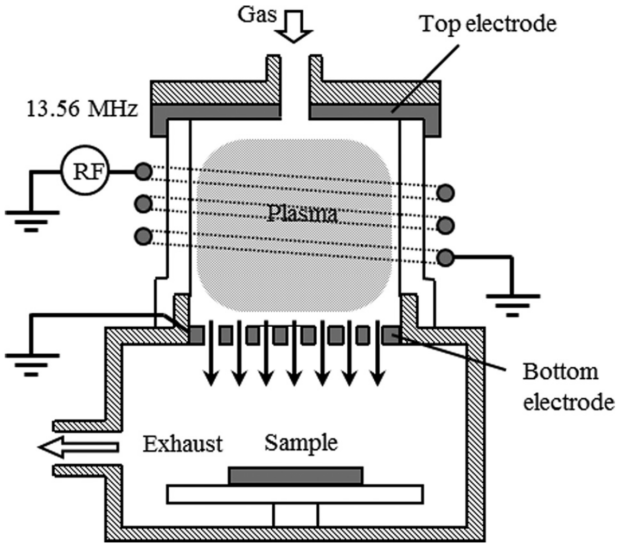

Figure 3.4: *Structure schematic of RIBE system with a clockwise rotation of 90 degrees [64].*

In this work, the RIBE system was employed to etch the RRAM device stacks of the interest. The metal oxides (e.g. TaO_x, TiO_x, and HfO_2) were etched by using CF_4 gas. The metal electrodes (e.g. Pt, Ta, and Ti) were etched mainly by using Ar gas.

3.1.2.3 Barrel plasma etching

Barrel plasma etching relies on a barrel reactor that generates a high-density plasma by either RF voltage biasing or microwave radiation. As an isotropic dry etching technique, it operates in a high pressure, typically larger than 0.2 mbar, because the resulting short mean free path of ions in the plasma critically mitigates the contribution of ion bombardment in the etching process, which is therefore focused on the chemical rather than the physical mechanism [65].

In this work, the barrel plasma etching was performed by the *GIGAbatch* microwave plasma system from PVA TePla, whose core concept of the system structure is illustrated in Figure 3.5. A magnetron generates microwaves, more specifically electromagnetic waves with the

frequency of 2.45 GHz. The microwaves are transmitted along a waveguide and emitted through an antenna to a barrel-type quartz chamber. The electrons oscillated by the microwaves collide with gas atoms or molecules, leading to ionization of the gas and, therefore, creation of plasma in the chamber. Subsequently via chemical reactions, the ions remove materials from the substrate surface and turn them into vaporous byproducts, which are pumped out from the end opposite to the gas inlet. The use of microwave in plasma generation, compared to RF voltage biasing, not only broadens the available pressure range, but also enables lower ion bombardment energies, thereby reducing the surface damages in the plasma etching process.

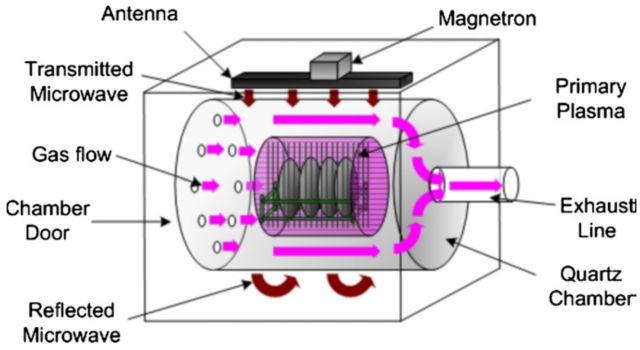

Figure 3.5: *Structure schematic of barrel microwave plasma etcher [66].*

Oxygen gas finds widespread application in barrel plasma etching for resist ashing, an effective method to remove lithography resists based on organic materials. Additionally, mild plasma etching utilizing oxygen gas is commonly employed for descum, which targets the removal of resist residue in developed areas after lithography. To improve the descum result by achieving gentler and more uniform etching, a Faraday cage in the chamber can be utilized, with the substrate positioned at the center. The Faraday cage, essentially a metal mesh cylinder, shields the substrate from direct plasma exposure, thereby minimizing the physical sputtering contribution to the etching process. In contrast, perforated channels on the Faraday cage allow reactive radicals from the plasma to diffuse through, which enables chemical reactions at the substrate surface to remove materials. Details regarding the parameters of the ashing and descum processes conducted in this work can be found in Table 3.3.

Table 3.3: *Process parameters of barrel plasma etching for ashing and descum processes.*

Parameters	***Ashing***	***Descum***
Power (W)	600	600
Gas flow (sccm)	200 (O_2)	600 (O_2)
Pressure (mbar)	0.4	0.9
Faraday cage	No	Yes

3.1.3 Lithography

Lithography is one of the cornerstone techniques in microelectronics manufacturing, pivotal for generating specific patterns on wafers. The general lithography process, depicted in Figure 3.6, unfolds as follows: Initially, a photon- or electron-sensitive material, commonly referred

to as resist, is uniformly applied in liquid form onto a substrate via spin coating. During this step, the substrate rotates at several thousand revolutions per minute (rpm) to ensure uniform distribution of the resist across the substrate surface. The resulting thickness of the resist primarily hinges on its viscosity and the spinning speed. Typically, a softbake or prebake step ensues to diminish solvent content in the resist film. This not only enhances adhesion and stability of the resist film but also forestalls bubbling caused by solvent evaporation in subsequent processes. Subsequently, exposure occurs by illuminating specific areas on the resist surface with high-energy photons or electrons. The exposure induces chemical reactions altering the polymer structure and consequently, the chemical properties of the resist. For certain resists, a post-exposure bake (PEB) or postbake is necessary after exposure to finalize the chemical reactions for the desired property alteration. Finally, development transpires as a wet process employing a designated chemical solution, termed the developer, to dissolve and remove either the previously exposed or unexposed parts of the resist film, contingent on the type of resist employed. Positive resists exhibit higher solubility in areas previously exposed during the exposure step due to the chemical reactions induced by exposure. Thus, the unexposed regions persist on the substrate after development. Conversely, negative resists exhibit reduced solubility in the previously exposed areas, leading to exceptional chemical resistance to the developer. The exposed regions, therefore, remain on the substrate after development.

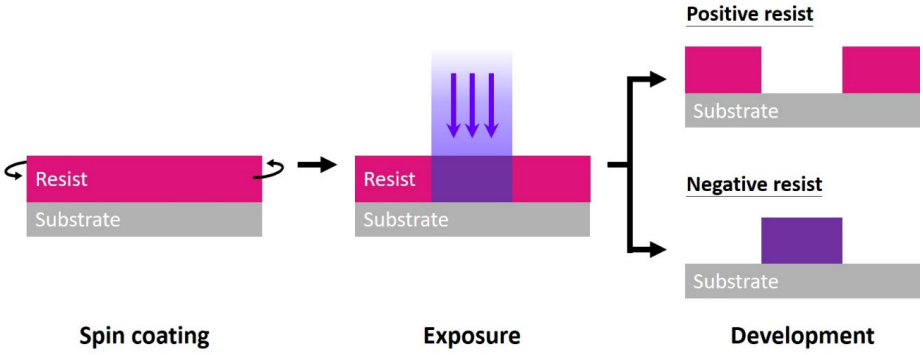

Figure 3.6: *General procedure of lithography in microelectronics manufacturing.*

With the patterned resist layer from lithography, the pattern can be transferred to a target layer by either lift-off or etching, as shown in Figure 3.7. For the lift-off approach, the use of negative resists is critical to obtain an undercut in the resist profile as the result of lithography, because negative resists are particularly designed with the chemical reaction mechanism serving the purpose. Deposition of the targeted thin film takes place after lithography such that it covers the resist surface and fill the open area. The last step is to remove the resist layer by acetone or other specific chemical solvents dedicated to certain resists. The undercut favorably creates a good exposure of resist sidewall to the chemical solvent, ensuring an excellent removal effectiveness. Due to dissolution of the resist layer, the deposited target layer on the resist is lifted off the substrate, which therefore gives the name, lift-off, to this process. Eventually, the resist pattern is reversely transferred to the target layer. In contrast, when the etching approach is adopted to pattern the target layer, the undercut feature is not required. Both positive and negative resists are compatible. Deviating from the lift-off approach, deposition of the targeted thin film must be completed before lithography. Based on the fact that resist materials are etching resistant, the patterned resist layer serves as the etching mask

such that wet or dry etching of the deposited target layer merely takes place in the open area. After removing the resist, the target layer previously covered by the resist remains on the substrate, implying the resist pattern is transferred identically to the target layer.

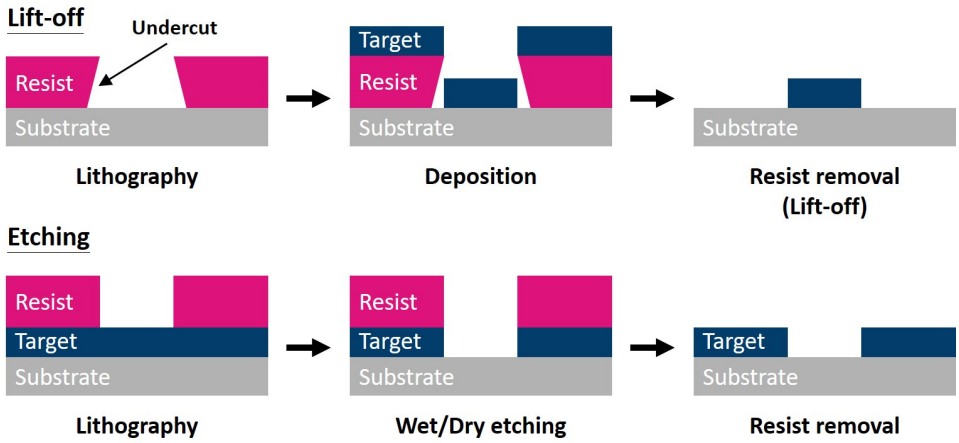

Figure 3.7: *General procedure of patterning by lift-off and etching approaches.*

3.1.3.1 Photolithography

Photolithography refers to the lithography methods utilizing photons in the exposure step [67]. To expose specific areas with photons, specialized exposure systems have been established and advanced over decades for light projection of patterns fabricated on photomasks. As the pattern projection involves refraction and diffraction when light traveling through optic lens and a photomask, the resolution limit can be derived from Abbe's formula or Rayleigh's criterion and commonly referred to as [68-70]:

$$CD = k_1 \frac{\lambda}{NA} . \tag{3.1}$$

Here, CD refers to critical dimension, which is the minimum feature size within the capability of photolithography. k_1 is a collective coefficient to correct the discrepancy from optical theory to realistic lithography technology, including optical aberrations, lens defects, resist behavior, pattern shape, etc. λ indicates the wavelength of light, and NA is the numerical aperture of optics, which describes the light-gathering ability. Generally, NA is defined by

$$NA = n \sin \theta , \tag{3.2}$$

where n is the refractive index of the media between the lens and the resist surface, and θ is the half-angle of the maximum light cone that can be focused by the optics. As CD is directly proportional to λ, a shorter λ leads to a smaller CD. In order to realize micro- and even nanoscale features, wavelengths within the ultraviolet (UV) light range, i.e., from 400 nm to 10 nm, are used in micro- and nanoelectronics fabrication with a continuous pursuit towards shorter wavelengths. Since the early 1960s, gas-discharge lamps using mercury, commonly called Hg lamps, have been used as the UV light sources for exposure in lithography by filtering out one of the spectral lines, e.g. the i-line at 365 nm. Based on excimer lasers invented in the 1970s, deep-ultraviolet (DUV) photolithography was developed in the early 1980s and has been massively used in semiconductor manufacturing nowadays by employing KrF or ArF excimer lasers with λ of 248 nm and 193 nm respectively [71]. Recently around 2020, extreme

ultraviolet (EUV) photolithography appears as the state-of-the-art exposure technology with λ near 13.5 nm enabled by a laser-pulsed tin droplet plasma [72].

In this work, photolithography was performed by using UV light sources with λ of 365 nm. Besides conventional mask aligners, which are standard exposure systems incorporating with photomasks, a maskless aligner (MLA) was employed as a technical solution for the encountered process issues, which saves the time and financial costs on photomask fabrication.

3.1.3.1.1 Mask aligner

A mask aligner is a typical exposure system for photolithography, where a substrate can be precisely aligned with respect to the mounted photomask for UV light exposure. Contact lithography, also known as contact printing, is adopted as the exposure method in this work, because it has an optic system less complex than projection lithography, thereby offering design simplicity and lower costs. In addition, it can achieve a better resolution than proximity lithography because the photomask is in direct contact with the photoresist layer coated on the substrate. If a gap appears in between, the resulting light diffraction during exposure would deteriorate the resolution limit. In contact lithography, parallel UV light emitted from the light source shines through a photomask to project the desired pattern on the photoresist layer, as shown in Figure 3.8a. The photomask is a fused silica plate that has one side covered with a desired pattern defined by a Cr film. This side is supposed to be in contact with the photoresist layer such that the Cr film can absorb the UV light and prevents the area underneath from exposure.

A *Süss MA6* mask aligner employed in this work is shown in Figure 3.8b. It is equipped with a UV light source based on a Hg lamp, of which the i-line was used in this work with the wavelength of 365 nm. The overlay of manual alignment can be less than 2 µm, and the feasible minimum of feature size is around 2 µm depending on the photoresist, the shape of desired pattern, and the substrate.

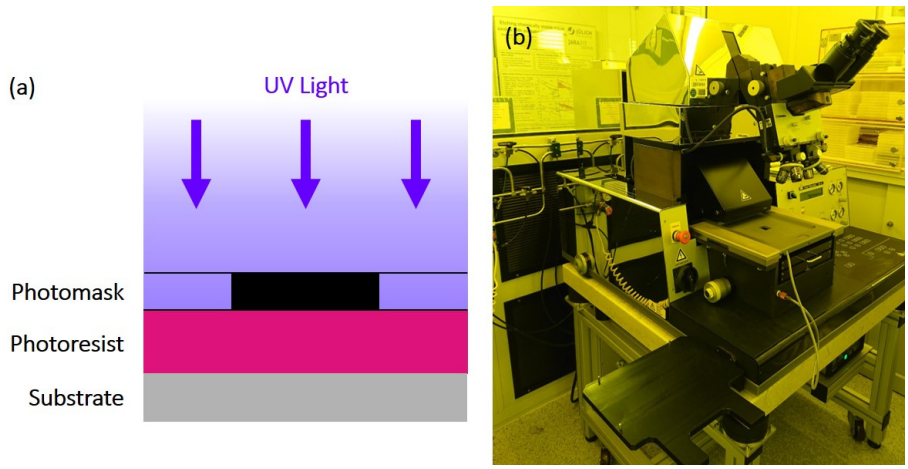

Figure 3.8: *Photolithography by using a mask aligner. (a) Schematic diagram of exposure process in contact lithography. (b) The Süss MA6 mask aligner in the clean room at PGI-7 in Research Center Jülich.*

According to the purpose of each patterning step, different photoresists are used in this work to perform photolithography by using the *Süss MA6* mask aligners. The parameters of complete photolithography process respectively for the different photoresists are listed in Table 3.4.

Table 3.4: *Process parameters of photolithography using mask aligners.*

Parameters	***UV6-0.6***	***AZ® MiR™ 701 14cps***	***AZ® nLOF™ 2020***	***AZ® 5214 E***
Type	Positive	Positive	Negative	Image reversal
Dilution with AZ® EBR Solvent	-	-	- 1 : 1	1 : 0.5
Spin coating speed	4000 rpm	4000 rpm	4000 rpm	5000 rpm
Thickness	0.5 µm	0.9 µm	2 µm 0.4 µm	0.55 µm
Soft bake	130°C, 60 sec	90°C, 60 sec	110°C, 60 sec	90°C, 120 sec
Exposure	4 mJ/cm^2	130 mJ/cm^2	40 mJ/cm^2	30 mJ/cm^2
Post exposure bake	130°C, 90 sec	115°C, 60 sec	110°C, 60 sec	115°C, 120 sec
Flood exposure	-	-	-	> 200 mJ/cm^2
Developer	MF-24A	AZ® 726 MIF	AZ® 326 MIF	AZ® 326 MIF
Developing time	45 sec	60 sec	40 sec	60 sec

3.1.3.1.2 Maskless aligner (MLA)

An MLA is a UV light exposure system for any defined pattern without the need of photomask. Unlike a conventional mask aligner based on projection of the pattern on a photomask, the MLA utilizes a direct writing approach, where light-emitting diodes (LEDs) or laser diodes are employed as the light source. The emitted UV light is manipulated by software control to scan over the photoresist layer coated on a substrate and specifically expose the area of predefined pattern. To complete an exposure step for a certain pattern, writing apparently requires much more time than projection. In order to boost the writing speed, a raster scan writing strategy is adopted for the maskless aligner employed in this work, which is the *MLA 100* from Heidelberg Instruments for entry-level research and development purposes. Figure 3.9 illustrates the working principle of the *MLA 100*, which has a 10-W LED light source with the wavelength of 365 nm. The desired pattern layout is converted into a pixel image for exposure. A spatial light modulator (SLM), more specifically a digital micromirror device (DMD), serves the role of a programmable photomask that modulates the UV light to generate each fraction of the converted pattern accordingly. Through a fixed optic system, the light pattern is shone on the corresponding area of the photoresist-coated substrate. Once the exposure dose is reached, the SLM changes the light pattern, while the wafer stage advances to enable the exposure on the next corresponding area. Owing to the raster scan writing strategy, the minimum feature size is limited to 1 µm as per the pixel size of the SLM. A top-view camera is used to identify marks on the substrate for alignment. In addition to the recognition of substrate location, the pixel image conversion from the pattern layout is corrected based on the detected rotation. Technically, the alignment with a deviation not larger than 1 µm is feasible. Owing to its efficiency and flexibility on layout iteration, the MLA is a

useful photolithography exposure tool especially for low- to mid-volume research and development activities.

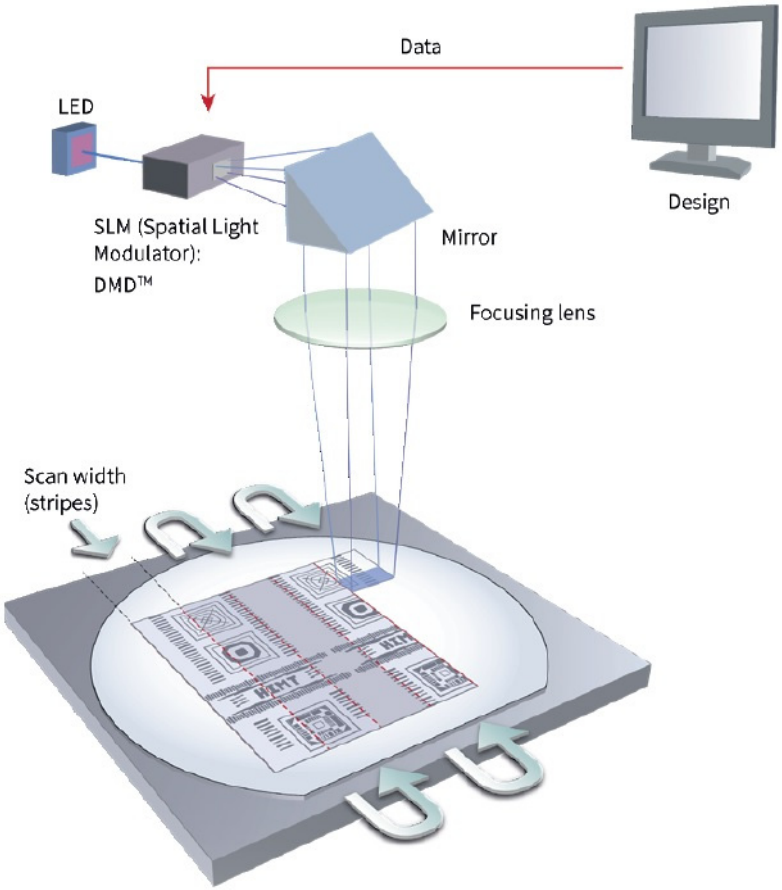

Figure 3.9: *Schematic diagram of the maskless aligner MLA 100 performing exposure by a UV LED light source [73].*

The parameter details of complete photolithography using the *MLA 100* are listed in Table 3.5, including a positive and a negative photoresists sensitive to the wavelength of 365 nm.

Table 3.5: *Process parameters of photolithography using the MLA 100.*

Parameters	***AZ® MiR™ 701 14cps***	***AZ® nLOF™ 2020***
Type	Positive	Negative
Spin coating speed	4000 rpm	4000 rpm
Thickness	0.9 µm	2 µm
Soft bake	90°C, 60 sec	110°C, 60 sec
Exposure	850 mJ/cm^2 *(Non-reflective substrate)* 560 mJ/cm^2 *(Reflective substrate)*	450 mJ/cm^2 *(Non-reflective substrate)*
Post exposure bake	115°C, 60 sec	110°C, 60 sec
Developer	AZ® 726 MIF	AZ® 326 MIF
Developing time	90 sec	40 sec

3.1.3.2 E-beam lithography (EBL)

The photolithography based on the UV light with the wavelength of 365 nm has a technical constraint of feature size around 1 µm. As suggested from the Equation 3.1, decreasing the wavelength is one of the most crucial factor to achieve a smaller critical dimension. However, advanced exposure systems using DUV or EUV light are aimed to the market of high-volume production. The access to those systems is extremely limited for research and development activities. Alternatively, e-beam lithography (EBL) has been developed with electrons as the exposure source to achieve nanoscale patterning, because electrons can exhibit wave-like behavior with a very short wavelength when they are accelerated by a very large electric voltage.

As a part of wave-particle duality, matter waves, also known as de Broglie waves, describe the wave-like behavior of all matter, including electrons. The wavelength of a de Broglie wave, λ, is associated with a particle of momentum, p, through the Planck constant, h:

$$\lambda = \frac{h}{p} = \frac{h}{mv}, \tag{3.3}$$

where the momentum, p, is further expressed by the product of the mass, m, and the velocity, v, of the particle.

An electron, a negative charge carrier, can be accelerated by the electric field induced by an applied voltage. The kinetic energy that the electron obtains after the acceleration (E_{acc}) is determined by the electric potential difference during the acceleration, which is equal to the applied accelerating voltage, V_{acc}. Considering a stationary electron, its kinetic energy, E_k, is equal to the E_{acc} after acceleration:

$$E_k = E_{acc} \tag{3.4}$$

$$\Rightarrow \frac{1}{2} m_e v^2 = eV_{acc} \tag{3.5}$$

$$\Rightarrow v = \sqrt{\frac{2eV_{acc}}{m_e}}, \tag{3.6}$$

where m_e is the electron mass, and e is the electron charge. Combining the Equation 3.3 and

Equation 3.6, the de Broglie wavelength of an accelerated electron can be expressed by

$$\lambda = \frac{h}{\sqrt{2m_e eV_{acc}}}, \tag{3.7}$$

suggesting a higher V_{acc} can lead to a shorter λ. Therefore, EBL can realize nanoscale patterning by applying an adequate V_{acc} ranging from few tens of kV to 100 kV.

In the EBL, the exposure is performed in a direct writing approach without the need for masks by using a designated system, known as the e-beam writer. However, it has a major limitation on its low throughput, which makes it only be widely used for low-volume research and development activities. The *EBPG 5000plus* from Vistec Electron Beam and the *EBPG 5200* from Raith GmbH were used in this work as e-beam writer systems for EBL. As they belong to the family *of EBPG* systems, they have similar configurations. Figure 3.10 shows a typical structure of such systems, from which one can find a significant similarity to the structure of a scanning electron microscope (SEM) displayed in Figure 3.13. To generate an e-beam, an electron gun is employed based on thermal field emission, which combines thermal and electric field effects, from a sharp tungsten tip coated with ZrO_2. The emitted electrons are accelerated by an accelerating voltage of 100 kV, forming an e-beam that is then focused, corrected, and deflected by electron optics to perform the desired exposure on the specimen. Among the different electron optics, a beam blanker is used to temporarily stop the e-beam exposure on the specimen without switching off the electron gun, when the focused spot of the e-beam is moving across the area that is not supposed to be exposed. There are two sets of blanking electrodes with a blanking aperture in between. When the e-beam is bent off the axis by the electric field generated from both sets of the blanking electrodes, it cannot pass the blanking aperture and reach the specimen. As a result, a temporary switch-off of the e-beam is realized until the electric field generated from the blanking electrodes is removed. In addition to several electromagnetic lens deployed to focus the e-beam in different stages, a stigmator, essentially a set of coils generating magnetic fields, is used to shift the e-beam with astigmatism back to its optimum shape. At last, electron deflectors above the objective lens apply corresponding electric fields to shift the e-beam in x and y directions such that the focused spot is positioned desirably on the specimen, in order to perform vector scans, where the e-beam is scanned over only the area to be exposed.

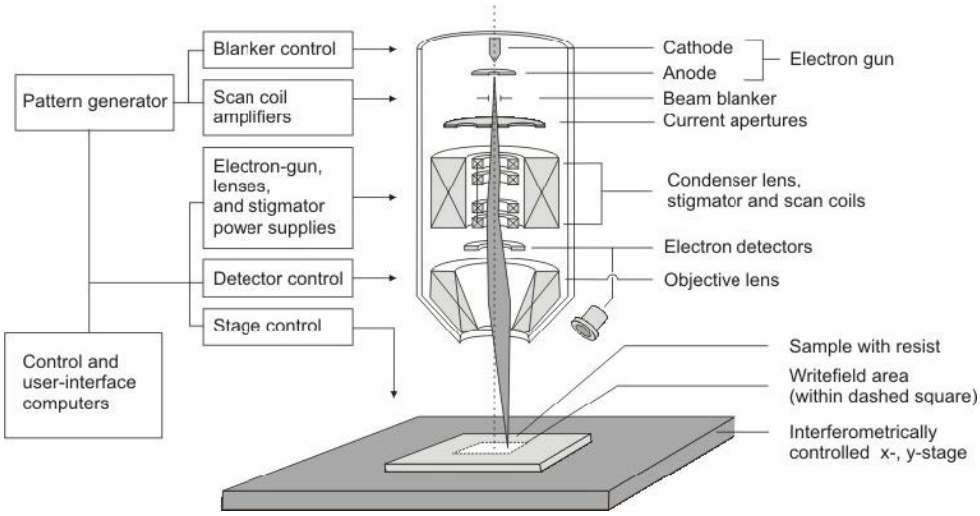

Figure 3.10: *Schematic diagram of e-beam writer for EBL [74].*

The EBL process parameters used in this work are listed in Table 3.6, including a positive and a negative e-beam resists. In line with the direct writing approach, the positive resist is mainly used in patterning by lift-off, while the negative resist is used in patterning by etching.

Table 3.6: *Process parameters of EBL.*

Parameters	***CSAR AR-P 6200.04***	***AZ® nLOF™ 2020***
Type	Positive	Negative
Dilution with AZ® EBR Solvent	-	1:2
Spin coating speed	4000 rpm	4000 rpm
Thickness	100 nm	250 nm
Soft bake	150°C, 1 min	90°C, 3 min
Exposure	650-800 μC/cm^2	55 μC/cm^2
Post exposure bake	-	110°C, 3 min
Developer	AR 600-55	AZ® 726 MIF
Developing time	1 min	3 min

3.2 Analytical methods

In the process of device fabrication and chip manufacturing, there is a variety of analytical methods that can be used for process control according to the purpose of inspection. Based on the analysis results, one can confirm the quality and yield of each process step, and more importantly identify material and structure defects that cause device malfunctions. Moreover, this enables further studies to pinpoint the major process issues. In this subchapter, the working principle of different analytical methods mainly used in this work will be explained.

3.2.1 Electron microscopy

Microscopy is the science of investigating small objects and structures using a microscope. Optical microscope (OM) utilizes photons to detect objects through optics. Like photolithography, OM is a diffraction-limited systems, whose resolution limit can be derived from Abbe's formula or Rayleigh's criterion, as expressed by the Equation 3.1. Hence, wavelength reduction is one of the greatest interests in the microscope development towards better resolutions. In this regard, electrons have been used to replace photons in OM based on the wave-particle duality of electrons, which opens a new field of electron microscopy and later stimulates the invention of EBL. To date, electron microscopes have the resolving capability from few nm down to 0.5 Å [75-79], enabling nanoscale studies on physical structures and atomic arrangements.

3.2.1.1 Transmission electron microscopy (TEM)

Based on the development of electron optics laid by the groundwork from the 1880s to the 1920s, the German physicist, Ernst Ruska, working with the electrical engineer, Max Knoll, developed the first prototype of electron microscope in 1931, which formulates the concept of

transmission electron microscope (TEM) [80]. Because of his groundbreaking contribution to electron microscopes, Ruska was awarded the Nobel Prize in physics in 1986.

A TEM works on similar principles to an OM but uses electrons in the place of photons and electromagnetic lenses in the place of glass lenses [81, 82]. It utilizes a high-voltage e-beam to illuminate a specimen and creates a visible image translated from the electron signal. The general structure of TEM is sketched in Figure 3.11. The illumination system at the top enables a focused e-beam illuminating the sample. The e-beam is generated by an electron gun, where emission and acceleration of electrons take place. The electrons are emitted from a sharp tip typically made of tungsten by either thermionic or field emission. They are subsequently accelerated by electric field under a very high voltage, normally from 40 kV to 400 kV. By using electromagnetic lenses and electron apertures, the e-beam is corrected and focused prior to the emergence of the sample, which is also called the lamella. Normally, this kind of sample is prepared in a dedicated approach by using focused ion beam to have the thickness below 100 nm. As the high-energy electrons are transmitted through the sample, they carry the structural information and travel through another set of electromagnetic lenses and electron apertures to form an image with correct focusing and magnification. On the one hand, the image can be projected on a fluorescent viewing screen coated with a phosphor or other scintillator material such as zinc sulfide for enabling direct observation of the image. To obtain the digital format of the image, on the other hand, the phosphor or scintillator can be coupled by means of an optical system to a digital camera. Alternatively, an electron detector can be used directly to convert the electron signals to a digital image.

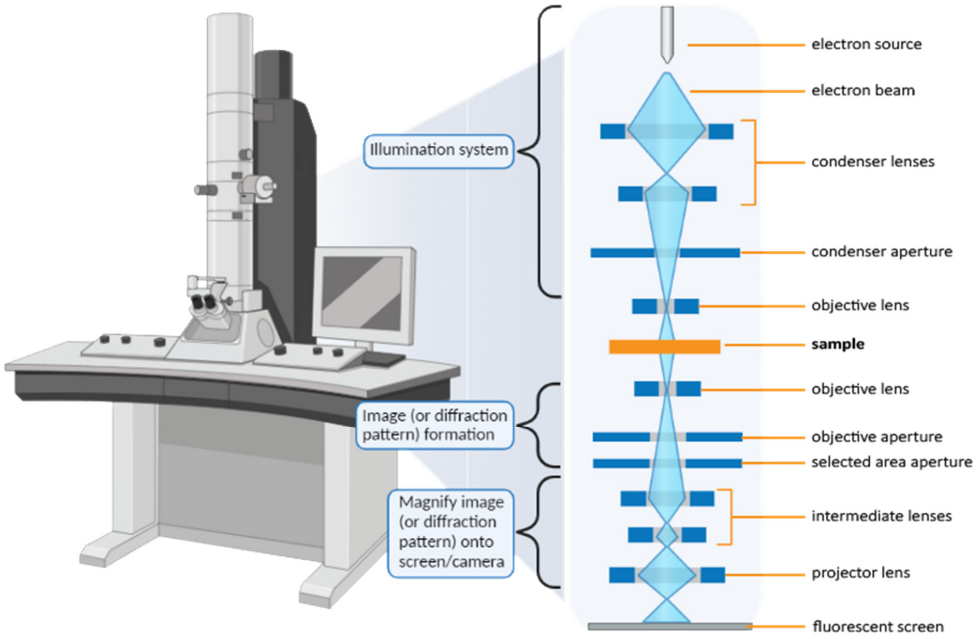

Figure 3.11: *Structure schematic of TEM [83].*

According to interactions of the e-beam and the sample, the contrast in a TEM image is constituted of mass-thickness contrast, diffraction contrast, and phase contrast. The mass-thickness contrast is literally attributed to the mass and thickness of the sample. In the area of higher mass or larger thickness, the incident electrons are scattered more, resulting in less

transmitted electrons that can be captured by the aperture. Thus, the area appears to be darker in the image due to the weaker electron intensity on the image plane. Besides, diffraction of electron can also influence the contrast. In the area with periodic crystal structure, diffraction of electron occurs according to Bragg's law, where the de Broglie wave of electron is considered. In bright-field imaging, which is based on the primary beam, the area where the diffraction takes place is darker if the diffraction angle is too large for the aperture to capture the diffracted electrons. In dark-field imaging, for which a diffracted beam is selected, the area where the corresponding diffraction takes place appears to be brighter, because the aperture only allows the electrons of a certain scattered angle to pass through. Lastly, the phase contrast comes from the phase shift of the de Broglie wave of electron resulted from the transmission of the sample. When the objective aperture is removed or a very large objective aperture is used, the most scattered electrons available can contribute to the imaging. The resulting image is an interference pattern that contains a lot of structural information encoded in the phase contrast. In this regard, the use of phase contrast can enable high-resolution transmission electron microscopy (HRTEM) with resolutions less than 1 Å. However, the interpretation of phase contrast in the TEM image is very complex, because multiple factors that can lead to phase changes need to be taken into account, e.g. sample thickness, lattice orientation, scattering factor, focus, and spherical aberration. Hence, a proper interpretation of an HRTEM image on an atomic scale requires simulations that is based on an adequate amount of given information regarding the sample and the TEM condition.

3.2.1.2 Scanning electron microscopy (SEM)

A few years after Ernst Ruska, who is in a group led by Max Knoll, developed the first prototype of TEM in 1931, Knoll first obtained scanned electron images in 1935 by the use of an electron beam scanner [84, 85]. In the late 1930s, Manfred von Ardenne investigated the possibility of using a scanned electron probe to avoid the chromatic aberration of objective lens in TEM with thick samples [86, 87]. In this regard, he established some fundamental principles of SEM including the formation and deflection of the electron probe, the positioning of electron detector, and the approach to amplify the very small signal current. However, he was unable to realize them in a complete system due to his limited time on the development. Meanwhile in America, Vladimir K. Zworykin and his team at the RCA Laboratories built an SEM with several important original features, but the performance could not show a promising interest compared to the production of commercial TEM [88, 89]. Charles Oatley at the Cambridge University revived the idea of SEM in 1948, and his research group had been dedicating to the succeeding SEM development over the next 15 years, which culminated in the production of the first commercial SEM later by the Cambridge Instrument Company in 1965 [90, 91].

SEM is nowadays an analytical technique massively used to study morphology and topography of solid-state materials and biological specimens with a resolution of around 1 nm. Despite a lower resolution in comparison to TEM, it has advantages of simple specimen preparation, capability of acquiring 3D information, and low cost. The basic principle of SEM is to scan over the specimen surface by a high-voltage e-beam with the acceleration voltage typically from 1 kV to 30 kV and simultaneously detect the electrons interacted with the specimen [82, 92, 93]. The detected current signal is synchronized to the sampled position on the specimen surface to generate each pixel of an SEM image.

Various results of high-energy electrons interacting with a specimen are illustrated in Figure 3.12 with the corresponding interaction volumes in the specimen. Apart from the transmitted, diffracted, and scattered electrons exiting from the bottom of the specimen, which are used for the imaging in a TEM, photons and other electrons exiting from the specimen surface also carry different valuable information that can be collected for material analysis. Secondary electrons

are the most common signal collected in an SEM for high-resolution topographical images because they are generated close to the specimen surface. The typical probing depth of secondary electrons ranges from 5 nm to 50 nm, depending on the kinetic energy of incident electrons [82]. Besides the probing depth, the interaction volume of secondary electrons is approximately the size of the e-beam diameter, which ensures an outstanding lateral resolution. Emission of a secondary electron results from inelastic scattering of an incident electron that involves an energy transfer from the incident electron to an atom in the specimen, which consequently excites an electron in the atom. When the electron is emitted with energy less than 50 eV, it is regarded as a secondary electron. To detect such low-energy electrons, an Everhart-Thornley (E-T) detector is utilized predominantly in an SEM. The E-T detector is composed of a scintillator inside a Faraday cage, where a low positive bias is applied to attract the low-energy electrons and guide them to the scintillator. When the scintillator converts the electrons into photons, a light guide then transmits the weak photon signals to a photomultiplier tube, in which the received signals are amplified and converted to the final voltage signal used to generate a digital image.

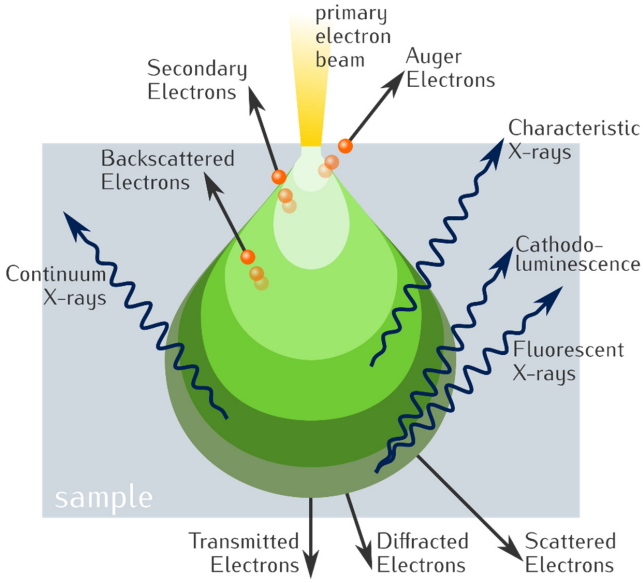

Figure 3.12: *Electron-matter interactions. High-energy electrons from an e-beam interact with a sample and generate various electron and photon signals, which correspond to different interaction volumes [94].*

Sharing a common task of scanning a focused high-voltage e-beam over the specimen surface, an SEM and an e-beam writer for EBL have similar structures. The structure diagram of a typical SEM is displayed in Figure 3.13, while it can be compared to the structure of an e-beam writer sketched in Figure 3.10. Similarly, both are built based on a vacuum chamber where an electron gun at the top generates an e-beam by either thermal or field emission and high-voltage acceleration. Through different electron optics, the e-beam is aligned and focused with stigmatism correction before deflected to scan over the specimen surface. Secondary electron detectors, commonly the E-T detectors, are deployed close to the specimen surface for imaging. Normally, an e-beam writer is also equipped with secondary electron detectors to enable real-time imaging when aligning the substrate to be exposed.

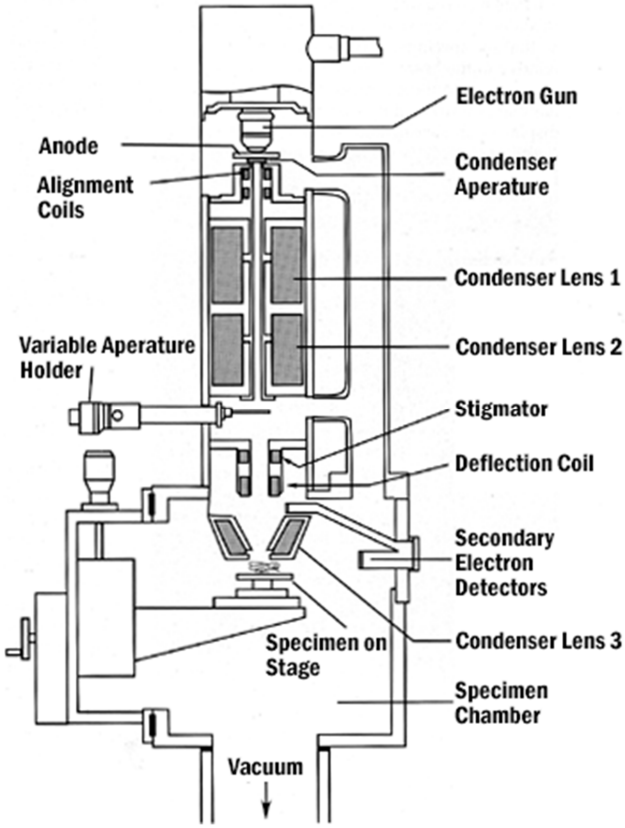

Figure 3.13: *Basic structure schematic of SEM [95].*

3.2.1.3 Energy-dispersive X-ray spectroscopy (EDX)

High-energy electron interactions with the specimen in electron microscopes includes photon emissions, as mentioned previously and shown in Figure 3.12. Among them, characteristic X-rays emitted from the specimen carry the information of chemical composition that is widely used for elemental analysis [81-83, 92]. When an incident high-energy electron excites an electron in the inner shell of an atom, the resulting vacancy in the inner shell can be filled by an electron in one of the outer shells. In such process, the energy difference of the two discrete energy levels can be released in the form of a photon in the X-ray range. The process of characteristic X-ray emission is illustrated in Figure 3.14. Due to the fact that discrete energy levels of atomic orbitals are distinct element characteristics, these X-rays appear to be the elemental fingerprint, which is the perfect information for elemental analysis.

As the e-beam used in an electron microscope, e.g. an SEM or a TEM, can simultaneously serves as the excitation source for characteristic X-ray emission, an energy-dispersive X-ray spectroscopy (EDX or EDS) is commonly installed in an electron microscope system to measure the emitted X-ray intensities of different energies for in-situ elemental analysis. Combining imaging of electron microscope and local elemental analysis of EDX, presence of one or multiple elements can be mapped out on an electron microscope image, which is a powerful analytical method commonly available in electron microscope systems, i.e., EDX mapping. Examples of an EDX installed in a TEM and an SEM are shown in Figure 3.15a and

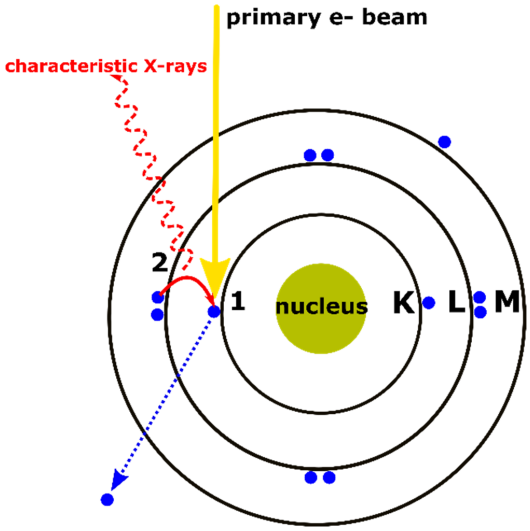

Figure 3.14: *Process of characteristic X-ray emission [96].*

Figure 3.15b respectively. The Si(Li) detector displayed in Figure 3.15a is a conventional solid-state X-ray detector based on Li-doped Si. Through the photoelectric effect, an X-ray photon is able to ionize a Si atom in the lattice and consequently generate electron-hole pairs, whose number is proportional to the energy of the X-ray photon. When the both charge carriers are drawn to the detector contacts by an applied electric potential, a voltage pulse corresponding to the X-ray energy is recorded. Eventually, the collection of these voltage pulse signals can form an energy histogram, which is the energy-dispersive X-ray spectrum. The Si(Li) detector is operated at a critically low temperature by liquid nitrogen cooling in order to reduce a dark current caused by thermally generated electron-hole pairs and prevent Li atoms from diffusing.

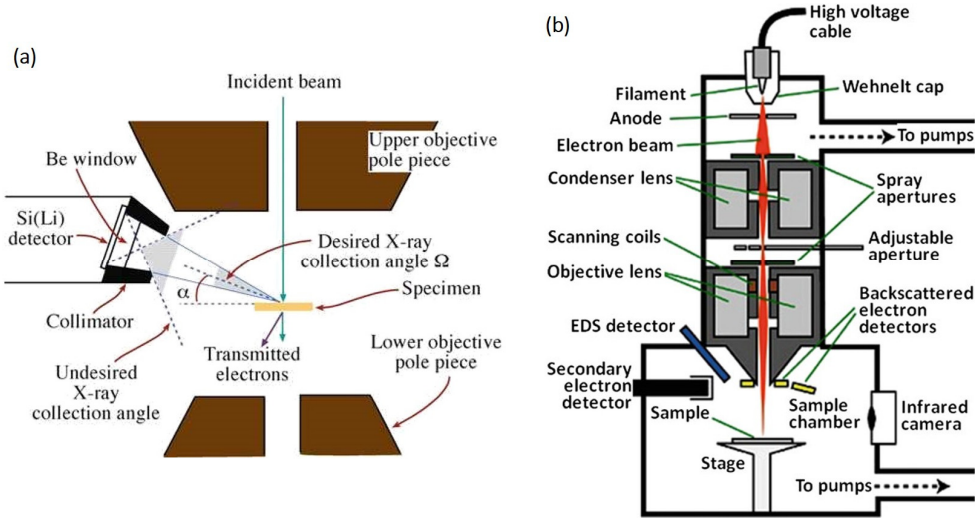

Figure 3.15: *Schematic diagram of EDX installed in electron microscopes. (a) EDX installed in TEM [81]. © 2009 by Springer. (b) EDX installed in SEM [97]. © 2017 by Springer.*

Besides, an X-ray transparent window normally made of Be isolates the Si(Li) detector from the electron microscope column to avoid condensation or contamination on the cooled detecting surface. Nevertheless, soft X-rays of energies less than 1 keV are strongly absorbed by the Be window, disabling the analysis for elements having atomic number less than 11 [82].

3.2.2 Atomic force microscopy (AFM)

Starting from the invention of scanning tunneling microscope (STM) in 1981 by Gerd Binnig and Heinrich Rohrer, who were consequently awarded the Nobel Prize in physics in 1986, the family of scanning probe microscopes (SPMs) has been developed for nanoscale surface analysis in various aspects [98]. In order to overcome the limitation that STMs can only measure conducting or semiconducting surfaces, an atomic force microscope (AFM) was first developed by Gerd Binnig, Heinrich Rohrer, and Calvin Quate in 1985 as an SPM that can image the topology of any type of surface on the nanometer scale [99].

The working principle of AFM is illustrated in Figure 3.16a. A flexible cantilever normally made of silicon or silicon nitride with a sharp tip is employed in an AFM as the scanning probe. While the sharp tip is raster-scanned over the specimen surface, it is affected by the interatomic force such as van der Waals force between the surface atoms of the tip and the specimen, which has a distance dependence described in Figure 3.16b. When the tip is brought closer to the specimen surface, it starts to sense an attraction from the specimen that has a maximum before increasing repulsion comes into place. During the raster scan, bending of the cantilever according to the specimen topography is recorded by a photodetector that tracks a laser beam reflected on the backside of the cantilever, which finally leads to generation of a digital image containing 3D topographical information [100-103].

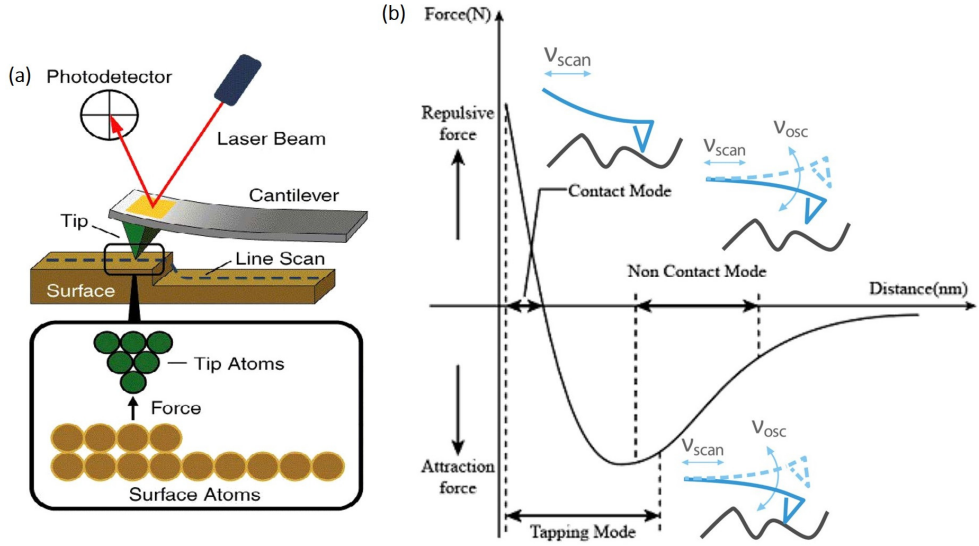

Figure 3.16: *Working principle of AFM. (a) Schematic diagram of AFM operation [103]. (b) Interatomic force as a function of distance. In addition to the regimes distinguished by attraction and repulsion, there are different regimes of interatomic force used in different operation methods in AFM, namely the contact, tapping, and non-contact modes. Adapted from [101].*

The original operation mode of AFM is the contact mode, where the tip is in contact with and therefore repelled by the specimen. Technically, there are constant height and constant force approaches. For the constant height approach, the absolute height of the cantilever is maintained, and the bending of the cantilever is recorded and directly used to generate images. The constant force approach, on the other hand, utilizes a feedback loop to maintain the cantilever at a specified bending condition. The compensating height adjustment of the cantilever is recorded to generate images. To avoid hard contact with the specimen and reduce fidelity loss for soft specimens, the tapping mode is commonly used. Driven by a piezoelectric actuator, the cantilever is oscillated at or near its fundamental resonance frequency. The tip is positioned to slightly touch the specimen surface at the lower end of the cantilever oscillation, whose amplitude is damped according to the distance between the tip and the specimen surface. A feedback loop is then utilized to maintain a specified oscillation amplitude and hence a constant interaction force. The compensating height adjustment of the cantilever is recorded for the use of image generation. Furthermore, the non-contact mode is operated also with cantilever oscillation, but without contact on the specimen surface. It offers the lowest possible tip damage caused by the interaction with the specimen surface, which preserves the tip sharpness and achieve high resolution. However, it is challenging to keep the tip in the attractive regime of interatomic force. A high performance feedback control is needed for the case.

In this work, AFM was mainly used as an inspection method for the process control when the lithography resist covers most of the area on the sample surface. The polymer-based resists are not electric conducting, thereby causing a strong charging effect in the imaging of electron microscopes. AFM can not only offer a nanoscale image to measure the overlay of lithography result, but also determine height differences of certain features on the sample surface by calculating from the recorded topographical data. Practically in this work, it is useful to measure resist thickness, deposition thickness, and etching depth.

3.3 Electrical characterization methods

To conduct electrical characterization of the fabricated electronic devices in this work, including MOSFETs and RRAMs, mainly two types of testing instruments were employed for electrical parameter analysis. Despite the accessory variations in specification, they are basically products from the series of *Keithley 2600* and *Keithley 4200-SCS*.

Designed for basic electrical testing, the *Keithley 2600* is a source measure unit (SMU) or also called sourcemeter, which is a compact instrument capable of precisely generating and simultaneously measuring electric voltage and current. Figure 3.17a shows an example of a two-probe testing setups based on Keithley 2600, which is situated in PGI-7 at Research Center Jülich. Through a homemade LabVIEW interface, one can command the two-probe testing setup to perform a quasi-static I-V measurement on an RRAM device that the current is measured in one or multiple DC voltage round sweeps. Practically, it can be used to observe SET and RESET processes of resistive switching, and test endurance by repeating the switching cycles.

On the other hand, *Keithley 4200-SCS* is an advanced and more comprehensive parameter analyzer that can be configured with multiple SMUs and ultra-fast pulse measure units (PMUs) [104]. Even though several measure units can be installed in the *Keithley 4200-SCS*, only four of them are allowed be operated at the same time due to the instrument specification, indicating they need to be fit into a four-probe station. In order to arrange all installed measure units to connect to four testing probes, a remote preamplifier/switch module (RPM) can be employed as a multiplexer to manage automatic switching between the use of an SMU and a PMU. With

four probes available, the *Keithley 4200-SCS* can not only measure RRAM devices, which have two terminals, but also characterize three-terminal test structures, including MOSFETs and 1T-1R unit cells, even with an additional bulk terminal of transistor. More importantly, it offers versatile possibilities to test simple circuits constituted of multiple devices as long as they only require four probing terminals or less. For examples, up to three devices in an RRAM array can be tested simultaneously, sequentially, or in any specific way to study circuit dynamics, crosstalk between devices, and CIM applications. Figure 3.17b shows an example of a four-probe testing setup based on Keithley 4200-SCS, which is situated in the Helmholtz Nano Facility (HNF) at Research Center Jülich. Through the Keithley external control interface (KXCI) with GPIB connection, the *Keithley 4200-SCS* can also perform programmable pulse measurements. In this work, the program control was developed in the programming language Lua. Any waveform can be defined individually for four different PMUs to execute. Additionally, the resistance calculated from the measured current and the applied voltage can be utilized to verify the resistance state of RRAM, which allows the program to make the decision on the next measurement step. By extending the idea further, the resistance state of RRAM can be programmed desirably to a specified range for a specific operation, which is a technique frequently used to control the initial condition of an experiment study.

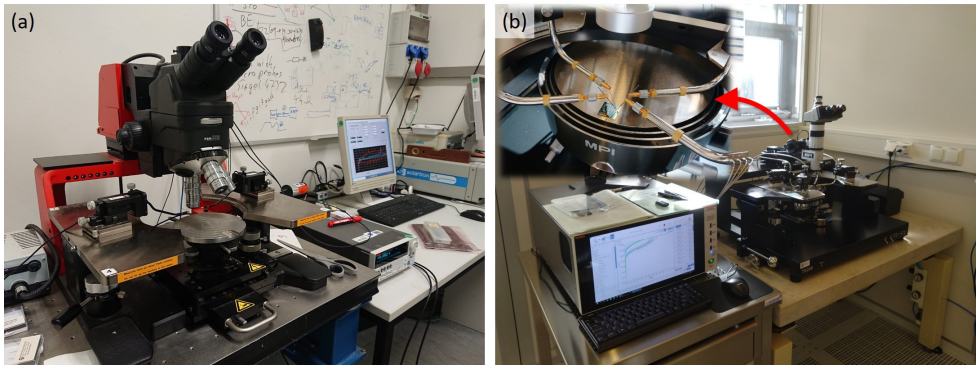

Figure 3.17: *Testing setups for electrical characterization. (a) A testing setup based on Keithley 2600 in PGI-7. With two testing probes available, it is capable of characterizing devices and test structures with only two contact terminals. A homemade LabVIEW interface enables an intuitive program control on characterizing memristors by quasi-static measurements. (b) A testing setup based on Keithley 4200-SCS in HNF. With four testing probes available, it is capable of characterizing devices and test structures with up to four contact terminals. Besides using the internal application program, Clarius, to conduct basic customizable measurements, it is possible to command this type of testing setup by Lua scripts via the KXCI for programmable measurements, which is available in PGI-7.*

Chapter 4
Active TaO_x RRAM arrays on silicon-on-insulator substrates

Given the mature and advanced state of CMOS technology, MOSFET is a suitable choice for the use in an active RRAM array as the selecting device. Scientists and engineers have dedicated significant efforts over decades to inventing and advancing CMOS technology, which provides a strong foundation for electrical performance tuning in order to achieve a good match between the MOSFET and RRAM. In this context, a study was initiated that involves designing, fabricating, and testing 1T-1R and 1T-nR arrays on an SOI substrate. The integration technology was developed at HNF through a close collaboration between two divisions of Peter Grünberg Institute (PGI), PGI-7 *Electronic Materials* and PGI-9 *Semiconductor Nanoelectronics*, in Research Center Jülich. The PGI-9 has provided technical knowledge of HKMG transistor fabrication, while the PGI-7 has contributed for TaO_x VCM-type RRAM.

4.1 Material selection of TaO_x RRAM's bottom electrode for co-integration

In order to realize 1T-1R and 1T-nR arrays, a critical building block is the integration of an RRAM and an MOSFET devices, which forms a 1T-1R standard unit cell depicted in Figure 4.1a. The bottom electrode of the RRAM cell connects in series to the drain terminal of transistor. A cross section of the physical structure is illustrated in Figure 4.1b. A typical planar-type transistor is shown with gate at the center and source and drain (S/D) symmetrically by the sides. To obtain ohmic contacts at the S/D, a pronounced CMOS process technique of silicidation is able to selectively form conductive NiSi, covering the S/D [105-108]. The RRAM cell is then fabricated directly on the NiSi drain electrode. For the TaO_x VCM-type RRAM used in this work, a Ta layer serves as an ohmic electrode on top of the TaO_x. Together with a Pt capping layer covered on the top, they form the so-called top electrode (TE). At the bottom side of the RRAM cell, Pt bottom electrode (BE) forms a schottky contact. Potentially, the NiSi drain electrode can replace the Pt BE and keep the schottky contact in place underneath the TaO_x thin film. This will reduce the process complexity and the potential yield loss of RRAMs integrated. In this regard, an experiment was performed to compare resistive switching properties of the TaO_x RRAM devices with Pt and NiSi bottom electrodes, while keeping the rest in the layer stack identical for both types of devices.

The crossbar device structure was utilized due to its simplicity. As depicted in Figure 4.2, the device design consists of two patterns intersecting at a certain overlapped area, where a complete MIM stack is formed. Due to the fact that the functional RRAM cell locates at the overlapped area or crossing point, devices of this structure are called crosspoint or crossbar devices. With electrode effective width designed to be 5 µm, the device area was 5×5 μm^2 in

this study. For the RRAM fabrication, BE was patterned as the first step, while the switching layer and TE were patterned simultaneously in the second step.

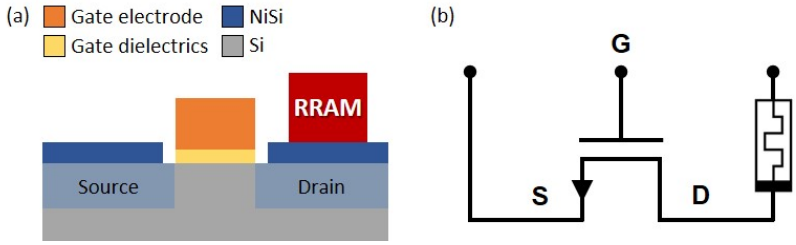

Figure 4.1: *Schematic diagrams of 1T-1R standard unit cell. (a) Cross-sectional schematic of a 1T-1R integration example as a unit cell. (b) Circuit schematic of a 1T-1R standard unit cell.*

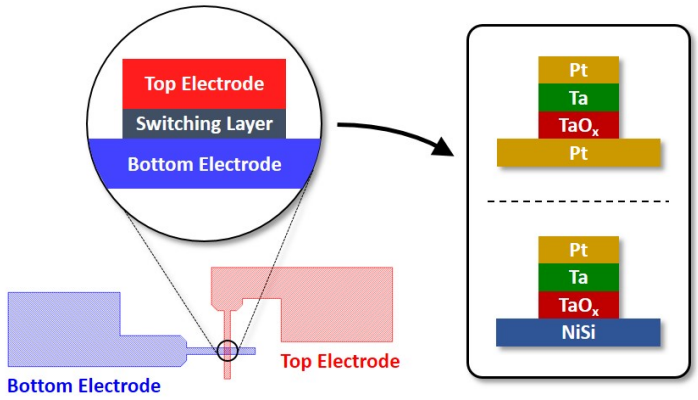

Figure 4.2: *Crossbar device structure of RRAM and the device stacks of TaO_x RRAMs to be studied for the BE material difference.*

To determine the effect of the BE material, Pt and NiSi electrodes were prepared separately. For the Pt BE, 30-nm-thick Pt on 5-nm-thick Ti was sputtered on a SiO_2 substrate, followed by photolithography patterning. The BE pattern was transferred into the sputtered layers using reactive ion beam etching (RIBE) with pure Ar gas. On the other hand, for the NiSi BE, the lithography step was performed on an SOI substrate having 60-nm-thick Si on top of 145-nm-thick SiO_2. The topmost Si was over-etched into SiO_2 using RIBE with pure Ar gas such that the BE pattern was transferred to the Si layer. Next, a 10-nm-thick Ni was sputtered and a rapid thermal processing (RTP) step followed at 500°C for 30 sec in the forming gas mixture of N_2 and H_2. Silicidation during the RTP step enabled a selective formation of the NiSi on the Si surface. The unreacted Ni at the surface was removed by subsequent wet etching in a chemical mixture of H_2SO_4 and H_2O_2, leaving behind only the NiSi on the patterned Si.

Following the BE fabrication, the in-situ deposition of switching layer and TE was synchronized for both Pt and NiSi BE devices. This process involved a series of depositions starting from reactive sputtering of 10-nm-thick TaO_x to RF sputtering of 15-nm-thick Ta, and ending with DC sputtering of 25-nm-thick Pt. Similar to the BE, the TE was patterned by using photolithography and RIBE process.

After the fabrication, quasi-static electrical measurements on both types of devices were performed by DC voltage sweep and the I-V characteristic of 10 switching cycles is displayed

in Figure 4.3. The standard TaO_x RRAM stack with Pt BE showed a reliable filamentary-type resistive switching with slight variations. Oppositely, the device with NiSi BE showed inconsistent switching with large variabilities in $V_{SET,C}$ and $V_{RESET,C}$, where the SET and RESET processes actually take place. Additionally, R_{LRS} and R_{HRS} were found with large variabilities. Instead of a gradual RESET process, typically found in the standard device with Pt BE, the device with NiSi BE showed an abrupt RESET process. As a result of the abrupt RESET process, a large variability in the SET process is also observed. This indicates an additional series resistance in the RRAM stack that could originate from the difference in device fabrication. In summary, the electrical measurements concludes that the NiSi is not suitable for the BE of the TaO_x RRAM stack.

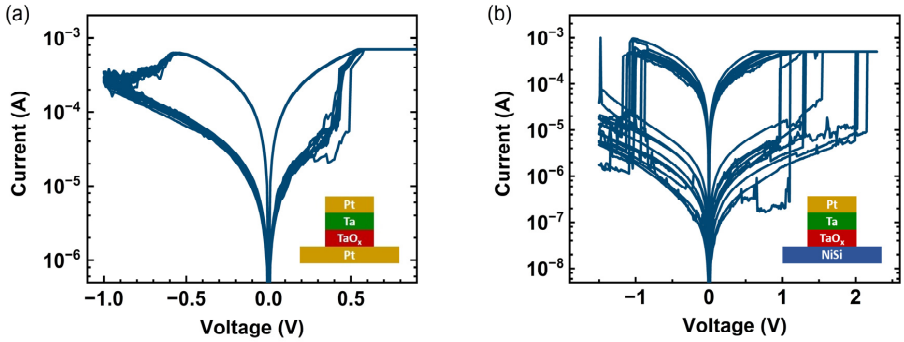

Figure 4.3: *I-V diagrams of NiSi/TaO_x/Ta/Pt RRAM device and Pt/TaO_x/Ta/Pt RRAM device for 10 bipolar resistive switching cycles when current compliance is applied in SET processes. (a) The TaO_x RRAM with Pt BE shows consistent switching. Both $V_{SET,C}$ and $V_{RESET,C}$ are below 1 V. (b) The TaO_x RRAM with NiSi BE shows inconsistent switching and abrupt RESET processes. Both $V_{SET,C}$ and $V_{RESET,C}$ have large variabilities and mostly larger than 1 V.*

To explore the underlying cause of switching issues in the NiSi/TaO_x/Ta/Pt RRAM devices, a cross-sectional TEM analysis was performed on the stack of an actual device. From the illustrative result shown in Figure 4.4, a complete cross section of the stack is clearly shown. The NiSi on the Si layer was produced by consuming Si directly from the SOI substrate through silicidation. Upon a closer examination of the interface between TaO_x and NiSi, an interfacial oxide layer was detected to be roughly 2.5 nm thick [109, 110]. Unlike inert metals such as Pt, NiSi reacts more readily with oxygen due to a lower activation energy. This interfacial oxide layer could have been resulted from the reactive sputtering of TaO_x on the NiSi BE, which occurred in the presence of an oxygen-containing gas mixture. As a result of this unintentional interfacial oxide, the NiSi/TaOx/Ta/Pt RRAM devices experience an additional series resistance that is likely within the range of their R_{LRS}. The voltage divider effect, as a consequence of the series resistance, is responsible for the abrupt RESET process along with high $V_{SET,C}$ and $V_{RESET,C}$ as well as the significant variability.

To prevent the negative impact of the interfacial oxide, an experiment was conducted with an additional Pt layer as the effective BE, inserted between the TaO_x and NiSi layers. The reactive sputtering of TaO_x could be performed on the Pt layer, which prevents the oxidation of NiSi surface. Besides, a reliable switching property could be expected as the Pt layer would work as the BE. To maintain the consistency of device fabrication process, the process remained the same until the NiSi BE was fabricated. The subsequent in-situ deposition began with 20 nm-thick Pt followed by 10-nm-thick TaO_x, 15-nm-thick Ta, and 20-nm-thick Pt. The

TE pattern was then transferred to the entire stack (Pt/TaO_x/Ta/Pt) using photolithography and RIBE process, while the NiSi layer was exposed.

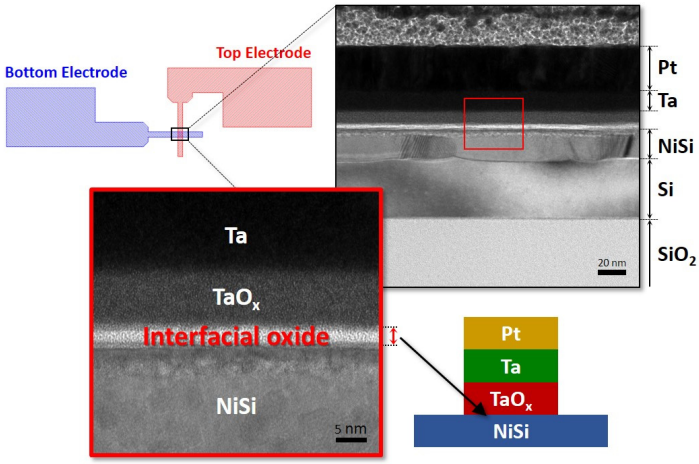

Figure 4.4: *Presence of interfacial oxide in the fabricated TaO_x RRAM device with NiSi BE. Cross-sectional TEM images show the device stack on the SOI substrate, in which an interfacial oxide layer with a thickness of roughly 2.5 nm can be recognized between the TaO_x and the NiSi layers.*

I-V characteristic of the NiSi/Pt/TaO_x/Ta/Pt device was plotted in Figure 4.5. A reliable resistive switching can be found comparable to the reference device (Pt/TaO_x/Ta/Pt), which is shown previously in Figure 4.3. For both of the devices, not only the ON/OFF ratios are larger than 7 but also the $V_{SET,C}$ are around 0.5 V when the I_{CC} of 500 µA was applied in the SET processes. The positive outcome of inserting the Pt layer between the TaO_x and NiSi layers confirms the elimination of unintentional interfacial oxide. This verifies and concludes the core concept of realizing 1T-1R integration in this work, where the entire RRAM stack (Pt/TaO_x/Ta/Pt) is integrated on the NiSi-based S/D ohmic contacts of the MOSFET device with Pt as the BE material.

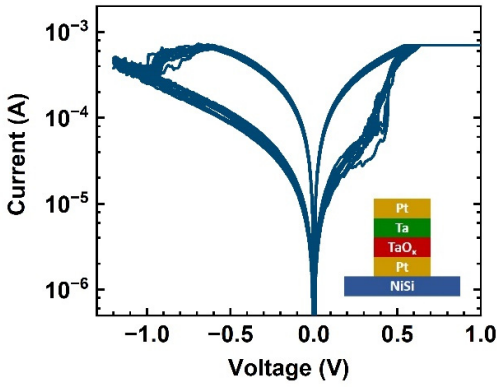

Figure 4.5: *I-V diagram of NiSi/Pt/TaO_x/Ta/Pt RRAM device for 10 bipolar resistive switching cycles.*

4.2 Layout of 1T-1R and 1T-nR arrays

Figure 4.6a displays the circuit layout of 1T-1R array, where each RRAM cell is connected in series to a transistor such that the bottom electrode of the RRAM cell is linked to the transistor's drain terminal. The gate terminals of the transistors are arranged in a column, forming a word line, while the top electrodes of the RRAMs are arranged in a row, forming a bit line. The source terminals of the transistors are connected in columns, with every two adjacent columns combined as one common source line. This common source line reduces the physical area of the array on the chip and lowers the series line resistance, when accessing the furthest cell in the array. Figure 4.6b shows the planar pattern layout of 1T-1R array, taking a 16×16 array as an example. The sixteen word lines can be accessed from the bottom, while eight of the sixteen bit lines can be accessed from the left and the other eight from the right. The eight source lines can be accessed from the top. Metal contact pads are designed based on the probing locations of a specific probe card with sixteen probes at each side. This probe card is compatible with a customized testing setup for RRAM arrays. Finally, Figure 4.6c and Figure 4.6d show the fabricated 1T-1R array of TaO_x RRAMs. The bit lines, word lines, and source lines are marked in Figure 4.6c with transistors and RRAMs illustrated, while Figure 4.6d displays the SEM view with a 45-degree rotation.

A new configuration of active RRAM arrays, called 1T-nR, has been designed in addition to the existing 1T-1R. Figure 4.7a displays the circuit layout of 1T-nR array, where multiple RRAMs have their bottom electrodes connected to a drain terminal of transistor. This enables each transistor to actively control the entire row of RRAMs. Hence, the required number of transistors for an array is equal to the number of rows, and does not depend on the number of RRAMs per row or the number of columns. Compared to the 1T-1R configuration, the 1T-nR configuration significantly reduces the required number of transistors. Owing to the fact that one transistor can control one row of RRAMs, one word line can be activated simply by controlling one transistor. The gate and source terminals of each transistor represent the word line and source line in this configuration, respectively. Figure 4.7b shows the planar pattern layout in the 1T-nR configuration, using a 16×16 1T-16R array as an example. Among the sixteen bit lines, eight can be accessed from the top, and the other eight from the bottom side. To arrange word lines and source lines, one of each are considered a group that refers to the gate and source terminals of a transistor. Eight groups are accessible from theleft side while the other eight from the right side. Metal contact pads are designed based on the probing locations of a specific probe card with sixteen probes at each side, similar to the pattern layout of the 1T-1R array. This probe card is compatible with the customized testing setup for RRAM arrays. Finally, Figure 4.7c and Figure 4.7d show the fabricated 1T-1R array of TaO_x RRAMs. The bit lines, word lines, and source lines are marked in Figure 4.7c with transistors and RRAMs illustrated, while Figure 4.7d displays the SEM view with a 45-degree rotation.

Designs of 1T-1R and 1T-nR arrays were combined on a single physical mask, featuring array sizes of 4×4, 8×8, and 16×16. The mask layout is displayed in Figure 4.8, which also includes 1T-8R, 1T-15R, and 1T-30R line arrays that have a single word line and several bit lines. Furthermore, the layout has stand-alone transistors and RRAMs for testing purposes and some test structures for process monitoring. In total, the layout comprises of eight layers, which means eight patterning steps were required to fabricate the arrays. The fabrication process details for patterning steps and other processes will be explained in the next section.

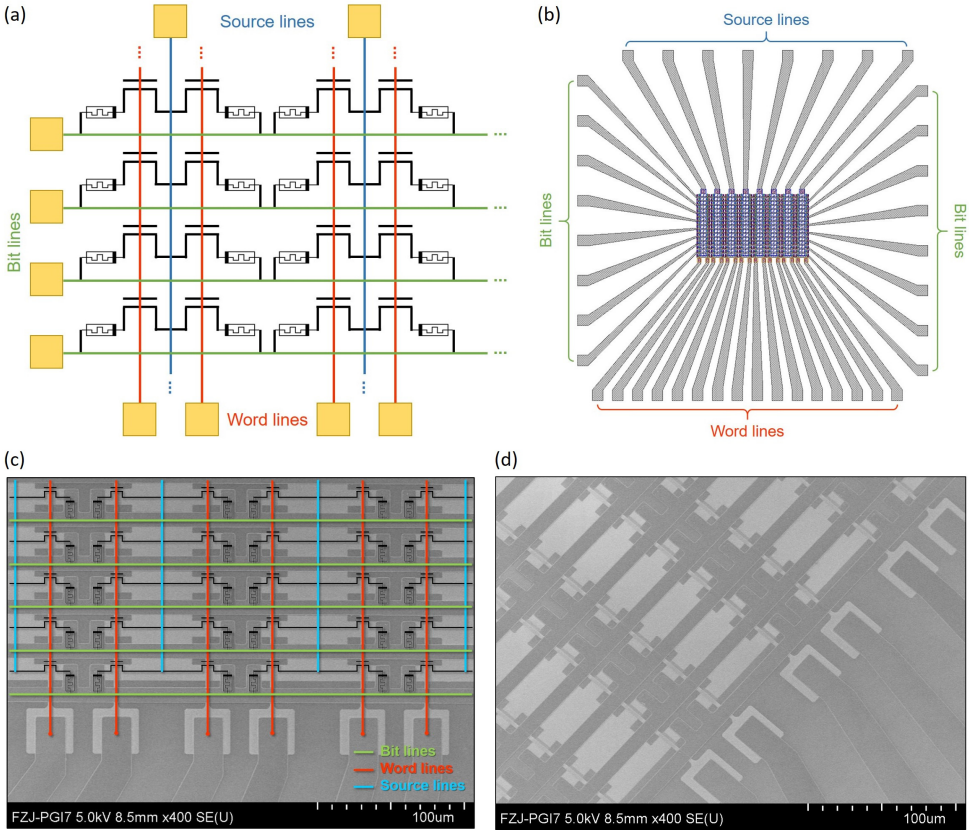

Figure 4.6: *Design of 1T-1R array. (a) Circuit layout of the 1T-1R array depicts the connection of RRAMs and MOSFETs and the arrangement of bit lines, word lines, and source lines. (b) Pattern layout of a 16×16 1T-1R array. For the bit lines, word lines, and source lines, the positions of metal contact pads are designed according to a specific probe card. (c) Top-view SEM image of the fabricated 16×16 1T-1R array with illustration of circuit schematic. (d) SEM image of the fabricated 16×16 1T-1R array with tilting and rotation.*

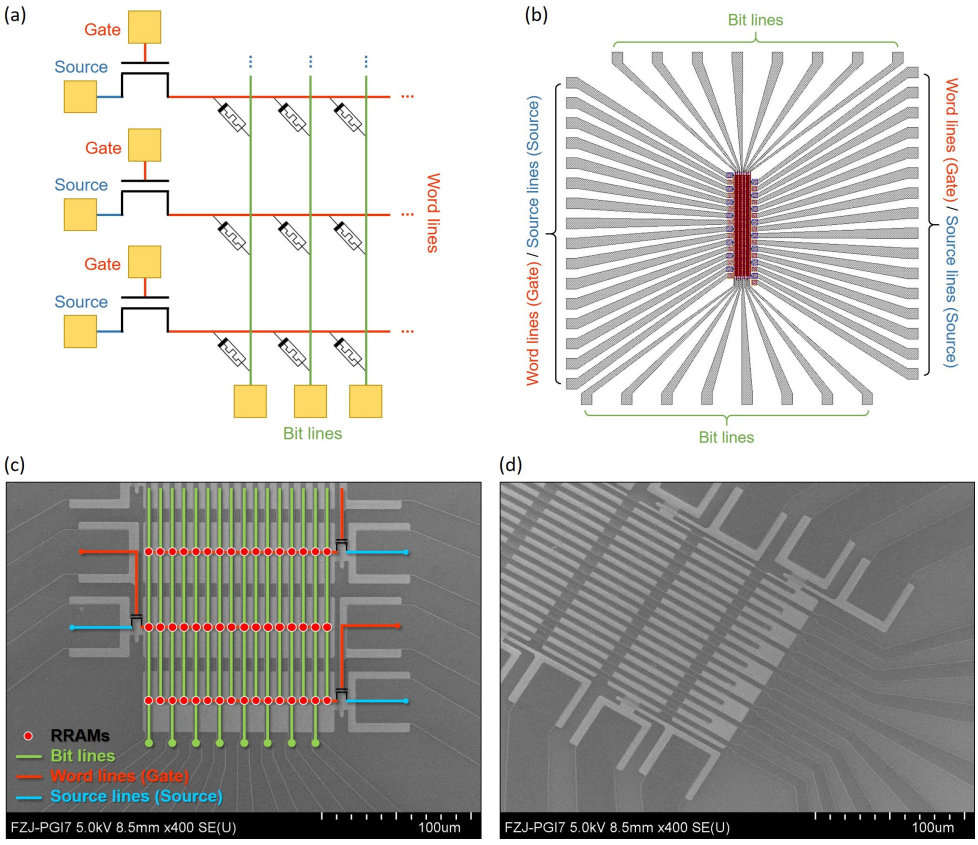

Figure 4.7: *Design of 1T-nR array. (a) Circuit layout of the 1T-nR array depicts the connection of RRAMs and MOSFETs and the arrangement of bit lines, word lines, and source lines. (b) Pattern layout of a 16×16 1T-nR array. For the bit lines, word lines, and source lines, the positions of metal contact pads are designed according to a specific probe card. (c) Top-view SEM image of the fabricated 16×16 1T-nR array with illustration of circuit schematic. (d) SEM image of the fabricated 16×16 1T-nR array with tilting and rotation.*

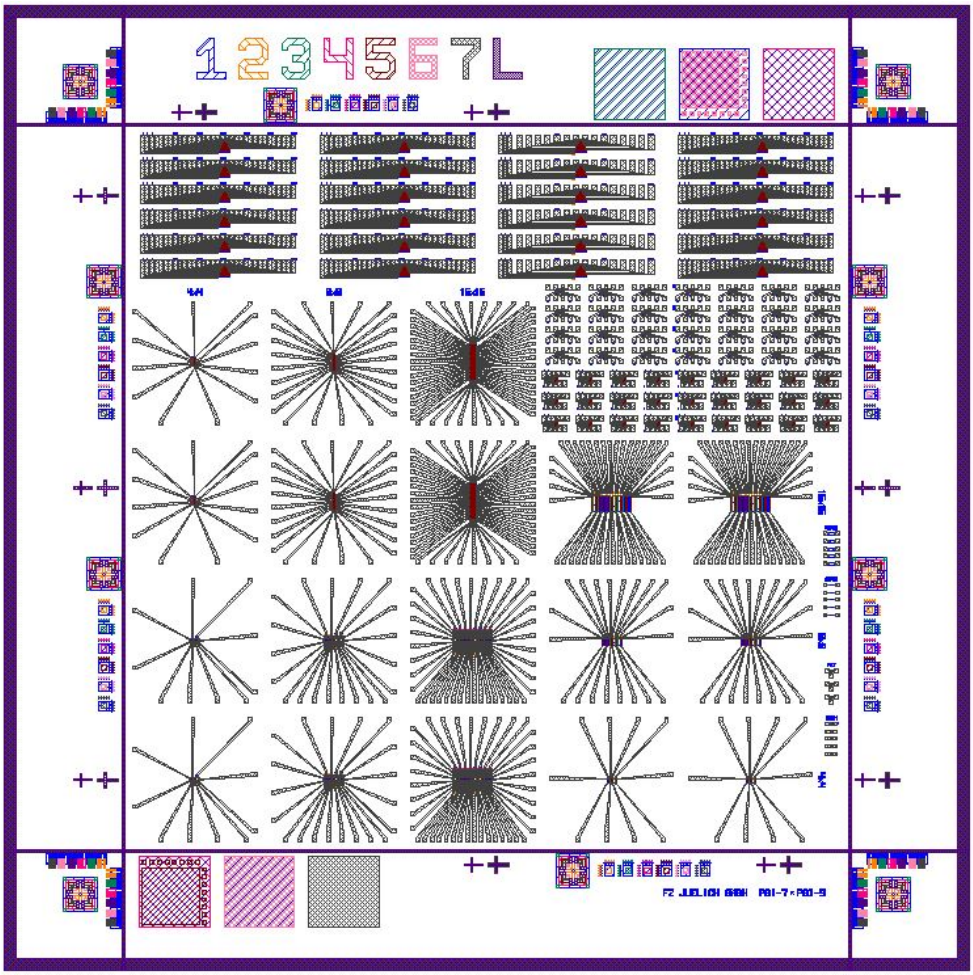

Figure 4.8: *Photomask layout with a substrate size of 2×2 cm² for 1T-1R and 1T-nR arrays. Besides different array sizes, several test structures and stand-along devices are included for process control. Multiple alignment marks are deployed at the corners and on the sides.*

4.3 Process flow

The fabrication process of the designed active RRAM arrays is comprised of two phases: MOSFET fabrication and RRAM co-integration. Initially, the fabrication of planar HKMG MOSFETs utilizes a gate-first method with a self-aligned process. To safeguard the transistors in subsequent process steps, a passivation stage follows prior to the RRAM co-integration. Next, the passivation layer is opened on drain contacts through a contact opening step, which allows for the creation of plug-type RRAM cells. The process concludes with metallization, which connects bit lines, word lines, and source lines, and establishes metal contact pads for future electrical characterizations. Figure 4.9 summarizes the process flow, while Figure 4.10a and Figure 4.10b portray a cross-sectional diagram and an equivalent circuit respectively of the 1T-1R unit cell.

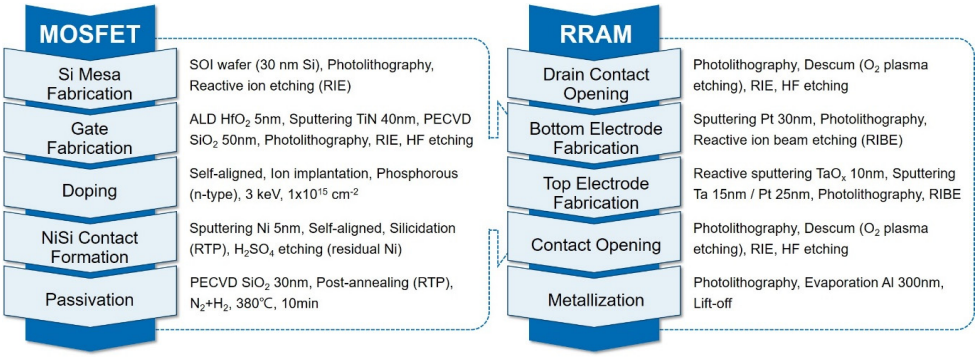

Figure 4.9: *Process flow of RRAM co-integration with planar HKMG MOSFETs.*

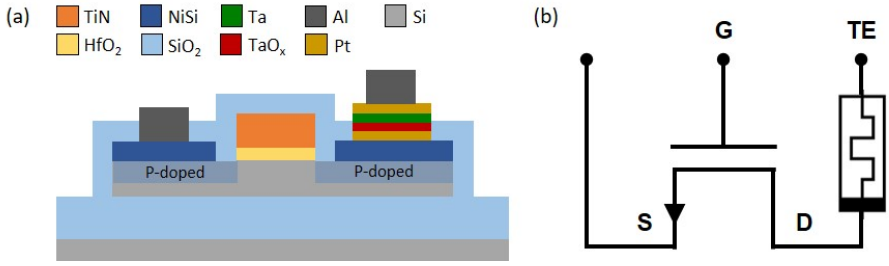

Figure 4.10: *Schematic diagrams of integrated 1T-1R unit cell. (a) Cross-sectional schematic of the 1T-1R unit cell. (b) Circuit schematic of the 1T-1R unit cell. The BE of RRAM is connected to the drain of N-channel MOSFET. The TE, gate, and source terminals are accessible to operate the 1T-1R unit cell.*

Careful considerations have been taken into account when designing the metal connections for the network of bit/word/source lines. For the 1T-1R configuration, the TiN layer for transistor's metal gate is directly used to connect the gates desirably, forming word lines of the array. Right after transistor fabrication, a passivation step is carried out to prevent any potential short circuit between the gate and S/D region caused by metal particle generated during further processing. Generally, metallization for connecting source lines and bit lines is therefore allowed only after the transistor passivation step. In the circuit layout of the 1T-1R array, as shown in Figure 4.6a, source lines and bit lines were designed to connect in different directions: source lines in the vertical direction and bit lines in the horizontal direction.

The crossing of source and bit lines necessitates a two-step metallization process with a dielectric layer passivating in between. To avoid complicating the process and adding unnecessary steps, a workaround was proposed and implemented that the source lines were connected without metallization while leaving only the bit lines connected by metallization. During transistor fabrication, source lines can be connected directly by the NiSi, which serves as the S/D ohmic contacts. Additionally, the concept of common source lines was applied by sharing one common source with two horizontally adjacent transistors, as shown in Figure 4.11. In summary, both word lines and source lines were formed without metallization before transistor passivation, while only bit lines were formed lastly by metallization.

For the 1T-nR configuration, the similar approach was adopted to limit the metallization steps. As shown in the circuit layout of 1T-nR array, previously in Figure 4.7a, word lines and bit lines connect in different directions, causing crossing of each other. Thus, NiSi was used to

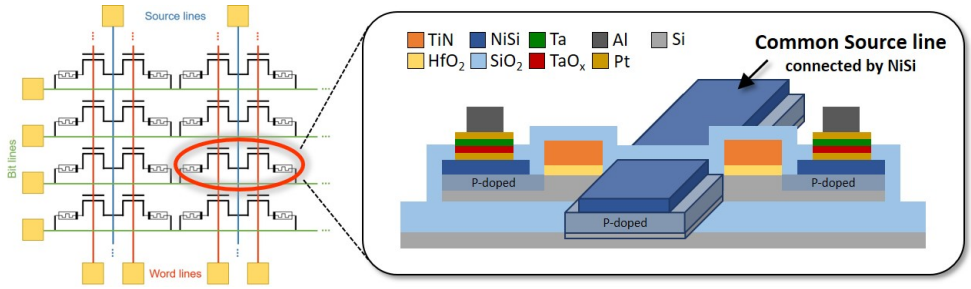

Figure 4.11: *Concept illustration of common source line in 1T-1R array*

connect BEs of RRAM cells, forming word lines before transistor passivation, while only the bit lines formed during aluminium metallization. This has been illustrated in Figure 4.12, taking the 1T-3R configuration as an example. The NiSi drain contact connects Pt BEs of three RRAM cells, simultaneously completing the connection of the word line.

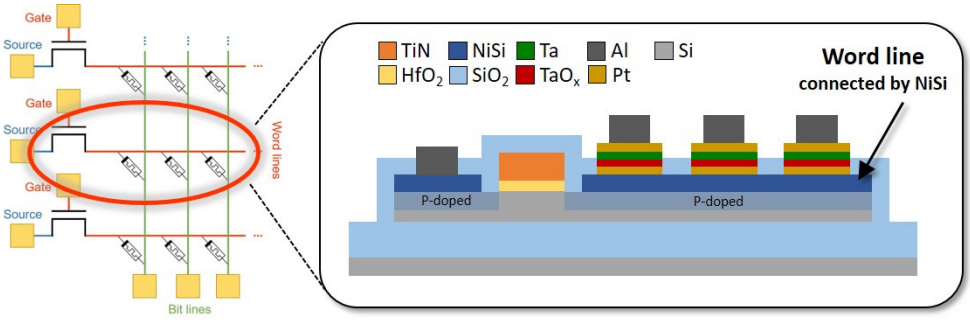

Figure 4.12: *Concept illustration of word line in 1T-nR array*

Following the concise overview of the process flow and the detailed explanation of the metal connection in the designed active RRAM arrays, the complete fabrication process will next be presented in a chronological order with comprehensive details.

4.3.1 Si mesa fabrication

The process of MOSFET fabrication started from an SOI wafer with 60-nm-thick Si on top of 145-nm-thick SiO_2 placed on bulk Si. The topmost Si layer was then thinned down to 30 nm through thermal oxidation and wet etching using 1% HF solution. After the wafer was diced into small pieces of 20×20 mm^2 substrates, a five-step RCA cleaning procedure followed. In the first step, the substrate was immersed in Piranha solution of H_2O_2 and H_2SO_4 in a ratio of 2:1 for 10 minutes to eliminate organic particles. A quick dip in a 1% HF solution was then performed for 10 seconds to remove an impurity-containing oxide thin film forming at the Si surface. Next, the RCA standard cleaning-1 (SC-1) was carried out by immersing the substrate in a mixture of deionized (DI) water, H_2O_2, and NH_4OH in a ratio of 20:4:1 at 60°C for 10 minutes to remove organic contaminants and traces of metals that had been transformed into metal complexes. Following a 10-second dip in 1% HF solution, which stripped a thin hydrous oxide as a result of SC-1, the cleaning procedure concluded with the RCA cleaning-2 (SC-2), using a mixture of DI water, H_2O_2 and HCl in a ratio of 20:1:1 at 60°C for 10 minutes to remove residual ionic contaminants and traces of heavy metals. As the clean substrate was prepared, Si

mesa was patterned using photolithography and RIE with the *UV6-0.6* positive photoresist. The RIE continued until a certain level of over-etching was obtained into the buried oxide (BOX) layer beneath the topmost Si layer. Finally, the photoresist was removed using a solution of dimethyl sulfoxide (DMSO) mixed with 0.03% cyclopentanone for 30 minutes, leaving behind the 30-nm-thick patterned Si on the BOX layer. This physical pattern defined the active area of transistors designed in the RRAM arrays, the channel width of the transistors, and the shape of the S/D regions. Figure 4.13 shows an example of the 16×16 1T-1R array after Si mesa patterning.

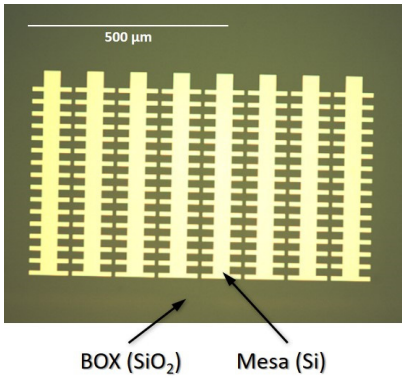

Figure 4.13: *OM image of the fabricated Si mesa pattern for a 16×16 1T-1R array.*

4.3.2 Gate fabrication

As the gate-first approach was employed in this work, the gate of MOSFET was fabricated right after the Si mesa. To ensure a contamination-free Si surface, the RCA cleaning procedure was performed again before depositing the gate dielectric. After a 10-second dip in 1% HF soluiton, a 5-nm-thick HfO_2 film was deposited using atomic layer deposition (ALD). Subsequently, a 40 nm-thick TiN metal gate was sputtered, followed by the deposition of a 50-nm-thick SiO_2 film through plasma-enhanced chemical vapor deposition (PECVD). The SiO_2 film on top served to prevent potential contamination issues during patterning, such as photoresist residue.

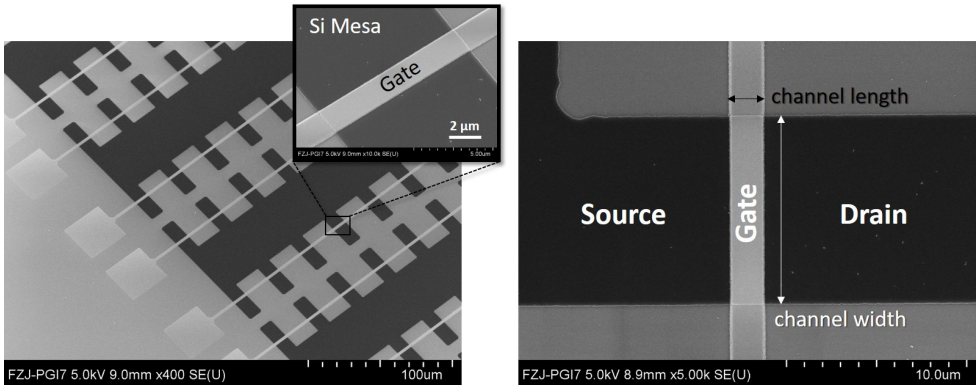

Figure 4.14: *SEM images of the fabricated gate pattern. The gate pattern crossing on top of the Si mesa is displayed from a tilting angle. The top-view image shows the definitions of MOSFET's channel length and width.*

Next, the gate of MOSFET was patterned using photolithography with the *UV6-0.6* positive photoresist and etching process involving RIE and HF etching. The RIE removed the 50-nm-thick SiO_2 and the 40-nm-thick TiN, while the 5 nm-thick HfO_2 was etched in 1% HF solution to prevent any damage potentially from the RIE process to the Si surface at the S/D regions. Finally, the photoresist was stripped using a solution of DMSO mixed with 0.03% cyclopentanone for 30 minutes. The result of gate fabrication is shown in Figure 4.14, taking the 16×16 1T-1R array as an example. By patterning the HKMG stack, the gate region on the Si mesa was defined, automatically defining the S/D regions. The gate width defined the channel length, resulting in a channel length of 2 μm and a channel width of 10 μm. The gates were fabricated in a long-line structure crossing the Si mesas in columns, as shown on the left in Figure 4.14. The TiN in the gate stack served as metal connections among the transistors in columns, indicating that the word lines of the array were already formed at this stage.

4.3.3 Doping of Source and Drain

After fabricating the gate of MOSFET, ion implantation was carried out to dope the S/D regions. The TiN metal gate fabricated can serve as a mask covering the channel region, enabling a self-aligned process that allowed a selective doping on the S/D regions. This particular advantage is a key feature of the gate-first approach. Here, an n-type dopant, Phosphorous (P), was implanted at a dose of 1×10^{15} cm^{-2} with an accelerating voltage of 3 keV. Subsequently, these dopants are activated by RTP in an Ar ambient at 1000°C for 15 seconds.

4.3.4 Nickel silicidation process for source and drain contacts

Once the implanted dopants were activated, the essential elements of MOSFET, namely the gate, source, and drain, were basically complete. However, metal electrodes with ohmic contacts are required to ensure a proper electrical access to these terminals. While the TiN used for the metal gate can provide an ohmic contact for subsequent metal connections, the S/D terminals need additional process steps to create ohmic contacts on them. To this purpose, metal silicides are a class of materials that are commonly adopted in modern CMOS technology. The metal silicides, such as the NiSi used in this work, can be selectively formed on the Si surface without the need of patterning. When the ohmic contacts on the S/D terminals are obtained by the metal silicide on the P-doped Si, the metal connections of the circuit can then be fabricated by normal highly-conducting metals, such as aluminum.

Before forming the NiSi on the S/D regions, a SiO_2 spacer enclosing the gate region was formed to avoid short circuits between the gate and S/D. Without the spacer, the NiSi formed on the P-doped Si would be right adjacent to the gate edge. Even a slight touch of NiSi on the edge of TiN would cause a short circuit between the gate and S/D, destroying the functionality of the MOSFET device. To fabricate the spacer, 50-nm-thick SiO_2 was deposited by PECVD at 280°C. A photolithography step with the *AZ® MiR™ 701 14cps* positive photoresist was performed and followed by RIE with CHF_3 gas and wet etching in 1% HF solution. The photoresist was stripped off afterwards in DMSO at 80°C for 4 hours. The fabricated spacer is illustrated at the top in Figure 4.15.

Next, 10-nm-thick Ni was sputtered after a 10-second HF dip and an in-situ Ar sputtering step to ensure oxide-free P-doped Si surfaces on the S/D regions. An RTP step was immediately carried out at 500°C for 30 seconds in a forming gas mixture of N_2 and H_2 to initiate silicidation at the interface of Ni and P-doped Si, resulting in the formation of NiSi. In the SEM image at the center of Figure 4.15, irregular islands of unreacted Ni can be observed outside the P-doped Si area after the RTP. These islands were formed by the Ni atoms situated on the non-Si surface, such as SiO_2, because they did not participate in the silicidation reaction during the RTP. In

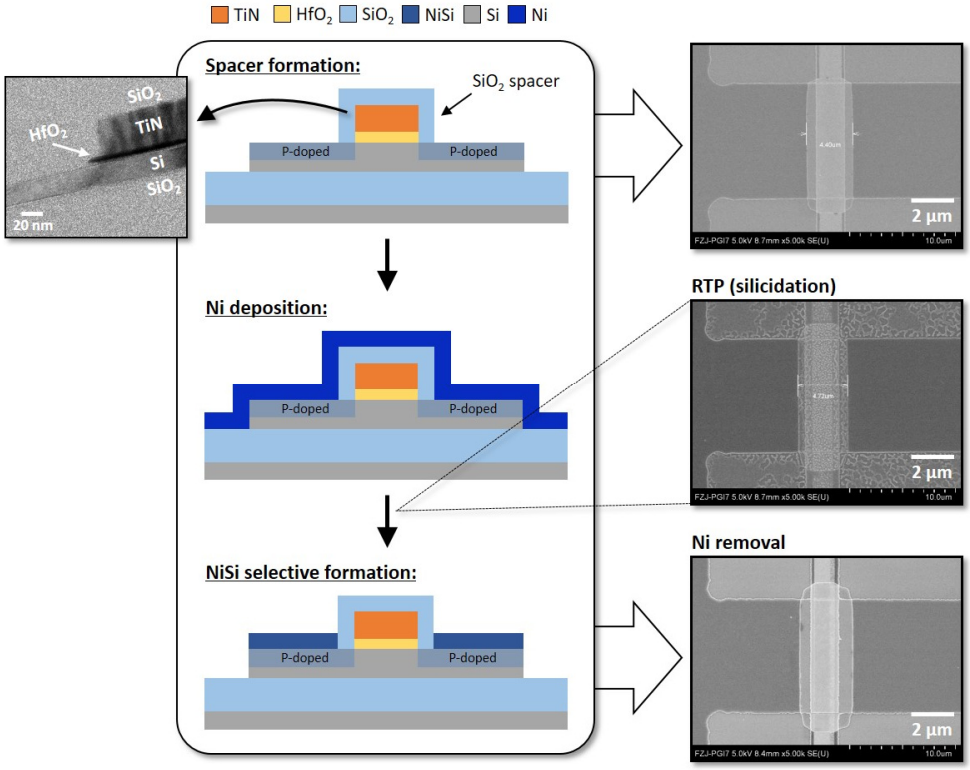

Figure 4.15: *Process of NiSi formation on source and drain contacts. A SiO_2 spacer is formed first to avoid the later formed NiSi layer causing short circuits between the gate and the S/D. The result is displayed by a top-view SEM image and a cross-sectional TEM image at the top. Ni deposition is immediately followed by RTP to enable scilicidation, whose result is shown by a SEM image at the center. After wet etching in the H_2SO_4 aqueous solution, the Ni remaining on top after the scilicidation is removed, and the NiSi layer is exposed. The final result is shown by a SEM image at the bottom.*

fact, they migrated and formed local clusters. In contrast, on the P-doped Si area, no such island was found as a smooth NiSi formation was observed with unreacted Ni homogeneously left on the surface.

For the removal of remaining Ni, a highly selective wet etching was performed using the H_2SO_4 aqueous solution in a ratio of $H_2SO_4 : H_2O = 1 : 2$. The use of H_2O_2 was avoided to prevent etching of the TiN metal gate. After the etching for 4 minutes, the effectiveness of the Ni removal process is presented in the bottom right SEM image of Figure 4.15.

4.3.5 Transistor passivation and annealing

After the NiSi was formed for the S/D ohmic contacts, a passivation layer of 30-nm-thick SiO_2 was deposited via PECVD at 280°C. To passivate any possible defects, especially oxygen vacancies, in the gate oxide and improve the transfer characteristic, an annealing process was conducted, subsequent to the transistor passivation, by an RTP step at 380°C for 10 minutes in a forming gas mixture containing N_2 and H_2 [111-115]. The effect of the annealing process on the transistor performance is illustrated in Figure 4.16.

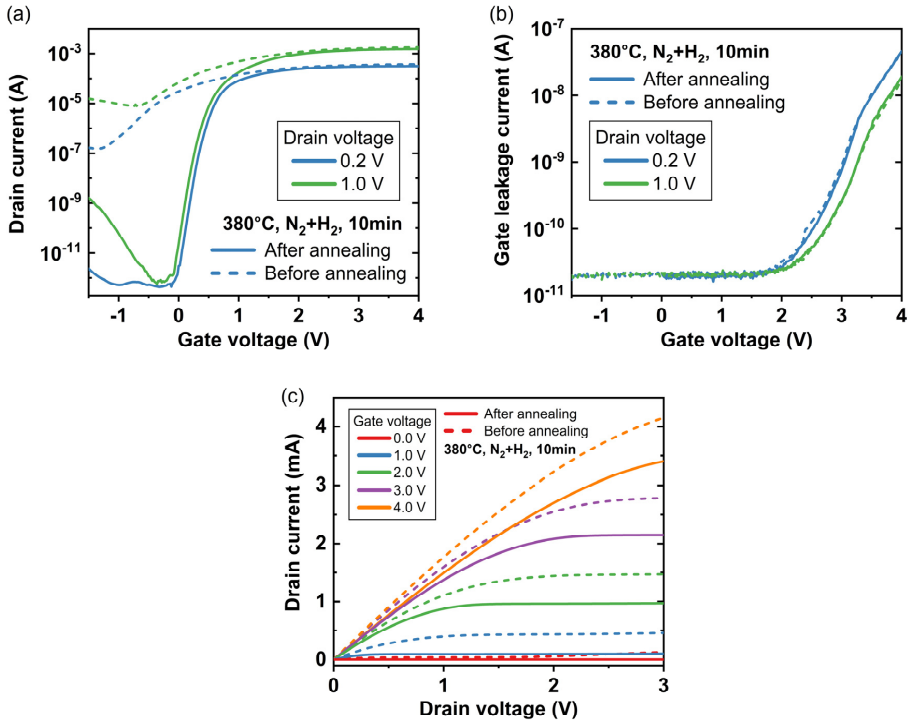

Figure 4.16: *Effect of annealing after passivation on transistor characteristics. (a) Transfer characteristic before and after annealing at two drain voltages, 0.2 V and 1 V. (b) Gate leakage current, as a function of gate voltage, before and after annealing at two drain voltages, 0.2 V and 1 V. (c) Output characteristic before and after annealing at gate voltages ranging from 0 V to 4 V.*

Figure 4.16a displays a remarkable improvement in the transfer characteristic of the MOSFET after annealing, where the subthreshold slope is significantly improved. With a considerable suppression of the OFF current, the ON current can still be kept at the same level, indicating a full control of the carrier channel by the gate voltage. The performance improvement is a result of the defect annihilation during the annealing process, passivating the defects in the gate oxide, which are likely to trap charges. When the defect concentration is reduced, the built-in potential caused by the trapped charges is alleviated. With a lower built-in potential adding to the gate voltage applied externally, the threshold voltage was slightly increased, and the OFF current was significantly suppressed, improving the overall subthreshold behavior of the MOSFET. In addition, Figure 4.16b shows a great overlap of the solid and dashed data curves, suggesting that a low gate leakage is nicely hold after the annealing process.

The comparison in the output characteristic shown in Figure 4.16c, specifically for the gate voltage in a range of 0 V to 4 V, indicates a decrease in the drain current after the annealing process, resulting from the slight increase in the threshold voltage. The drain current, despite the decrease, can achieve the range of mA within the operational limit of the gate voltage. This current level is critical for the RESET process of the RRAM device in 1T-1R configuration, where a high current is required.

4.3.6 Drain contact opening for RRAM integration

The passivated transistor, shown in Figure 4.17 on the left, had all terminals buried under the passivation layer of a 30-nm-thick SiO_2 layer. To integrate an RRAM device on the drain terminal, in the case of 1T-1R configuration, it is necessary to open a contact hole such that the NiSi drain contact was accessible from the surface. The designed contact hole has a width of 2 µm and a depth of 30 nm, as illustrated in Figure 4.17 on the right.

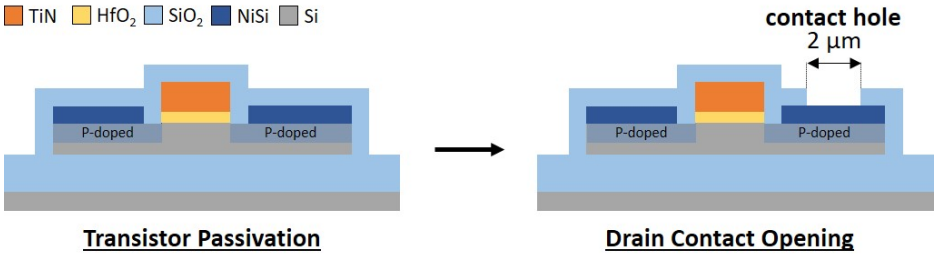

Figure 4.17: *Cross-sectional schematic of drain contact opening for a 2-µm-wide contact hole.*

To pattern the 2-µm-wide contact holes, a photoresist mask was patterned first using photolithography with the *AZ® MiR™ 701 14cps* positive photoresist in combination with an inverted exposure. Next, an descum step was carried out to ensure that there was no resist residue left in the hole pattern after development due to non-ideal exposure or development. The descum process is typically an O_2 plasma etching step that is a well-established technique for etching organic materials, particularly lithography resists. During the O_2 plasma etching at a low etching rate, approximately 20 nm of the photoresist was etched away, given the original thickness of 850 nm. With the hole pattern on the photoresist, the 30-nm-thick SiO_2 passivation layer was etched via RIE with CHF_3 and wet etching in 1% HF solution in order to expose the NiSi surface. The etching result is shown in Figure 4.18, taking the 16×16 1T-1R array as an example, with clear 2-µm-wide contact holes visible next to the gate of each transistor. Finally, the photoresist was stripped off using DMSO at 80°C for 4 hours.

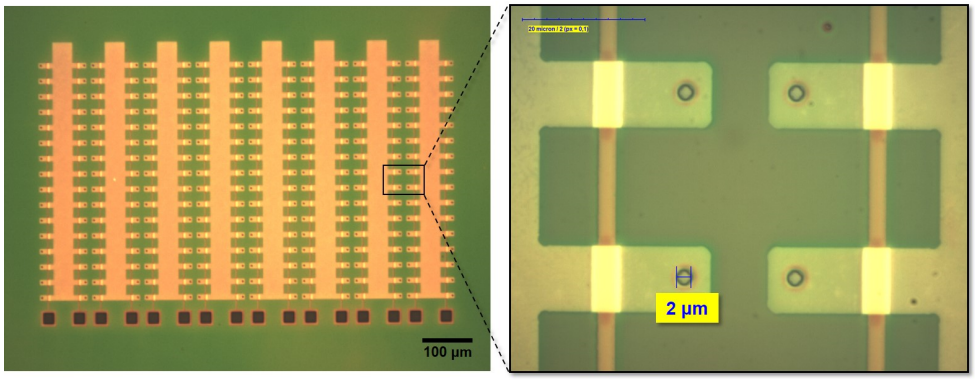

Figure 4.18: *2-µm-wide drain contact holes patterned after etching with the photoresist mask.*

4.3.7 Fabrication of TaO_x RRAM devices

With the access to the drain terminals of transistors, TaO_x RRAM devices were fabricated directly on the contact holes. The fabrication of Pt/TaO_x/Ta/Pt RRAM devices involved two main parts, in which the first part was dedicated to BE while the other to TE of the RRAM devices. Both parts follow the same sequence of process steps, including deposition, photolithography, etching, and photoresist removal. The deposition was carried out through sputtering, while etching through RIBE. The *AZ® MiR™ 701 14cps* positive photoresist was utilized for the photolithography and stripped off after the RIBE by DMSO at 80°C for 4 hours.

Right before the sputtering of the Pt BE, the native oxide layer on the exposed NiSi surface was removed by a 10-second dip in 1% HF solution. Subsequently, 30-nm-thick Pt was deposited using DC sputtering, which was patterned using RIBE with Ar. On the other hand, the TE fabrication involved a series of in-situ depositions, starting from reactive sputtering of 10-nm-thick TaO_x to RF sputtering of 15-nm-thick Ta, and ending with DC sputtering of 25-nm-thick Pt. The TaO_x/Ta/Pt stack was patterned using RIBE first with Ar for Pt and Ta and later with CF_4 for TaO_x.

A top view of the Pt/TaO_x/Ta/Pt RRAM device fabricated under SEM is shown in Figure 4.19 with a cross section illustrated on the right, providing a clear representation of patterning the integrated RRAM device in practice. The 30-nm-thick Pt layer was patterned in a size larger than the contact hole (region A), forming the Pt BE (region B) of the RRAM device. The stack of 10-nm-thick TaO_x, 15-nm-thick Ta, and 25-nm-thick Pt were patterned together to fully cover the BE (region B), forming the TE (region C) of the RRAM device.

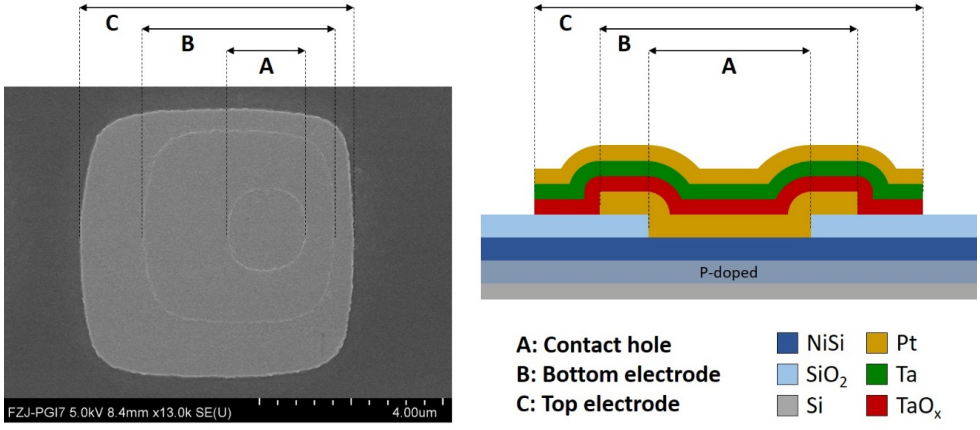

Figure 4.19: *Structure of the plug-type TaO_x RRAM device patterned by photolithography and RIBE. A top view of the fabricated device is displayed by the SEM image on the left. A cross-sectional schematic of the actual device structure is shown on the right. The dimensions of contact hole, bottom, and top electrodes are indicated by the regions A, B, and C respectively.*

4.3.8 Contact opening for metallization

Once the RRAM devices were fabricated, the final task until completion was to establish inter-connection for the network within the RRAM array and form metal pads for electrical characterizations. Prior to this stage, however, only the TEs of the RRAM devices were accessible from the surface, while the passivation layer of 30-nm-thick SiO_2 covered the remaining area. Therefore, another contact opening step was necessary before metallization,

enabling access to the gate, source, and some drain terminals. The same procedure as the drain contact opening for RRAM integration, described earlier, was applied at this stage.

4.3.9 Metallization for inter-connection and probing pad formation

The final step in the fabrication process was metallization, which created a patterned metal layer to connect desired array circuits and form probing pads. To reduce series resistance of inter-connection in the circuit, a thick metal layer is typically used for this purpose. The most efficient and effective way to pattern a thick metal layer is a lift-off process of photolithography. Initially, photolithography was carried out using the *AZ® nLOF™ 2020* negative photoresist, allowing the desired pattern to be transferred as a hollow pattern to the photoresist. To remove the native oxide on the NiSi and TiN surfaces in the opening contact holes, a 5-second dip in 1% HF solution and an in-situ Ar sputtering step prior to metal deposition were carried out. Finally, 300-nm-thick Al was deposited via an evaporation process, and the lift-off step was performed in the *TechniStrip® NI555*, a photoresist remover dedicated to the *AZ® nLOF™ 2020*, for 4 hours. Figure 4.20 shows the 16×16 1T-1R array after the metallization step.

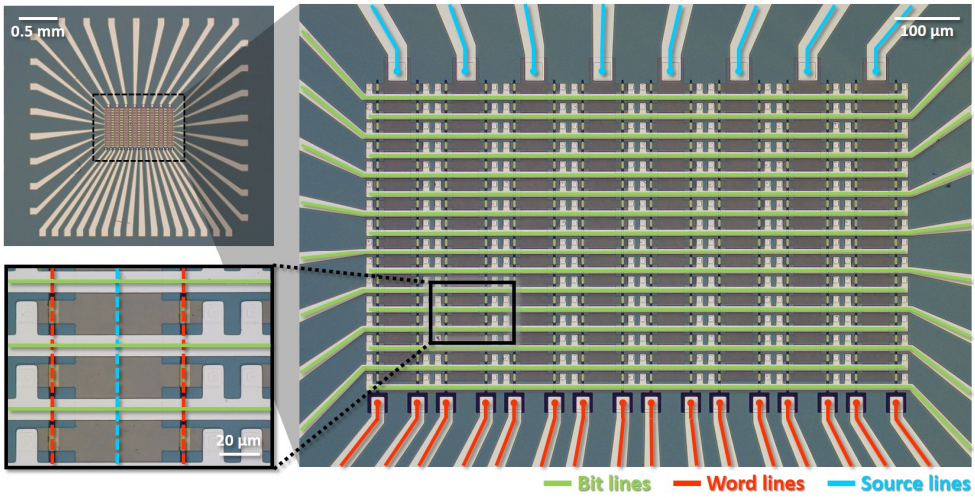

Figure 4.20: *The completely fabricated 16×16 1T-1R array after metallization. The metal connection of bit lines, word lines, and source lines are marked on the OM images. The dashed lines indicate the connection completed below the SiO_2 passivation layer, which prevents short circuits between bit lines and word/source lines.*

In order to better understand the metallization layout depicted in Figure 4.20, further explanation is provided. The word lines (orange) and source lines (blue) were formed already at the stage of transistor fabrication and buried under the passivation layer. As a result, only probing pads need to be formed and connected to the open terminals of the word lines and source lines in the array circuit. On the other hand, the bit lines (green), connecting RRAMs in rows, were formed together with their probing pads during metallization.

When conducting the electrical characterization, metal probes are used to connect a testing instrument to certain terminals of a fabricated circuit via a physical contact on the probing pads. In order to ensure a proper contact of the metal probes, a dimension of 75 µm is recommended for the metal pads. Compared to the micro devices in the array, the metal pads require a much

larger area. Consequently, the metal pads were positioned at a good distance from the connected terminals in the array circuit. For the array characterization, a probe card consisting of multiple probes with a fixed spatial arrangement is used. Therefore, the probing pads must be formed at relative positions according to the probe card layout. The available probe card for this work has 16 probes at four sides with a pitch of 150 µm, forming a square shape. Correspondingly, it can be found in Figure 4.20 that the probing pads surrounding the 16×16 1T-1R array are well distributed at four sides, also forming the square shape. In total, there are 16 word lines (orange) connecting to the probing pads at the bottom, 8 source lines (blue) connecting to the probing pads at the top, and 16 bit lines (green), of which eight connect to the probing pads on the left while the rest on the right.

4.4 Process issues

In the stage of process development, various process issues appeared in different forms. To tackle each process issue, a series of experiments were required to study the possible causes and validate the proposed solutions. The upcoming sections will detail the most challenging and enlightening issues during the process integration in a chronological order.

4.4.1 Photolithography of 2-µm-wide contact holes

Lithography is a substantial technology bottleneck especially for feature downscaling in semiconductor manufacturing. In general, the resolution limit of lithography is not only restricted by the light source's wavelength but also affected by factors such as the photoresist material, substrate material, and exposure equipment. When the feature size approaches to the resolution limit, the fidelity loss becomes more pronounced. Owing to this phenomenon, one of the biggest process challenge was found when patterning the smallest feature of this work by using the *Süss MA6*, a research-use mask aligner, for exposure under the i-line (365 nm) of Hg lamp. The smallest feature among the entire fabrication process is the contact hole with areas of 2×2 µm^2, which exposes the NiSi drain contact underneath the passivation layer such that the RRAM device can be fabricated on top.

The first attempt on the photolithography of 2-µm-wide contact holes was based on the *AZ® nLOF™ 2020* negative photoresist. The photoresist was diluted with the *AZ® EBR Solvent*, which is a photoresist thinner recommended from the supplier especially for *AZ®* photoresists, in a ratio of 1:1. The diluted *AZ® nLOF™ 2020* photoresist has a thickness of 400 nm, given a spin coating speed of 4000 rpm. The result after development is shown in Figure 4.21. The OM images were taken with a light filter, resulting in a greenish effect. In addition to small holes of 2×2 µm^2 for fabricating RRAM devices, big holes of 30×30 µm^2 created simultaneously for other irrelevant purpose can be found in the OM image. The big square holes are overall in a good shape, indicating a proper development result. However, the shape of small holes appears to be a cross with a dot in the center, severely deviated from a square, as shown in the SEM image on the bottom right. To explain the underlying cause, a hint can be taken from the observation on the corner of the big square hole, shown in the OM image on the top right. The corner actually turns out not a sharp right angle and even not a truncated right angle. Details can be carefully recognized from the pattern boundary that two blue arrows point at concave curves and a red arrow points at a convex curve. When the square size shrinks from 30×30 µm^2 to 2×2 µm^2, four corners are brought closer. Owing to the fact that the corner distortion scales larger than 1 µm and the adjacent corners have a distance of 2 µm, the corner distortions merge together eventually causing a cross shape.

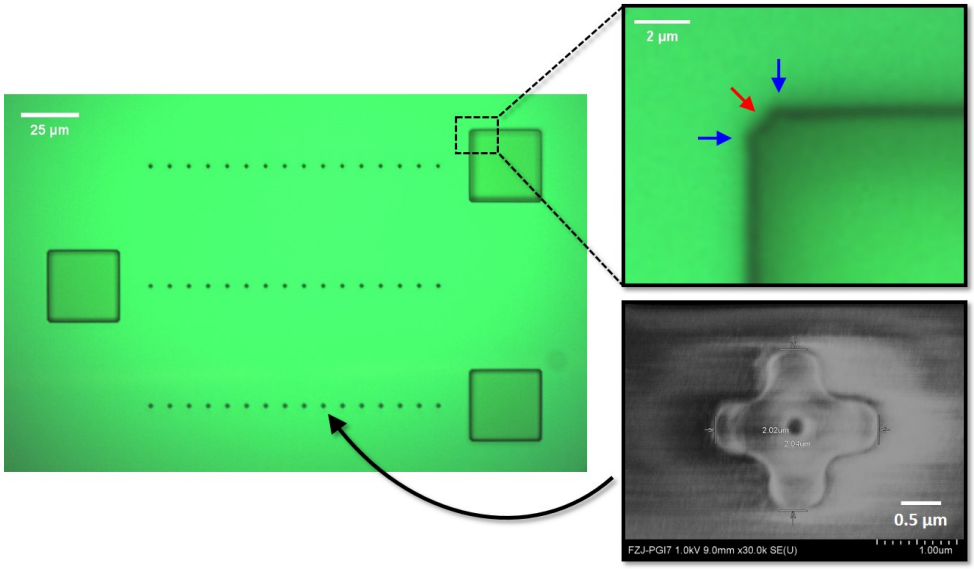

Figure 4.21: *Photolithography pattern distortion using the diluted AZ® nLOF™ 2020 photoresist. Each corner of large square holes is distorted and deviated from a sharp right angle. Two blue arrows mark concave distortions and a red arrow marks convex distortion. Small square holes with areas of 2×2 µm² become cross-shaped patterns with single dots at the centers, as shown in the SEM image.*

The corner distortion is attributed to the light interference during exposure, which is derived from the non-ideal exposure at the pattern edge [116, 117]. Even when the photomask is in contact with the photoresist, a certain degree of exposure at the masked area around the edge should be considered in actual circumstances, as illustrated in Figure 4.22. When two edges of the masking pattern encounter at a corner, the non-ideal exposure area around the edge overlaps, inducing light interference under the photomask. The exposure with the light interference at the corners, therefore, leads to the result previously shown in Figure 4.21.

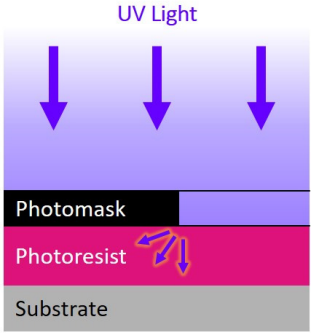

Figure 4.22: *Schematic diagram of non-ideal exposure in contact lithography using a mask aligner.*

To solve the issue, the *AZ® 5214 E* image reversal photoresist was studied as a possible alternative photoresist based on a different photochemical mechanism that shows a different

degree of the impact from the light interference [118, 119]. The *AZ® 5214 E* utilizes photochemical reactions based on diazonaphthoquinone (DNQ), unlike chemically amplified photoresists such as the *AZ® nLOF™ 2020* negative photoresist [117, 120]. With exposure on the *AZ® nLOF™ 2020*, a carboxylic acid forms and activates a melamine cross-linker. In the subsequent post-exposure bake (PEB), short phenolic resin units are cross-linked into longer chains, which decreases the solubility in the developer by several orders of magnitude. On the other hand, the *AZ® 5214 E* has a more complex mechanism. The DNQ sulphonate in the *AZ® 5214 E* is an established photo-active substance that is converted into an indene carboxylic acid upon exposure by releasing a nitrogen molecule and binding with a water molecule, as depicted in Figure 4.23, whereby the solubility in the developer increases by several orders of magnitude. In the subsequent PEB, or so-called image reversal bake, the indene carboxylic acid is neutralized with a special cross-linking agent activated at the elevated temperature, which brings the solubility back down and eliminates the photosensitivity. Herein, the previously unexposed area remains as a normal positive photoresist. Simply upon a flood exposure without a photomask, this area becomes highly dissolvable in the developer, overall creating a negative pattern after development.

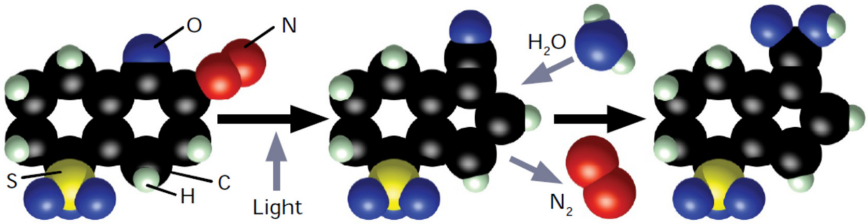

Figure 4.23: *Photochemical reactions in the AZ® 5214 E photoresist. The DNQ sulphonate is converted into an indene carboxylic acid upon UV light exposure by releasing a nitrogen molecule and binding with a water molecule [117].*

By using the *AZ® 5214 E* diluted with the *AZ® EBR Solvent* in a ratio of 2:1, the photoresist thickness is 550 nm, given a spin coating speed of 5000 rpm. The result displayed in Figure 4.24 shows small holes in the shape of a rounded square. It is a clear improvement from a cross with a dot in the center found previously in the result using the diluted *AZ® nLOF™ 2020*.

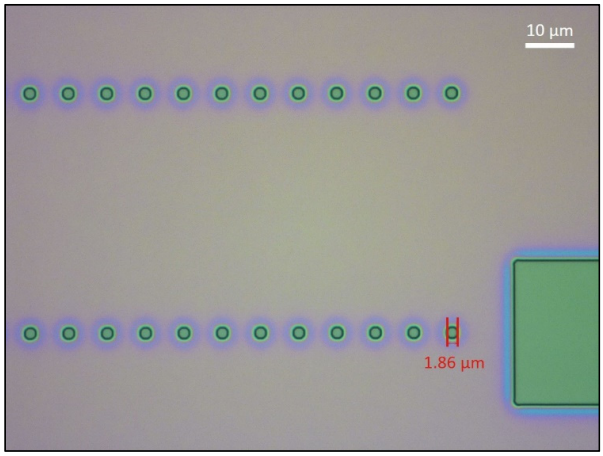

Figure 4.24: *Photolithography result using the diluted AZ® 5214 E photoresist.*

The result of the diluted *AZ® 5214 E* was investigated further in detail by SEM. With the surface covered mostly by the photoresist, an electrically insulating material, imaging by any electron microscopy is challenging due to the charging effect. Thus, the pattern on the photoresist should be transferred to the SiO_2 substrate so that it could then be examined by SEM after photoresist removal. The hole pattern was transferred by RIE using CHF_3, followed by O_2 plasma etching to remove the photoresist. As shown by SEM images in Figure 4.25, the shape of small holes massively transformed from a truncated square to a cross with a dot in the center. The variation from the location on the substrate can be deduced from the thickness variation of the coated photoresist, which is an inherent issue of spin coating that the thickness increases radially outward from the center in a closer look. As a conclusion, the use of the *AZ® 5214 E* photoresist alleviated the impact of light interference on the exposure result, but the location-related pattern distortion owing to the nature of photoresist thickness variation in spin coating still remained unsolved.

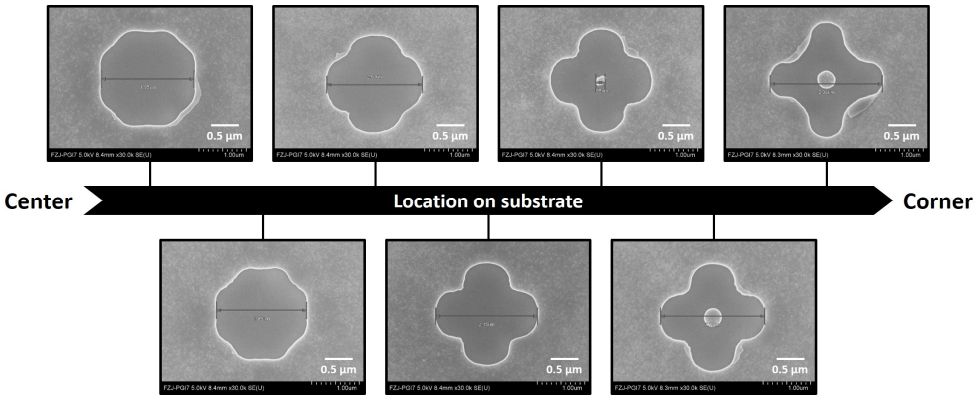

Figure 4.25: *Spatial dependence of photolithography result using the diluted AZ® 5214 E photoresist. Based on the fact that the designed pattern is square holes of 2×2 µm², larger pattern distortion is found further from the center and closer to the corner of the substrate. The result is taken after RIE and photoresist ashing on a SiO_2 substrate. All the SEM images are under an identical magnification.*

Moving the study focus away from photoresists, which is the recipient in the exposure process, the issue could be viewed from the perspective of light, which is the triggering source in the exposure process. Exposure by a mask aligner requires a photomask to define the exposure area. There are two types of photomask, namely bright field and dark field mask. Employed in the study above, a bright field mask only blocks the light on the desired pattern and most of the area on the mask is therefore transparent. On the other hand, a dark field mask only allows the exposure on the desired pattern, which makes most of the area on the mask not transparent. The corresponding circumstances of patterning a 2-µm-wide contact hole are depicted in Figure 4.26. The case shown on the most right is an advanced exposure technology by direct writing, which does not require a photomask and offers excellent flexibility. The exposure is fully controlled by programme in a scanning manner.

As identified and discussed previously, exposure with a photomask would suffer from the issue of undesired exposure under the photomask around the edge, which leads to pattern distortions particularly at the corners of desired pattern. For bright field mask, it affects inside the desired pattern. Whereas for dark field mask, it affects outside the desired pattern. Concerning the current case that the distortion cannot be reduced to an acceptable level, it is better to have the distortion outside than inside the desired pattern because the center part can

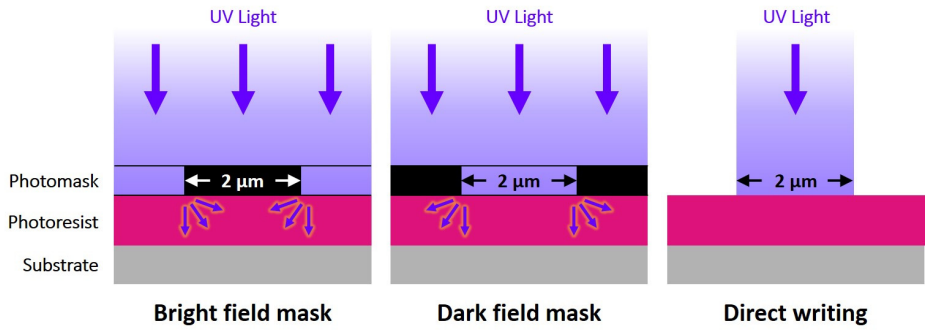

Figure 4.26: *Schematic diagrams of different UV exposure methods for a 2-µm-wide contact hole.*

remain unaffected with a good photolithography result secured. When no photomask is used in the exposure by direct writing, the issue of undesired exposure does not exist. Additionally, thanks to the maskless feature, it saves time and cost simultaneously because no photomask needs to be produced. Thus, direct writing was chosen for patterning the contact holes.

The direct writing in this work was implemented by the *MLA 100*, a commercial maskless aligner from Heidelberg Instruments. The technical details on exposure can be referred to Section *3.1.3.1.2 Maskless aligner*. By using the *AZ® MiR™ 701 14cps* positive photoresist, the area of desired contact holes was written by UV light, which was inverted from the exposure of a negative or image reversal photoresist with a bright field mask. To better ensure a clear opening in the contact hole, a descum step after development was performed that approximately 20 nm of the photoresist was etched away, given the original thickness of 850 nm. As shown by the OM images in Figure 4.27, the hole pattern appears to be a rounded square with no appreciable shape difference in regard to the location on the substrate. After all, direct writing was concluded as the solution to the challenging photolithography issue on 2-µm-wide contact holes. The exposure by direct writing perfectly avoid the drawback of using photomasks when the feature size is critically small.

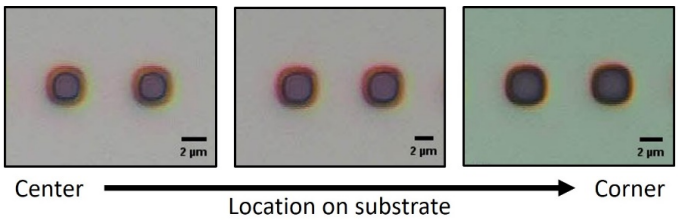

Figure 4.27: *Photolithography result of 2-µm-wide contact holes using the direct writing approach for exposure.* The *MLA 100 maskless aligner was used together with the AZ® MiR™ 701 14cps positive photoresist. No considerable spatial dependence is found in the result.*

4.4.2 Reactive-ion-etching of 2-µm-wide contact holes

After completing the entire fabrication, the TaO_x RRAM device in a 1T-1R unit cell was characterized individually with the electrical access to the drain terminal, where the BE connects. According to the quasi-static measurement by voltage sweep with current compliance applied in SET operations, inconsistent bipolar switching was observed for 30 cycles, as shown in Figure 4.28. Instead of gradual RESET processes typically found in VCM-type RRAMs,

abrupt RESET processes were observed. Due to lack of controllability in abrupt processes, an extremely high variability, around three orders of magnitude, was introduced to the HRS. Despite the relatively consistent $V_{RESET,C}$ at around 0.45 V, $V_{SET,C}$ widely varies from 2.2 V to 3.6 V, which is in general higher than the normal $V_{SET,C}$ below 1 V. Moreover, the $I_{RESET,max}$ can be up to 4.5 mA, while the I_{CC} during SET process was set as 0.5 mA. This suggests a large current overshoot during the SET process because the current compliance was not properly applied to the RRAM device. Overall, the observation from the switching behavior implies the possibility of an equivalent passive electronic component presenting in series to the RRAM device.

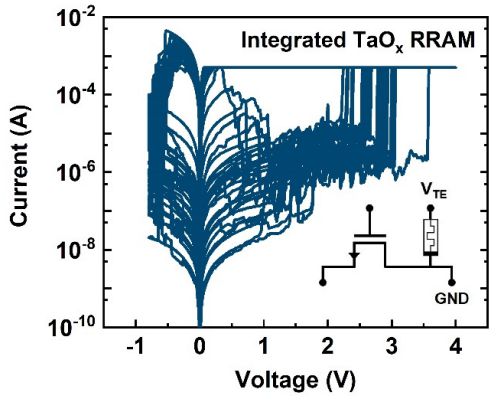

Figure 4.28: *I-V diagram of 30 inconsistent bipolar switching cycles from an integrated TaO_x RRAM. The measurement was carried out individually on the TaO_x RRAM integrated with a MOSFET as a 1T-1R unit. The RRAM's TE was voltage biased, while the BE connected to the ground, as illustrated by the inserted circuit schematic.*

To identify the issue of the integrated RRAM device by physical analysis, a cross-sectional TEM image was obtained together with an EDX line scan, which provides a depth profile of elemental characterization. The TEM image in Figure 4.29 shows the cross section of an integrated RRAM device with the EDX result attached on the right. The detected counts in EDX spectra were normalized to track each element simultaneously along the depth. Combining the EDX depth profile, the RRAM stack shown in the TEM image can be clearly distinguished from the capping electrode (Pt), the ohmic electrode (Ta), the switching layer (TaO_x), to the bottom electrode (Pt). In between the bottom electrode and the NiSi drain contact, an unknown layer can be found with a thickness of around 7 nm. According to the EDX result, a strong trace of O element is recognized by an intensity peak in the unknown layer, indicating an oxide layer. The unknown oxide layer, sandwiched by Pt and NiSi as electrodes, becomes a parasitic electronic component in series to the RRAM device that could behave as a resistor, a capacitor, or even a memristor in the system. Hence, the abnormal switching found in the integrated RRAM can be ascribed to the presence of the unknown oxide layer.

Before any solution can be proposed, the cause of the unknown oxide layer needs to be identified first. According to the previous study on TaO_x RRAM with NiSi BE in the subchaper 4.1, the thickness of interfacial oxide was around 2.5 nm when the TaO_x layer was reactively sputtered on the NiSi surface with an oxygen-containing ambient. Thus, for the current case, the unknown oxide with a thickness of around 7 nm is too thick to be an interfacial oxide or a native oxide on the NiSi when the Pt was sputtered with an Ar ambient after an HF dip. Taking

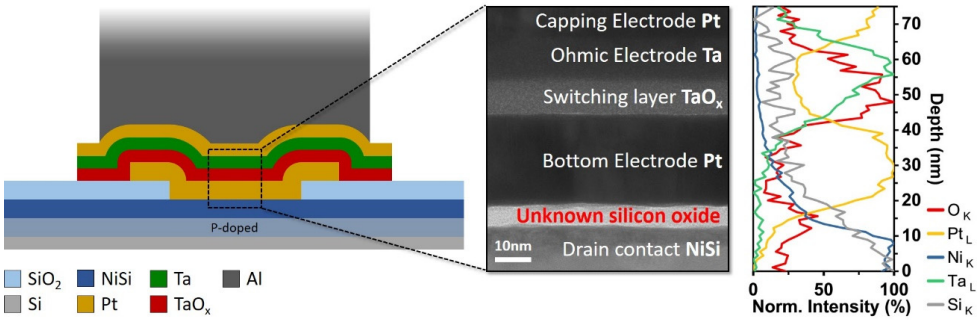

Figure 4.29: *EDX depth profile across the integrated TaO_x RRAM. The intensity of each characteristic X-ray is normalized to scale it by percentage. The depth axis is aligned to the cross-sectional TEM image at the side such that the result can directly relate to the layers observed in the material stack. In comparison to the schematic diagram of the ideal cross section on the left, an additional layer is found between the Pt BE and the NiSi drain contact. The elemental analysis from EDX suggests it is a layer of silicon oxide.*

one step back from the Pt deposition, the 2-µm-wide contact hole was etched via RIE with the photoresist mask. This etching step could possibly turn out insufficient and a part of SiO_2 passivation layer remained on top of the NiSi. According to the EDX result shown in Figure 4.29, the unknown oxide contains Si and possibly Ni. Considering the intensity decay rate of Si and Ni when entering the unknown oxide layer from the NiSi, Si shows a mild decay followed by a rapid drop when moving into the Pt layer. On the other hand, Ni shows a sharp decay right from the beginning. Consequently, it can be deduced that the unknown oxide is a silicon oxide that most likely does not contain Ni. Overall, the SiO_2 passivation layer remained after the insufficient etching is identified as the cause behind the unknown oxide layer found between the Pt BE and the NiSi.

During the RIE using CHF_3 for patterning the contact holes, a dummy SiO_2 substrate deposited by PECVD under the same condition as the passivation layer was used to monitor the RIE process, which confirmed an over-etching of 7 nm. In addition, the subsequent HF dip for 10 seconds before the Pt deposition etched another 4 nm of SiO_2. Based on the monitoring result of the etching process, no SiO_2 from the passivation layer was supposed to remain in the contact hole. However, this speculation is built on the assumption that the etching rate stays identical for the 2-µm-wide holes patterned by a photoresist layer. In the case of a large hole, it can be considered as an approximation of the etching scenario without a photoresist mask. The disparity to this scenario increases, when the hole size becomes smaller. Eventually, this brings out the microloading effect or known as aspect ratio dependent etching or RIE lag which has been observed and widely studied for decades [121-126]. The consequence of microloading effect is shown in Figure 4.30a. Under the same etching condition, the smaller hole ends up with a shallower etching depth, denoting a slower etching rate.

The proposed mechanisms of microloading effect are summarized in to four different models as shown in Figure 4.30b [122, 125, 127]. Firstly, the angle distribution of incident ions limits the number of ions that can reach the bottom of hole. Smaller opening leads to less ions attacking the surface at the bottom, thereby reducing the etching rate. Secondly, the desorption of volatile etching product is limited by the pressure difference from inside to outside the hole, obstructing the active ions from further reacting with the etching target and slowing down the overall etching process. Thirdly, polymerization takes place during etching that hydrocarbons from the etching gas and the etching products released from the photoresist form a polymer film on the sidewall and even at the bottom of hole. The surface passivated by the polymer

film, therefore, requires extra time to be etched through. Lastly, charges build up in the photoresist during etching deviates the traveling path of the active ions from straight to the bottom of hole.

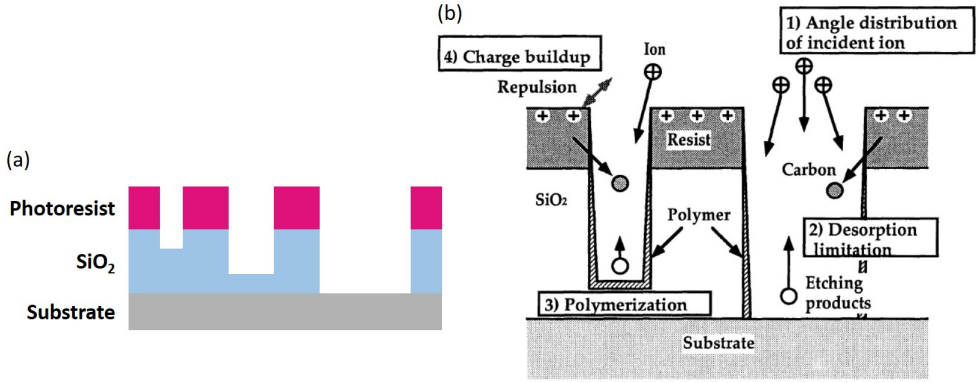

Figure 4.30: *Microloading effect in RIE. (a) Typical consequence of microloading effect. Photolithography patterns with different opening areas give different etching depth in a RIE process. The smallest pattern has the smallest etching depth and, therefore, the lowest etching rate. (b) Mechanism of microloading effect. There are four models proposed to explain the phenomenon [127].*

The solution to the microloading effect is rather simple for the current case because of the hole size is identical despite a minor variation when producing the 2-µm-wide contact holes. Thus, the etching rate is supposed to be identical for all the contact holes on the substrate. The reduced etching rate caused by microloading effect can be simply compensated by increasing etching time. Subsequent to the RIE, a short wet etching step, bypassing the microloading effect, can be added, for example 30 seconds in 1% HF solution, to ensure a reliable result. The pattern broadening, as a common result of wet etching, is tolerable here because the lateral dimension of contact hole does not influence the function-wise outcome.

Collected from literatures, several paths to alleviate the microloading effect can be studied potentially as long-term solutions [127-129]. Firstly, lowering pressure can increase the mean free path, thereby releasing the desorption limitation. Instead of using merely CHF_3, introducing CF_4 to the etching gas lowers the ratio of C/F, which has been found to alleviate the microloading effect. Moreover, adding a small amount of O_2 to the etching gas is able to mildly etch organic materials including the polymer film formed by polymerization during RIE. In addition to the pressure and the gas species, a lower temperature is favorable due to a smaller degree of polymerization. Lastly, using a hard mask in RIE is less prone to the microloading effect compared to a photoresist mask. However, to pattern and remove the hard mask complicates the process and brings in other compatibility issues to the entire fabrication process.

4.4.3 Overlay tolerance and pitch size limit

From a simple device to a complex circuit containing different devices and inter-connections, the required number of pattern layers or lithography steps increases dramatically. To achieve the functionality as per design, the placement of different pattern and the spatial relationships between them are critical. Overlay engineering is thus brought to the spotlight.

The overlay stands for the interplay of layer-to-layer alignments with practical imperfection. The overlay tolerance depends on the requirements from the pattern layout in order to implement the functionality of circuit designed. For example in the case of two different patterns in different layers, the requirement could be two patterns with no overlapping, any degree of overlapping, or a complete overlapping.

In this work, the overlay challenges were observed in the stage of RRAM integration. An ideal representation of the design is shown in Figure 4.31 under a cross-section view and a top view. According to the layout design, the bottom electrode (B) must cover the contact hole (A) to avoid the TaO_x layer in direct contact with the NiSi. Additionally, the top electrode (C) must cover the bottom electrode (B) to prevent a short circuit possibly caused by the Al layer for metal connection (D) in contact with the Pt bottom electrode (B). Whereas, the metal connection (D) only needs to overlap with the top electrode (C) to any degree in order to attain the connection to the top electrode.

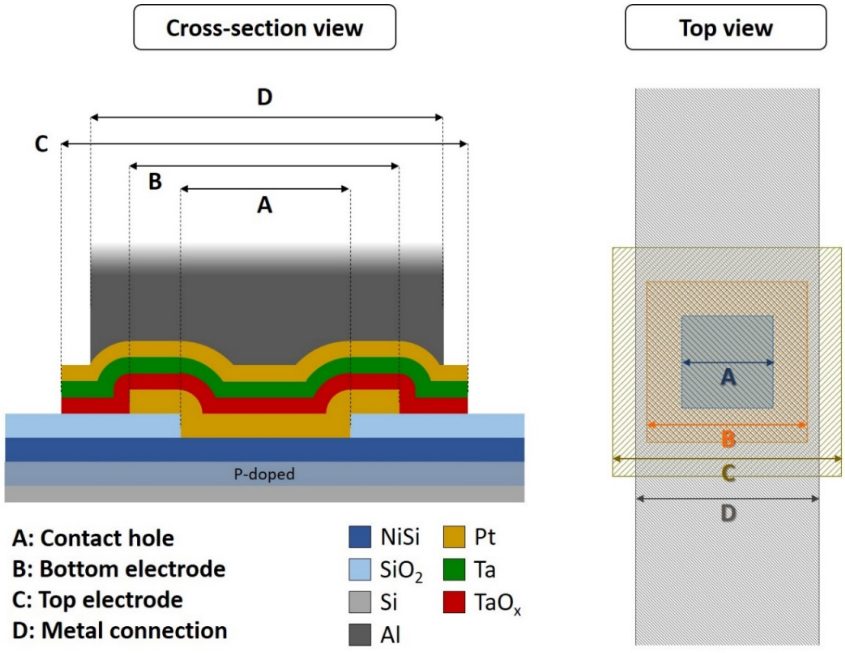

Figure 4.31: *Layout schematics of a TaO_x RRAM integrated on the NiSi drain contact of a MOSFET in a cross-section and a top views. The contact hole, bottom electrode, top electrode, and metal connection are indicated by the regions A, B, C and D respectively.*

Among all the overlay challenges, the simplest cases are limited to merely two pattern layers. The following example shown in Figure 4.32 is taken from the contact hole (A) and the bottom electrode (B). Ideally, A and B are centered to fulfill the prerequisite statement that A is fully covered by B. In general, an overlay failure appears when the statement is invalid by any mean. When aligning the pattern B to the pattern A, a large misalignment in +x direction can lead to an overlay failure when the overlay distance (d_{OL}) is larger than the buffer distance (d_{bf}). Considering the manufacturing imperfection on defined pattern dimension, overlay failures can also appear even when d_{OL} is smaller than d_{bf}, if the pattern A is enlarged or the pattern B is shrunk. Therefore, overlay control should include the discrepancy of pattern dimension from design to practice in addition to the misalignment.

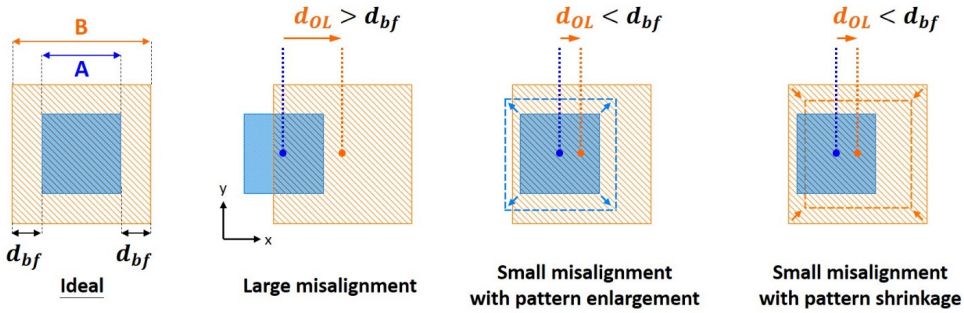

Figure 4.32: *Overlay issues of two pattern layers.*

Adding another pattern layer into the consideration brings out an interplay of misalignment in multiple layers. A practical scenario to address the interplay is simply taking the layer of top electrode (C) into account. Ideally, A, B, and C are centered to fulfill the prerequisite statement that A is fully covered by B and B is fully covered by C, as shown in Figure 4.33 on the left. When an alignment strategy using a floating baseline, is applied, each layer is aligned with respect to the previous one. No matter in which direction the misalignment (d_{OL}) is, the prerequisite statement stays valid when d_{OL} is smaller than d_{bf}. However, it is possible that the misalignments of multiple layers are in the same direction and the later produced pattern is then farther away from the initial baseline. The example shown in Figure 4.33 on the right demonstrates misalignments in +x direction for both the pattern B and C. The more the pattern layers are, the farther the pattern can be away from the initial layer. This can be problematic in manufacturing of integrated circuits where a significant amount of layers are commonly involved.

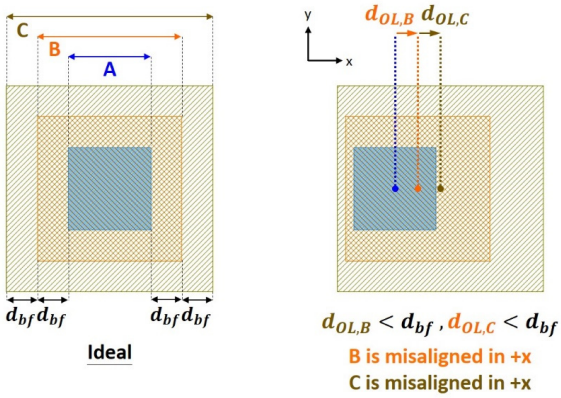

Figure 4.33: *Overlay issue of three pattern layers with misalignments in the same direction using the alignment strategy of floating baseline.*

To avoid the drawback from the use of floating baseline, another alignment strategy using a constant baseline was employed in this work that all subsequent layers are aligned with respect to the initial layer. The previous example is adapted and illustrated in Figure 4.34. After the pattern A is produced, the pattern B and C are both aligned with the pattern A. Nonetheless, an overlay failure can appear even though $d_{OL} < d_{bf}$, when the two layers are misaligned in opposite directions and the sum of $d_{OL,B}$ and $d_{OL,C}$ is larger than d_{bf}, as shown in Figure 4.34.

The pattern B is misaligned in +x direction while the pattern C in –x direction with respect to the pattern A, which disables a full coverage of the pattern B by the pattern C. Considering a d_{OL} maximum coming from a certain mask aligner, the d_{bf} must be at least two times of the d_{OL} maximum in order to ensure a successful result that fulfils the overlay prerequisite. Therefore, a larger d_{bf} is needed for the use of constant baseline, compared to the floating baseline.

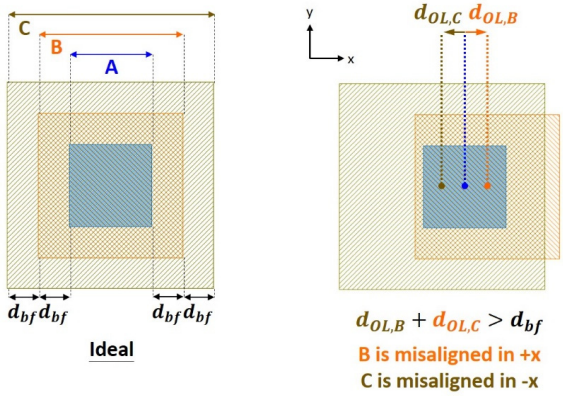

Figure 4.34: *Overlay failure of three pattern layers with misalignments in the opposite directions using the alignment strategy of constant baseline.*

Lastly, the pitch of duplicate patterns must be taken into account to avoid misconnections due to overlay. An example taken from the bottom electrode (B), the top electrode (C), and the metal connection (D) is shown in Figure 4.35. For one single set of patterns, the ideal design and a possible overlay result are sketched in Figure 4.35a. Ideally, B, C, and D are centered to fulfill the prerequisite statement that B is fully covered by C and D overlaps C to any degree. By aligning the pattern C and D both with respect to the pattern B, the example shows that the

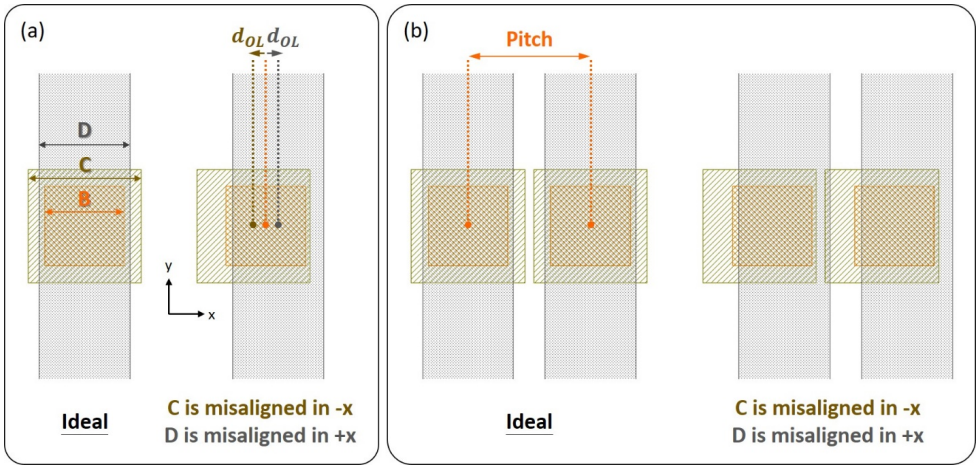

Figure 4.35: *Overlay issue related to pitch size. (a) Example of overlay result for a stand-along set of patterns. (b) Example of overlay failure for two sets of patterns with a certain pitch size.*

pattern C is misaligned in –x while the pattern D in +x direction both for a distance of d_{OL}, which still holds the overlay prerequisite. When the patterns are duplicated and placed right next to each other, the pitch size steps into play. Considering two identical sets of patterns with a certain pitch, the ideal design and the same overlay result are sketched in Figure 4.35b. Here, an overlay failure appears that the pattern D overlaps the pattern C next to it, causing a misconnection that leads to a short circuit between the top electrodes (C) of two adjacent RRAM devices. Hence, an appropriate pitch size is needed to tolerate the possible overlay. The worst overlay scenario should be taken into account when determining the pitch size in pattern layout design.

As the conclusion of the overlay study, the buffer distance and the pitch size in the layout design are extremely critical to the overlay tolerance. Practically, overlay issues arise from the misalignments due to manual operations and instrumental limits. In addition, the discrepancy of pattern dimension from the ideal to the practice must be taken into account together with the misalignment. Importantly, the worst overlay scenario needs to be tested and of a good awareness already at the layout designing stage so that a proper overlay tolerance can ensure a generous process window and a high process yield.

4.5 Electrical characterization of 1T-1R unit cells

After the entire fabrication was completed, electrical characterization was carried out on a 1T-1R unit cell to examine the fundamental property of bipolar resistive switching. Starting from the MOSFET characteristics as the foundation, the understanding of resistive switching in 1T-1R configuration is built up by a comprehensive comparison with 1R configuration and its capability of multi-level switching.

4.5.1 MOSFET characteristics in 1T-1R unit cells

The fabricated 1T-1R unit cell has four externally accessible terminals: gate, source, RRAM's TE, and a monitoring terminal on drain and RRAM's BE. With the access to the source, gate, and drain, the transfer and the output curves were measured when the RRAM's TE was left floating. In Figure 4.36a, the desired transfer characteristic can be found that the gate voltage (V_g) at 0 V turns off the carrier channel, giving a very low current. Above the threshold voltage (V_{th}) around 0.7 V extracted from the transfer curve, a high current over 1

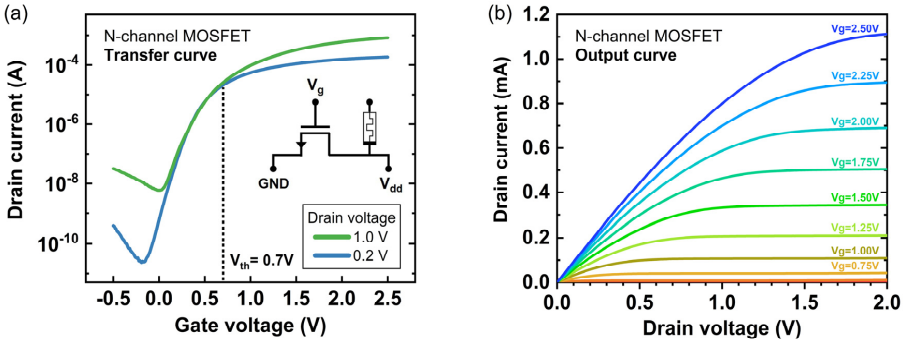

Figure 4.36: *Characteristics of N-channel MOSFET in the fabricated 1T-1R unit cell. (a) Transfer characteristic at two drain voltages, 0.2 V and 1 V. The extracted V_{th} is around 0.7 V. (b) Output characteristic at gate voltages ranging from 0.5 V to 2.5 V.*

mA can be delivered. In the 1T-1R unit cell, V_g at 0 V limits the current down to a very low level to block the unintended operation on the RRAM cell, while V_g above 0.7 V turns on the channel allowing high current passage to enable the intended operation on the RRAM cell. In Figure 4.36b, the output curve describes the drain-source I-V relation under a constant V_g, which can be divided into a linear and a saturation region. The saturation current forms a current limit for the RRAM cell connected in series in the 1T-1R unit cell. Higher V_g gives larger saturation current and therefore higher current compliance to the RRAM cell. According to the output curves of V_g applied up to 2.5 V shown in Figure 4.36b, the corresponding levels of current compliance can be referred to different V_g.

4.5.2 Comparison of resistive switching between 1R and 1T-1R configurations

A quasi-static characterization of resistive switching by DC voltage sweep was performed on a 1T-1R unit cell to compare its outcome with and without the use of transistor. The RRAM cell integrated in the 1T-1R unit cell can be individually characterized by biasing on the RRAM's TE, while grounding the MOSFET's drain, which connects to the RRAM's BE. The resistive switching with a current compliance of 100 µA applied during the SET processes directly by the sourcemeter was recorded for 60 cycles in Figure 4.37a.

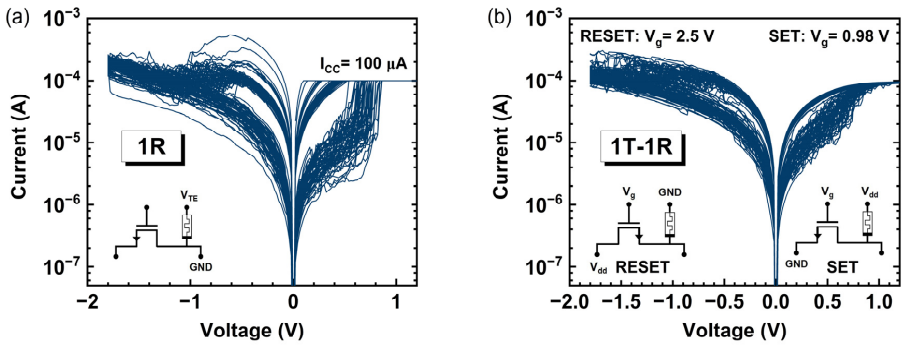

Figure 4.37: *Resistive switching property comparison of the integrated TaO_x RRAM device in 1R and 1T-1R configurations. (a) I-V diagram of 60 switching cycles in 1R configuration. The RRAM's TE was voltage biased, while the BE connected to the ground, as illustrated by the inserted circuit schematic. A current compliance of 100 µA was applied to the SET processes. (b) I-V diagram of 60 switching cycles in 1T-1R configuration. Two different testing schemes were used for the SET and RESET processes respectively, as illustrated by the inserted circuit schematics. The V_g was 0.98 V for the SET processes, and 2.5 V for the RESET processes. The $V_{g,SET}$ of 0.98 V is equivalent to a current compliance of around 100 µA.*

To characterize the entire 1T-1R unit cell, three of the four terminals were used. For SET operations, the RRAM's TE and the MOSFET's gate were positively biased respectively by the operating voltage (V_{dd}) and V_g, while the source terminal was grounded, and the drain terminal was left floating. For RESET operations, a larger V_g is normally applied to open the channel to its feasible maximum so that the saturation current is high enough to not be a current constraint hindering the RESET operation. The higher V_g also moves the saturation region towards higher drain voltage as shown in Figure 4.36b. This allows the transistor to work rather in the linear region such that it can leave a major portion of voltage drop on the RRAM cell to enable the RESET process. Theoretically, negative voltage biasing on the TE is the most direct

approach for RESET operations, but it was avoided to protect the gate dielectric from breakdown. If the TE was negatively biased, the drain terminal, connecting to the BE, would have a negative electrical potential with respect to the grounded source terminal. As a result, the potential difference from gate to drain is larger than V_g and more likely to exceed the breakdown voltage, which causes gate breakdown. Thanks to the symmetry feature in the physical structure of a transistor, drain and source can be swapped equivalently. Therefore, the RESET operation can be realized by grounding the TE and positively biasing the terminal on the other side of the transistor by V_{dd}. The resistive switching of the 1T-1R unit cell was recorded for 60 cycles in Figure 4.37b with the testing configuration noted accordingly. The V_g for SET operations ($V_{g,SET}$) was set at 0.98 V to attain the same I_{CC} at 100 μA as for the 1R case. The V_g during for RESET operations ($V_{g,RESET}$) was set at 2.5 V, and it should be noticed that the positive V_{dd} applied for RESET operations is marked with a negative sign in Figure 4.37b for expression.

With the current compliance and the RESET-stop voltage ($V_{RESET-stop}$) controlled at 100 μA and -1.8 V respectively, the comparison of switching property between 1R and 1T-1R configurations can be fairly addressed. As illustrated in Figure 4.37a and Figure 4.37b, the 1T-1R configuration has more self-consistent switching cycles than the 1R configuration despite a certain level of variability. A clear difference can be observed that the 1R configuration exhibits abrupt SET processes while the 1T-1R configuration shows more gradual SET processes. For filamentary-type switching to which VCM belongs, the SET process is a self-accelerating process involving thermal runaway, resulting in an abrupt current change. In the 1T-1R configuration, the abrupt current change is alleviated by the constraint from the transistor because the current can only follow the output curve of the transistor, going from the linear region gradually to the saturation region.

The relatively gradual SET process in the 1T-1R configuration indicates better current control that allows more effective implementation of current compliance to mitigate unfavorable current overshoot. Ideally, the abrupt current increase during the SET process is cut off once the predefined current value is reached by current compliance mechanism to control the maximum current loading on the RRAM device. In reality, a current spike over the compliance value commonly appears, which refers to current overshoot phenomenon due to non-ideal current compliance mechanism [130-134]. No current over compliance value is shown during the SET operations in Figure 4.37, because the current overshoot phenomenon typically cannot be observed directly from the measured current unless utilizing a dedicated instrument for fast current sensing. In this regard, the first-shown maximum current during the succeeding RESET process ($I_{RESET,max}$) is a useful indicator of the current overshoot [135, 136]. As a larger cross-sectional area of conductive filament allows more charge carriers to flow through per unit of time, a thicker filament can be expected from a higher current reached during the SET process. Generally, to initiate the RESET process of the thicker filament requires a higher current for enabling filament rupture. As a result, a larger current overshoot during the SET process can lead to a higher $I_{RESET,max}$ in the subsequent RESET process. Fetching from the measurements shown in Figure 4.37, the $I_{RESET,max}$ was recorded for the 60 cycles and plotted in Figure 4.38a. According to the $I_{RESET,max}$ distribution, the 1T-1R configuration has a lower median value and a smaller standard deviation than the 1R configuration, indicating minor current overshoots during SET processes with a smaller variability.

For the 1R configuration, current compliance is implemented externally by a sourcemeter that is also utilized for voltage sourcing and current sensing in measurements. Once it senses the current hitting the predefined compliance value, the corresponding electronics embedded in the system need a certain amount of reacting time to have the actual current compliance

function in place, causing the current overshoot in the SET process. As it originates from the instrument fault of sourcemeter, the current overshoot is an uncontrollable phenomenon with a stochastic nature. For the 1T-1R configuration, the transistor internally implements current compliance by utilizing its saturation feature of output current, which leads to the relatively gradual current increase during SET process. In addition, the 1T-1R configuration features a much less reacting time, because only one single transistor in proximity is employed for implementing the current compliance, unlike the 1R configuration relying on a complete electronic subsystem. Collectively, it offers a more effective current compliance, which favorably leads to reduced current overshoot with smaller variability.

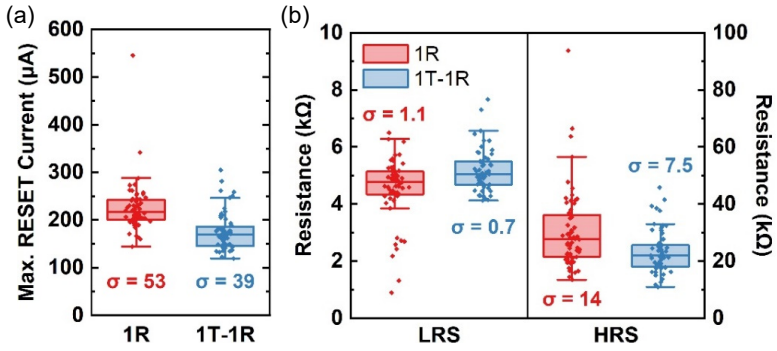

Figure 4.38: *Statistics of (a) $I_{RESET,max}$, (b) R_{LRS} and R_{HRS} over 60 switching cycles measured from the same RRAM device in 1R and 1T-1R configurations. The resistance state was read at 0.2 V, and the channel resistance of MOSFET was deducted.*

Given that a higher current reached during the SET process results in a thicker conductive filament, leading to a lower resistance of the LRS, a larger current overshoot under a constant I_{CC} causes a lower LRS. In Figure 4.38b, the LRS and HRS over 60 cycles were plotted with the corresponding standard deviations labelled, according to the DC measurements shown in Figure 4.37. In agreement with the observation on the current overshoot, the 1R configuration exhibits lower LRS with larger variability, compared to the 1T-1R configuration. The highest current overshoot reported corresponds to the lowest LRS even at 0.9 kΩ, which is more than 5 times smaller than the median value. Since the RESET process starts from the LRS, the variability of LRS can be transmitted to the HRS through the RESET process under a constant $V_{RESET-stop}$. Therefore, a much larger HRS variability was observed from the 1R configuration.

As a conclusion, the variability of current overshoot directly reflects on the LRS variability and subsequently affects the HRS variability [132, 133]. In order to obtain better control on current overshoot, the 1T-1R configuration is more promising for consistent and reliable resistive switching behavior [130, 136].

4.5.3 2-bit multi-level switching by gate voltage modulation

Multi-level switching can be demonstrated in a 1T-1R unit cell by modulating V_g. Switching cycles including SET and RESET processes are shown with the electro-forming step in Figure 4.39, where four different $V_{g,SET}$ were used from 1 V to 2 V, while $V_{g,RESET}$ fixed at 2.5 V. As the V_g during electro-forming ($V_{g,FORM}$) at 0.75 V was applied, the equivalent I_{CC} was around 65 μA. By increasing $V_{g,SET}$, the higher current output of transistor with a higher saturation current allows a higher current passing through the RRAM device during the actual transition

process from HRS to LRS. Similar to increasing I_{CC}, it results in a lower LRS, followed by a larger $I_{RESET,max}$ and a higher $V_{RESET,C}$, as shown in Figure 4.39. To conclude, multiple LRS modulated by $V_{g,SET}$ was achieved for multi-level switching or the concept of 2-bit per cell.

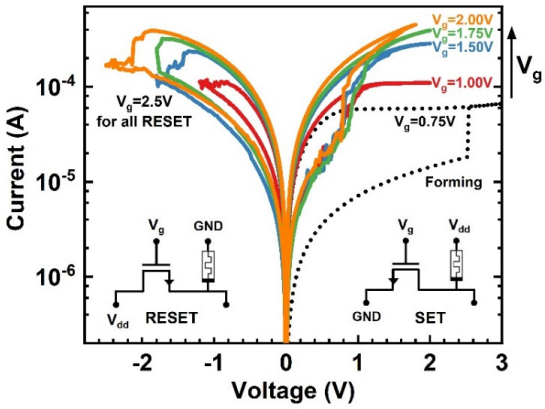

Figure 4.39: *2-bit multi-level switching of the integrated TaO_x RRAM in 1T-1R configuration. Two different testing schemes were used for the SET and RESET processes respectively, as illustrated by the inserted circuit schematics. The multi-level switching was enabled by the $V_{g,SET}$ modulation ranging from 1 V to 4 V. The $V_{g,RESET}$ was 2.5 V for all the RESET processes. The initial forming process under the $V_{g,FORM}$ of 0.75 V is also displayed. Reprinted from [137].*

4.6 Computing-in-memory: IMP stateful logic

According to the previous results and discussions of electrical characterization, the transistor can not only be a selecting component in active RRAM arrays but also provide a better control on current overshoot in the RRAM device and a possibility of multi-level switching by $V_{g,SET}$ modulation. Taking advantages of these features, stateful logic was demonstrated and studied in active RRAM arrays as CIM implementation.

The stateful logic is a novel concept that utilizes resistance instead of voltage and charge as the physical state variable, which was proposed and firstly demonstrated by Borghetti et al. in 2010 via material implication (IMP) [138]. Unlike conventional logic gates relying on transistors, stateful logic gates are built up by memristors such as RRAMs incorporated within an appropriate circuit. The memristors in a stateful logic gate performs the desired logic operation with the input directly given by their resistance states and subsequently the output again stored as their resistance states. Simultaneously, the same memristors are employed as the memory unit to store input/output data and as the processing unit to perform logic computation, which makes the stateful logic a perfect example of CIM that literally executes computation on memory devices where data can be stored instantly. Without actual data movements, it saves a substantial part of electric power required for operating. Plus, RRAM is a non-volatile memory featuring low power consumption among others. Collectively, the implementation of stateful logic creates a new building block for an energy-efficient computing architecture.

Normally, a stateful logic gate is realized by a simple circuit shown in Figure 4.40a, where multiple memristors incorporate with their bottom electrodes connecting together to a fixed

resistor in series. Different voltages can be biased individually from the top electrodes of different memristors. Depending on the type of logic, different amount of memristors, different voltage scheme, and different series resistance are used [139-141]. The active RRAM array in 1T-nR configuration fabricated previously in this work provides a possibility to replace the series resistor normally in the circuit of stateful logic gate by a transistor, as shown in Figure 4.40b. The output of transistor is equivalent to a variable resistor modulated by V_g, which allows a spontaneous control on the series resistance according to the desired logic. Moreover, the transistor provides a better control of current overshoot in RRAM devices. The data can, consequently, be written into RRAM devices as their resistance states with a reduced variability, either when a stateful logic is performed or when the RRAM devices are written for assigning input data. Based on the desired logic and the voltage scheme in use, different HRS and LRS can be chosen respectively by using different V_{RESET} and $V_{g,SET}$. Overall, a single 1T-nR memrsitive array can provide various possibilities on operating different stateful logics, and the resulting flexibility can enable comprehensive computations via cascading logic.

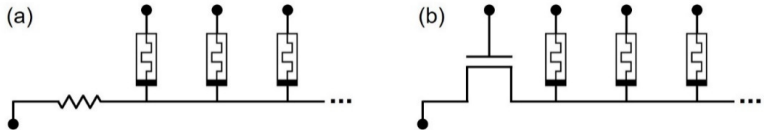

Figure 4.40: *Circuit schematics of memristor-based stateful logic gates. (a) Conventional stateful logic gate using memristors and a fixed resistor. (b) Novel stateful logic gate using memristors and a transistor.*

4.6.1 Realization of IMP stateful logic in 1T-nR arrays

In this work, the IMP stateful logic was realized in the 1T-nR array to verify the validity of using a transistor to replace the fixed resistor in a conventional stateful logic gate. It can be found in Figure 4.40 that a typical stateful logic gate is in the form of RRAM line arrays that have only one word line with multiple bit lines. Among the designed 1T-nR arrays, line arrays ranging from 1×8 to 1×30 were used to implement the IMP stateful logic. The fabricated 1T-15R line array is shown in Figure 4.41 as an example. For probing during testing, all open terminals including the drain/source/gate of transistor and the top electrodes of 15 RRAM cells were connected to larger metal pads that cannot be shown in the same SEM image.

A truth table of the IMP logic operation is shown in Table 4.1 including all possible input cases with corresponding outputs. There are two inputs (p and q) and one output, named q' because the result of pIMPq is assigned back to the q that is overwritten as the q'. As a result,

Table 4.1: *Truth table of IMP logic operation*

IMP logic operation

Case	Input		Output
	p	q	q'
1	1	1	1
2	1	0	0
3	0	1	1
4	0	0	1

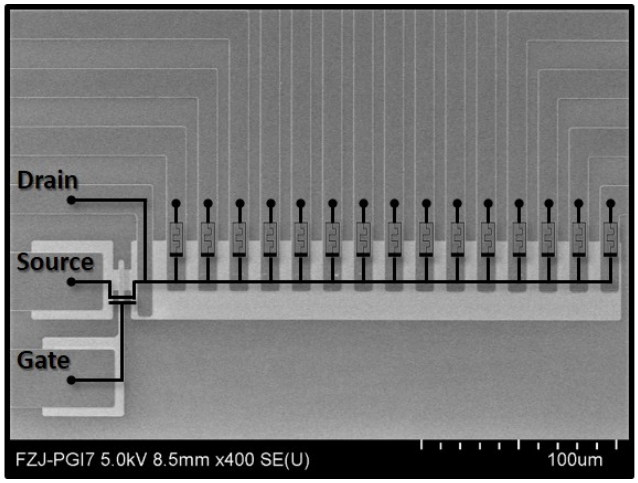

Figure 4.41: *A stateful logic gate based on the fabricated 1T-15R line array. The equivalent circuit and the arrangement of metal connection are sketched on top of the SEM image. Reprinted from [142].*

only two bits, namely *p* and *q* bits, are required to fulfill the task of giving the inputs (*p* and *q*) and storing the output (*q'*). Practically in the IMP stateful logic gate, two RRAM cells were employed as the *p* and *q* bits, resulting in a 1T-2R line array, as shown in Figure 4.42. The binary data is memorized as the non-volatile resistance state of RRAM cell that LRS and HRS represents logic 1 and logic 0 respectively. For example in Table 4.1, input case 1, (*p*/*q*=1/1), describes that the *p* and *q* bits are both in the LRS initially. After the IMP logic operation, the *q* bit stays in the LRS, indicating *q'* equals 1. Notably, a successful logic operation is assumed that $p' = p$ because there is no new value assigned to the *p* bit. In other words, the same resistance state should be retained in the *p* bit after the logic operation. By taking a closer look in Table 4.1, only input case 4, (*p*/*q*=0/0), leads to a bit flip on the *q* bit after the logic operation, which is essentially a SET process, bringing the *q* bit from HRS to LRS. Therefore, it is considered as a SET-based stateful logic gate, which operates by triggering SET processes on certain bits under corresponding input conditions.

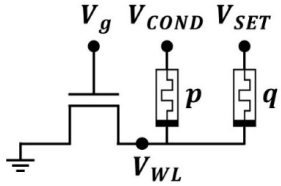

Figure 4.42: *IMP stateful logic gates using two RRAMs and a transistor. Each RRAM is regarded as one bit that stores a binary data by its resistance state. Corresponding to the p and q values used in the IMP logic, the RRAM biased by the V_{COND} is the p bit, while the RRAM biased by the V_{SET} is the q bit. The V_{WL} is a floating electric potential on the common word line, which is a shared node of the drain of transistor and BEs of the p and q bits.*

To operate the IMP logic gate, it requires three different operating voltages including the gate voltage on the transistor (V_g), the condition voltage on the *p* bit (V_{COND}), and the SET

voltage on the q bit (V_{SET}), as shown in Figure 4.42. For a single logic operation, they are applied only once via square pulses. In principle, a valid logic gate can process all input cases under an identical operating condition. Hence, the same set of operating voltages is used no matter which inputs are given. The only operating variables appear to be the resistances of p and q bits (R_p and R_q) besides the channel resistance of transistor depending on the actual circuit dynamics. With given V_g, V_{COND}, and V_{SET}, the electric potential on the common word line (V_{WL}) is determined by the R_p and R_q. The voltages across p and q bits (V_p and V_q), derived from the floating V_{WL} and the given V_{COND} and V_{SET}, are critical for resistive switching. Taking the input case (p/q=0/0) as an example, the V_q must exceed the $V_{SET,C}$ to trigger the SET process, bringing the q bit to the LRS, while the V_p needs to be lower than the $V_{SET,C}$ to prevent the SET process, retaining the HRS in the p bit. Therefore, the floating V_{WL} actually serves as the core logic operator controlling the resistive switching to occur under the correct condition. Experimental and theoretical analysis of stateful logic in passive and active crossbar arrays for computation-in-memory

With HRS and LRS defined according to the actual switching property of RRAM cells, the combination of operating voltages (V_g, V_{COND}, and V_{SET}) is the key to allowing the V_{WL} properly control the V_p and V_q during logic operations. To begin with the discussion on the V_{COND} and V_{SET}, the V_g is assumed as a constant. For the input case (p/q=0/0), which contains a bit flip, the V_{SET} must be larger than V_{COND} to fulfill $V_q > V_{SET,C} > V_p$. However, too high V_{SET} and too low V_{COND} are not favorable for the input case (p/q=1/0). Moving from the input case (p/q=0/0) to (p/q=1/0), the input p bit changes from the HRS to the LRS, causing an increase of V_{WL} according to the change in voltage divider configuration. To keep $q' = q = 0$ by fulfilling $V_q < V_{SET,C}$, the increase of V_{WL} favorably leads to a decrease of V_q, which can prevent the SET process on the q bit and hold the HRS as the correct output. Nevertheless, if the V_{SET} is too high, the V_q could be still larger than $V_{SET,C}$. On the other hand, V_q can be raised by decreasing V_{COND} due to a decrease in V_{WL}. If the V_{COND} is too low, V_q could also be larger than $V_{SET,C}$. The study on V_{SET} and the V_{COND}/V_{SET} ratio can be referred to a couple of published works [141-144]. As a benefit from the use of transistor, when the p bit is in the LRS, the V_{WL} is considerably close to the V_{COND} such that the transistor can take over the major portion of voltage drop to limit the loading current. As a beneficial consequence, the upper limit of V_{SET} and the lower limit of V_{COND} are both released.

Since the series resistance (R_{series}) in the stateful logic gate and the R_{LRS} resulted from SET processes are simultaneously determined by the V_g, both need to be taken into considerations for the V_g selection. Normally, the R_{series} must be chosen such that $R_{HRS} > R_{series} > R_{LRS}$ [138]. When the p or q bit is in LRS, the series resistor can take over a major voltage drop if $R_{series} > R_{LRS}$. The resulting increase in V_{WL} can favorably prevent the SET process on the other bit. On the other hand, for the input case (p/q=0/0), a major voltage drop can appear across the q bit, desirably triggering the SET process, if $R_{HRS} > R_{series}$. With the use of transistor, $R_{HRS} > R_{series} > R_{LRS}$ can be easily fulfilled. When one of the bits is in the LRS, the transistor can work in the saturation region, where the channel resistance significantly increases, to hinder further current increase, which naturally fulfills $R_{series} > R_{LRS}$. For the input case (p/q=0/0), where the two bits are both in the HRS, the transistor can work in the linear region, referring to a low channel resistance that can easily fulfill $R_{HRS} > R_{series}$.

When the R_{HRS} was defined within a range of 90 kΩ to 140 kΩ in this work, 1 V was chosen as the V_g giving the channel resistance of 3.3 kΩ in the linear region and the R_{LRS} of around 10 kΩ, referring to an ON/OFF ratio above 9. In addition, the V_{SET} of 1.1 V with a V_{COND}/V_{SET} ratio of 0.78 was selected according to the relevant published work [142]. To perform the IMP logic operation, the V_{COND} of 0.86 V, the V_{SET} of 1.1 V and the V_g of 1 V were applied

accordingly to the stateful logic gate, as the actual pulse scheme is sketched in Figure 4.43. The V_{SET} and V_{COND} have a pulse length of 100 µs while the V_g has a longer pulse length of 300 µs, allowing the transistor to work in a stable condition already before the V_{SET} and V_{COND} are applied on the RRAM cells.

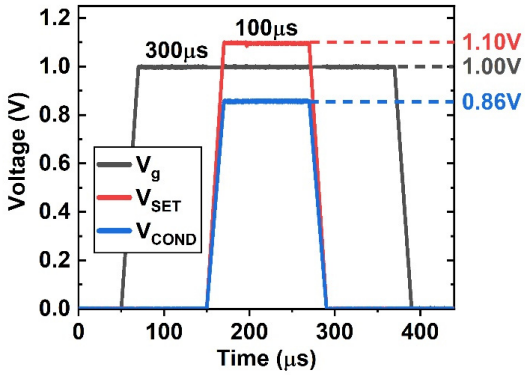

Figure 4.43: *Voltage pulse scheme for the proposed IMP stateful logic gate. The actual logic operation takes place during the 100 µs when the V_{COND} and the V_{SET} are applied. The V_g is applied for 300 µs to prevent the actual logic operation influenced by switching on and off the transistor.*

The logic operation result of different input cases is shown in Figure 4.44, where the transient currents of p and q bit are plotted together in a blue (I_p) and a red (I_q) curves respectively. For the input case 1, 2, and 3, no current change within the pulse duration was found, indicating no switching in resistive states. In line with the truth table of IMP logic operation, the outputs remain equal to the inputs. For the input case 1, I_p was found negative because V_{WL} was raised

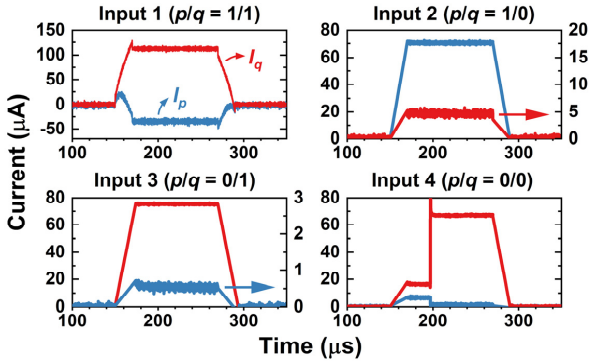

Figure 4.44: *Current readouts from the p and q bits during IMP logic operations of four different input cases. While applying V_{COND} to the p bit and the V_{SET} to the q bit, I_p, colored in blue, and I_q, colored in red, are recorded during the IMP logic operation. For the input cases 1, 2, and 3, the p and q values are kept after logic operation because no resistance change is noticed from any change in I_p and I_q. For the input case 4, an abrupt I_q increase indicates a transition towards the LRS taking place in the q bit, causing a change in the q value from 0 to 1. Reprinted from [142].*

above V_{COND}. When the p and q bit are both in LRS, high currents collected on the common word line shift the transistor output to the saturation region, allowing the transistor to take over a major voltage drop. The resulting V_{WL} increase limits the I_q at around 100 μA and causes the I_p flowing in the opposite direction. For the input case 4, a distinct current change within the logic pulse duration was observed both for the p and q bits. The q bit switched to the LRS within 40 μs, causing a huge jump in the I_q from one stable state to another. Meanwhile, the I_p decreased because of the V_{WL} increase resulted from the resistance change in the q bit. Eventually, the p bit remained in HRS while the q bit switched to LRS, leading to the correct logic output (p'/q'=0/1). Moreover, it can be noticed that the output (p'/q'=0/1) in the input case 4 shows the I_p and I_q comparable to the input (p/q=0/1) in the input case 3. In summary, the proposed IMP stateful logic gate in the 1T-2R configuration was experimentally verified in the fabricated 1T-nR line array with an appropriate combination of the operating voltages.

4.6.2 Impact of device-to-device variability on IMP stateful logic

In the IMP stateful logic gate, there are two RRAM devices among which one works as the p bit while the other works as the q bit. The V_{COND} is supposed to be applied on the p bit whereas the V_{SET} on the q bit. Depending on which cells the V_{COND} and V_{SET} are applied respectively to, the p and q bit are automatically assigned. For example, an IMP stateful logic gate sketched in Figure 4.45 can technically has the cell A or the cell B as the q bit, depending on which cell the V_{SET} is applied to. As the device-to-device (D2D) variability between the cell A and B comes into play, having different cell as the q bit can influence the outcome of the IMP stateful logic gate.

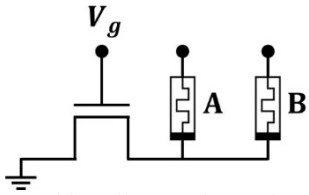

Figure 4.45: *Interchangeable cells A and B in the IMP stateful logic gate.*

In collaboration with Dr. Christopher Bengel, a simulation study was carried out using the physics-based compact model *JART VCM v1b var*, which is capable of reproducing a comprehensive switching variability of VCM-type RRAM devices for a variety of measurements [142, 145]. By introducing the D2D variability to the RRAM cells employed in the IMP stateful logic gate, the corresponding performance was simulated and evaluated in terms of stable input and correct output, as shown in Figure 4.46. The same set of operating voltages (V_{SET}=1.1 V, V_{COND}=0.86 V, V_g=1 V) was used in the simulation. The stable input represents the result that the p bit retains its logic value after a logic operation because there is no output value assigning back to the p bit. The correct output refers to the result that the q bit stores the correct logic value after a logic operation. For each level of D2D variability, 100 random combinations of simulated devices were assigned as the p and q bits for every input case. The successful incidents of stable inputs and correct outputs were calculated as success rates. It can be found from the most critical input case, (p/q=0/0), that a higher D2D variability leads to a lower success rate. Hence, a noticeable impact on performance decline from the D2D variability was verified.

In this context, the D2D variability in the fabricated 1T-nR array needs to be eliminated in order to pursue a superior performance. However, D2D variability is an inherent issue of

RRAM technology especially due to the stochastic nature of switching mechanism for filamentary-type RRAMs. A direct way to advancement is apparently to reduce the D2D variability down to an acceptable level. From an opposite perspective instead, an advancement can also come from a beneficial exploitation of the D2D variability. If some particular RRAM applications can benefit from the D2D variability, then it is considered as a favorable feature instead of a shortcoming. In this regard, an experimental study was conducted on proposing a possible approach of utilizing D2D variability in the IMP stateful logic gate.

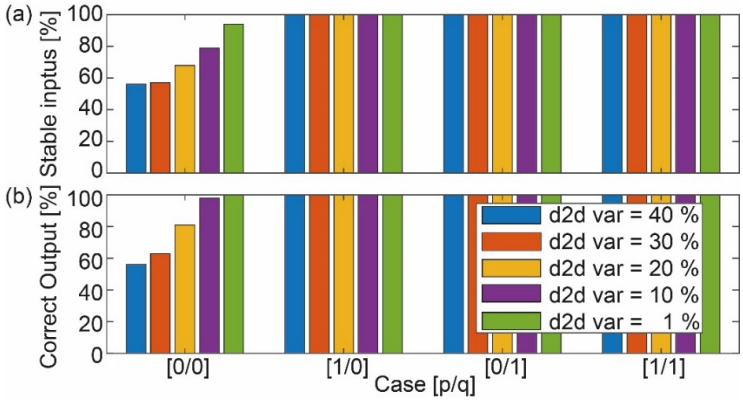

Figure 4.46: *Simulation result on the impact of D2D variability towards IMP stateful logic operation accuracy. Different amounts of D2D variability were introduced into the simulation model to study their impacts for four different input cases. Each scenario includes 100 combinations of random simulated RRAM devices. (a) Probability of stable input after logic operation. Only p value is of the concern, because no output value is assigned to p after logic operation, implying the p value should stay constant. (b) Probability of correct output after logic operation. Only q value is of the concern, because the only output value is assigned to q after logic operation. Reprinted from [142].*

As one of SET-based stateful logic gates, the IMP stateful logic gate operates by triggering SET processes correspondingly under certain input conditions. Thus, the D2D variability in SET process was of the major study interest. It has been learned that the SET process is triggered when the applied V_{SET} exceeds the $V_{SET,C}$. The D2D variability of $V_{SET,C}$ was investigated between two RRAM cells in an IMP stateful logic gate. Controlling the I_{CC} at 100 µA and the R_{HRS} at 120 kΩ, the cell A and B individually performed 50 SET processes in a quasi-static manner by DC voltage sweep. The recorded $V_{SET,C}$ of each SET operation were analyzed by a histogram shown in Figure 4.47 to depict the probability density function. Described by a normal distribution or Gaussian distribution, the $V_{SET,C}$ shows a mean difference ($\Delta V_{SET,C}$) of 0.186 V between the two cells, suggesting that the cell A requires a lower V_{SET} to trigger a SET prcoess.

To benefit from the existing D2D variability of the $V_{SET,C}$, one possible approach was proposed here. As the only bit flip after a logic operation appears in the input case (p/q=0/0), where the q bit switches from the HRS to the LRS through a SET process, an absolute advantage can be obtained when the cell with a lower $V_{SET,C}$ is assigned as the q bit. In this case, $V_q > V_{SET,C}$ can be fulfilled more easily to trigger the SET process desirably in the q bit in comparison with the opposite case. An experimental verification was carried out correspondingly that IMP logic with the input (p/q=0/0) was operated 50 times, when the cell A and B were respectively assigned as the q bit and the same set of operating voltages (V_{SET}=1.1

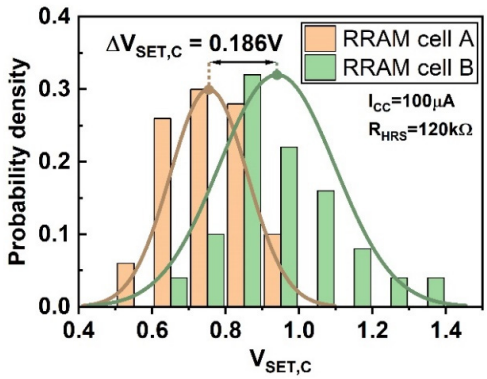

Figure 4.47: *D2D variability of $V_{SET,C}$ in RRAM cells A and B. The statistic is from the experimental data of 50 SET processes for each RRAM cell, when I_{CC} at 100 μA was applied and the starting HRS was fixed at 120 kΩ. A mean difference of 0.186 V in $V_{SET,C}$ was determined. Reprinted from [142].*

V, V_{COND}=0.86 V, V_g=1 V) was used. A flow chart of the programmed measurement is shown in Figure 4.48. Literally, every IMP logic operation is performed after the resistances of p and q bits are evaluated and programmed to a defined HRS within a range from 90 kΩ to 140 kΩ.

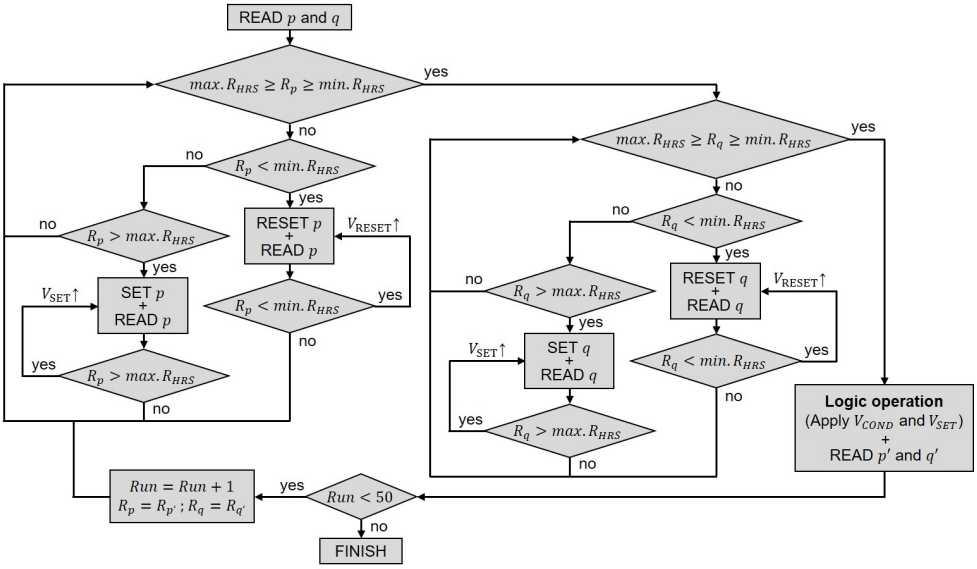

Figure 4.48: *Programmed measurement flow chart for success rate assessment on IMP stateful logic operation of the input case (p/q=0/0). 50 attempts of logic operations in one measurement is shown as an example.*

In Figure 4.49, the actual resistances of the p and q bits before and after the IMP logic operations total are plotted over 50 attempts under the input case (p/q=0/0). When the cell A, showing the lower $V_{SET,C}$, was assigned as the q bit, the output p' showed resistances in the

defined HRS range, which points out the fact that the *p* bit correctly kept the logic 0 after the logic operations. Besides, 68 % of the output *q'* showed resistances around 10 kΩ, indicating the correct output (*q'*=1) as a result of the successful resistive switching from the HRS to the LRS. Compared the overall success rate of 68 % to the case that the cell B, showing the higher $V_{SET,C}$, was assigned as the *q* bit, a pronounced difference was observed. The resistances of the output *p'* had a wide spread, where few cases falling below the HRS range. Importantly, the *q* bit consistently stayed in the HRS after the logic operations, indicating the wrong output (*q'*=0). The overall success rate of 0 % clearly concludes the significant impact of the D2D variability in $V_{SET,C}$ when assigning the *p* and *q* bits.

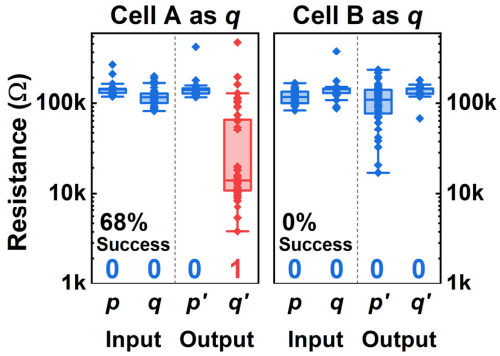

Figure 4.49: *Success rate difference of the cell A and B serving as the q bit in IMP stateful logic operations of the input case (p/q=0/0) based on experiment data. Reprinted from [142].*

In contrast to the random assignment of the *p* and *q* bits in the IMP stateful logic gate, the use of D2D variability can lead to a better performance by intentionally assigning the RRAM cell of lower $V_{SET,C}$ as the *q* bit. Depending on usage scenarios and application context, it is possible to exploit the D2D variability in the RRAM array rather than merely pursuing ultra-low D2D variability.

Chapter 5
Monolithic integration of nanoscale RRAMs with CMOS circuitry for neuromorphic computing

In the era of AI, data explosion presents significant challenges for the speed and energy efficiency of modern computing technology. To improve the AI computing, there is a growing need for computing accelerators. Alongside the powerful AI applications enabled by high-performance computing in centralized architectures, applications in 5G, internet of things, and autonomous driving drive a rising trend in edge AI, where data is analyzed in edge devices where the data is collected. As the edge devices have relatively limited power resources, the energy efficiency of AI accelerators becomes more demanding than the speed. In collaboration with PGI-14 *Neuromorphic Compute Nodes* in Research Center Jülich and Hewlett Packard Labs, this study focuses on the development of an RRAM technology platform for enabling a brain-inspired energy-efficient AI accelerator called the dot product engine (DPE) [146, 147].

The platform is aimed for accelerating matrix operations, allowing a performance boost in matrix-heavy applications such as speech recognition, image and video classifications. Within the framework of neuromorphic computing architecture, the DPE not only benefits from the use of RRAM, a low-power analog memory device with non-volatility, but also from the concept of CIM, where computation takes place directly in memory without data movement. Physically, the DPE is a 1T-1R array working as a synaptic array that has the neural network

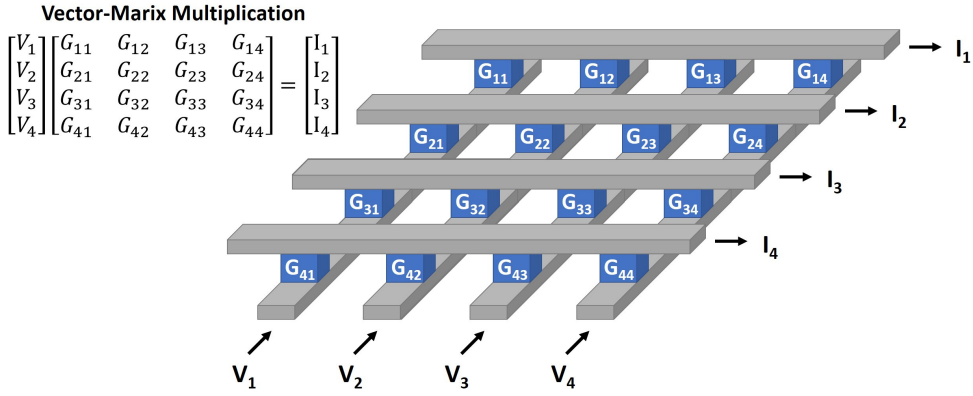

Figure 5.1: *Schematic diagram of vector-matrix multiplication intrinsically operated in an RRAM array. The voltage vector input enters from each row. Through the conductance matrix of RRAMs in the array, the current collecting from each column forms a vector output as the multiplication result of the voltage vector and the conductance matrix.*

weight stored as the RRAM conductance. By electrically reading the array and utilizing Ohm's law and Kirchhoff's law, it can directly operate vector-matrix multiplication (VMM) with analog-voltage-amplitude vector and analog-conductance matrix, enabling parallel analog computing in the array, as illustrated in Figure 5.1. To drive and sense the analog circuit of RRAM array, digital control and analog sensing circuits are integrated on chip such that digital-to-analog converters can drive digital voltage input to the array and analog-to-digital converters can deliver digital results from sensing the current output. Together with on-chip signal amplifiers, the driving/sensing circuitry and N-channel metal-oxide-semiconductor (NMOS) transistors in the front-end-of-line (FEOL) for 1T-1R arrays were taped out with TSMC 180 nm technology node. Subsequently, this work involved monolithic integration of nanoscale RRAM devices on the topmost metal vias passivated in the foundry BEOL process, followed by electrical testing, packaging, and chip verification on a customized operating hardware system.

5.1 Layout design for RRAM integration

The design of 180 nm CMOS dies taped out at TSMC, as shown in Figure 5.2, includes three 64×64 1T-1R arrays for three individual DPEs. Connecting to the layer Metal 4, the W vias at the topmost layer are arranged in a group of two, of which one must connect to the BE of an RRAM, while the other to the TE. Eventually, the BE is connected to the drain terminal of an NMOS transistor in the FEOL, whereas the TE to a signal input/output (I/O) terminal.

The strategy of RRAM integration in the BEOL is illustrated in Figure 5.3, which describes how a nanoscale RRAM device in a crossbar structure was designed to connect its top and bottom electrodes to the corresponding vias. The BE pattern covers the corresponding W via and gives a thin line with a 100-nm width. In the same patterning step, the other W via is covered by a rectangular pattern, which serves as a metal pad. The TE pattern also has a thin line with a 100-nm width crossing the BE from the top, but it does not cover the corresponding W via because the resistive-switching metal oxide (MO) layer is patterned together with the TE, following the device fabrication technology established in PGI-7. Thus, a metal connection is required to link the TE to the W via.

The RRAM integration in the BEOL requires an alignment with respect to the CMOS die. To this purpose, alignment marks are designed at four corners of the die with a pattern shown in Figure 5.4 on the right. The marks were produced in the BEOL process by spatially arranging metal vias to form the certain cross shape. For the fabrication of nanoscale RRAMs, EBL employed in this work was manually aligned with these marks because the e-beam writer, from the EBPG series of Raith GmbH, can only execute auto alignment with square marks. The manual alignment was performed by using the built-in SEM function to pinpoint the marks on the die with the corresponding coordinates in the mask design

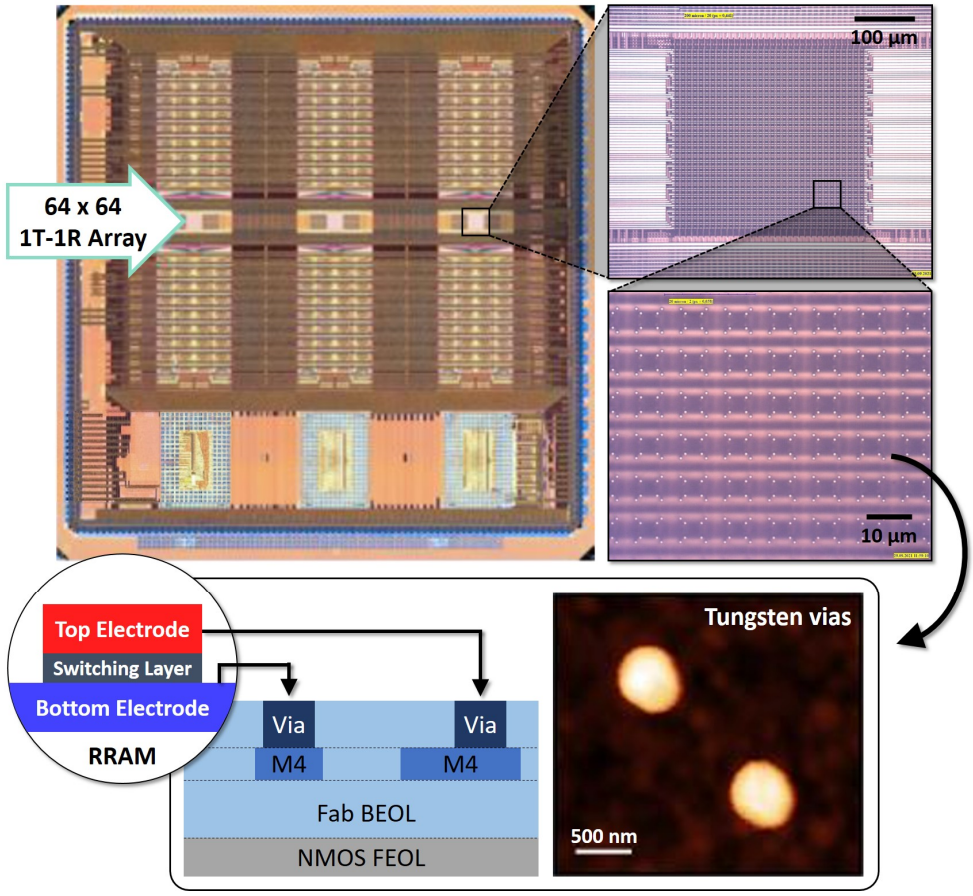

Figure 5.2: *Monolithic integration concept for fabricating RRAMs on the 180 nm CMOS die taped out at TSMC to realize the RRAM-based DPEs. There are three 64×64 1T-1R arrays serving as individual DPEs on one die. The TE and BE of each RRAM are designed to be later connected to the topmost W vias.*

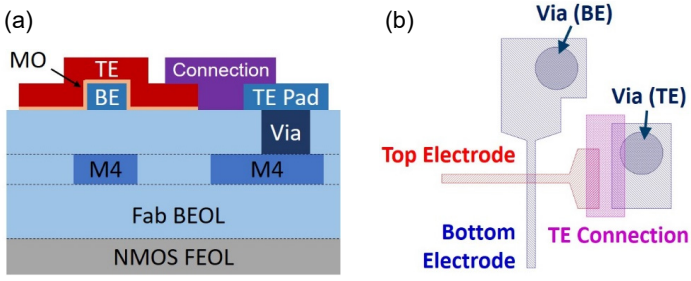

Figure 5.3: *Structure schematics of the integrated RRAM on the CMOS die in (a) a cross-sectional view and (b) a top views*

.

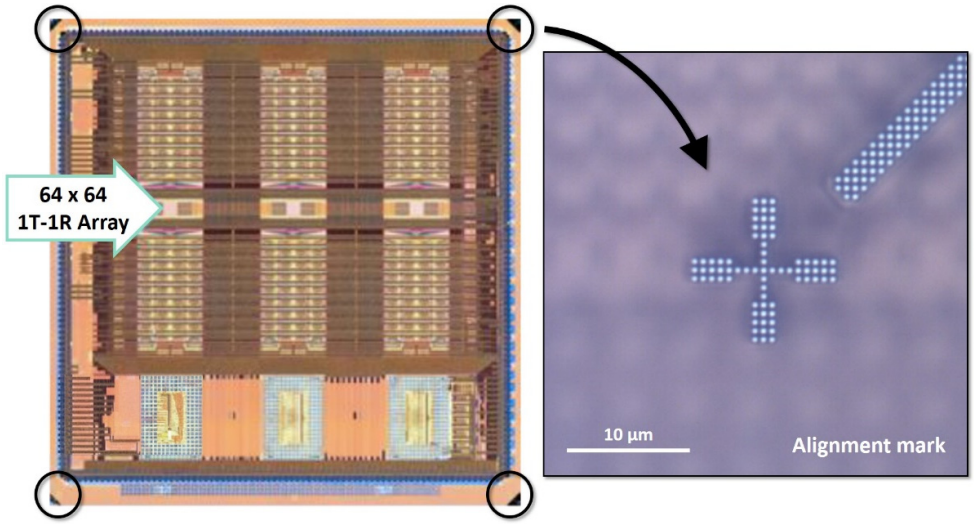

Figure 5.4: *EBL alignment marks available on the CMOS die.*

5.2 Process flow of RRAM integration

With the physical layout design established, a process flow has been developed to implement the monolithic integration of RRAMs by using the micro- and nanoelectronics fabrication facilities in PGI-7 and HNF at Research Center Jülich. Starting from a CMOS die, Figure 5.5 summarizes the entire process flow until it is fully functional with the integrated RRAMs. The RRAM integration result is shown in Figure 5.6. The OM images display the top view of an integrated 1T-1R array, where an integrated RRAM with the device size of 100×100 nm^2 is shown in the SEM images.

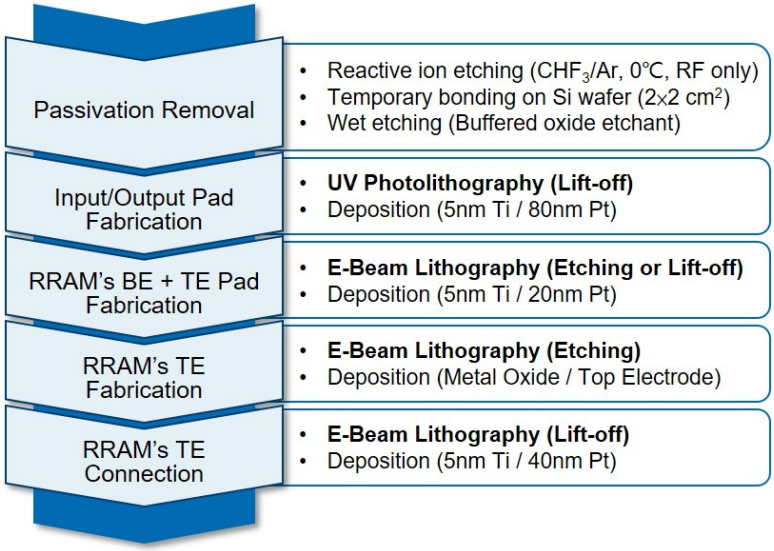

Figure 5.5: *Process flow of RRAM integration in the BEOL on the CMOS die.*

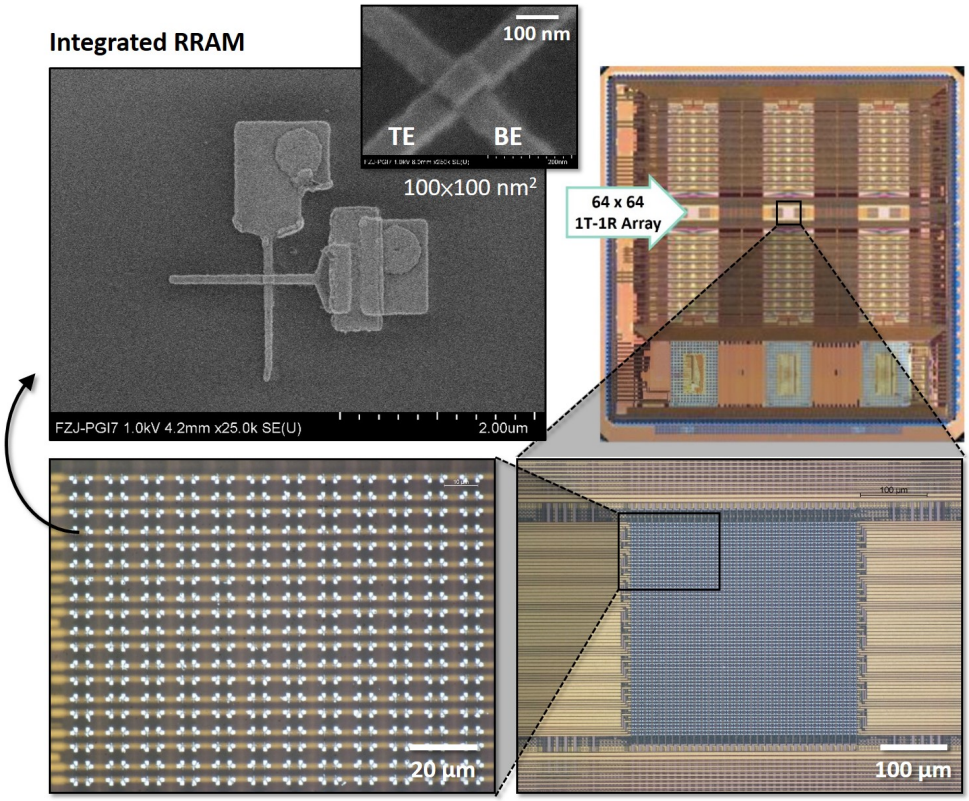

Figure 5.6: *OM and SEM images of BEOL-integrated RRAMs on the CMOS die.*

In the following part of this subchapter, each fabrication step will be introduced chronologically in detail.

5.2.1 Passivation removal

As the CMOS dies were passivated in the foundry, the passivation layer, comprising a 0.6-µm-thick Si_3N_4 layer on a 1.15-µm-thick SiO_2 layer, must be removed to expose the topmost W vias on the surface for the subsequent RRAM integration process. In addition, a via protrusion of 30 - 60 nm was achevied to ensure the adequate contact surface. The passivation

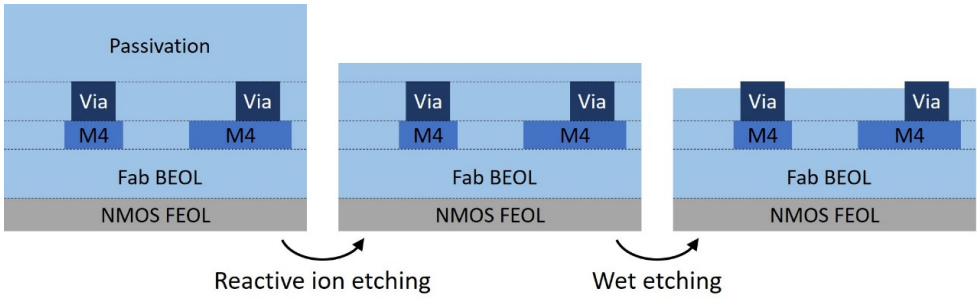

Figure 5.7: *Schematic diagram of passivation removal process*

removal process involved a RIE step and a wet etching step, as illustrated in Figure 5.7. Using a gas mixture of CHF_3 and Ar, the RIE step was carried out at 0°C to uniformly etch through the Si_3N_4 layer until roughly 200-nm-thick SiO_2 remained, which prevents the potential damages on the W vias caused by the plasma. Before the wet etching step, the CMOS die was temporarily bonded to a carrier substrate to facilitate handling during the wet etching and subsequent fabrication steps.

A liquid adhesive *WaferBOND® HT-10.11*, dedicated to the use in BEOL processes up to 220°C, was utilized as the temporary bonding material. After applying a small amount onto a clean Si substrate of 2×2 cm^2 that had been dehydrated on a hot plate at 120°C for 5 minutes, the CMOS die was placed directly onto the adhesive. Subsequently, hot plates were employed for a two-step heat treatment, where 5 minutes at 120°C for were followed by 10 minutes at 180°C, during which a gentle press was applied shortly on the CMOS die to evenly distribute the pre-heated adhesive and establish a firm contact. At last, cooling on a metal surface ensued at room temperature for 1 minute. The result of a bonded die is depicted in Figure 5.8.

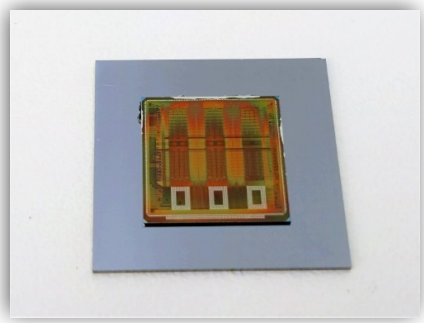

Figure 5.8: *CMOS die fixed on a Si carrier substrate of 2×2 cm^2 by temporary bonding.*

With the CMOS die boned on the carrier substrate, the wet etching step was performed in a buffered oxide etchant (BOE), which is a solution of buffered hydrofluoric acid solution that has a ratio of HF : NH_4F = 1 : 7, at room temperature for approximately 80 seconds. To ensure a via protrusion of 30 - 60 nm, additional wet etching for a few tens of seconds might be necessary depending on the result of each die, which can be evaluated by AFM. An example illustrating the desired outcome is displayed in Figure 5.9, where the results of line profile scan and morphology 3D mapping from AFM confirms a via protrusion of approximately 40 nm.

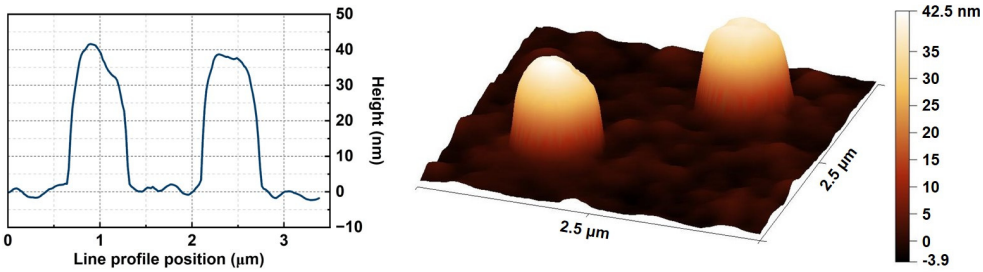

Figure 5.9: *AFM line profile and morphology 3D mapping of protruding W vias after passivation removal. The line profile is extracted directly from the 3D mapping data of the same W vias.*

5.2.2 Input/output pad fabrication

Once the passivation layer was carefully removed, contact pads were formed on the I/O ports of the CMOS die. The physical layout of the I/O pads is illustrated in the upper part of Figure 5.10. In the UV photolithography to pattern the I/O pads with areas of 70×70 μm^2, the *MLA 100* maskless aligner was employed with the manual exposure alignment conducted at the four corners indicated by red arrows in Figure 5.10. Taking the bottom right spot as an example, the OM image on the bottom right shows a square array of metal vias, where the I/O pad should be formed. Practically, it was used as the alignment mark for exposure such that its center was pinpointed with the corresponding coordinates in the mask design. Based on the aligned coordinates, the exposure was then carried out in a direct writing approach.

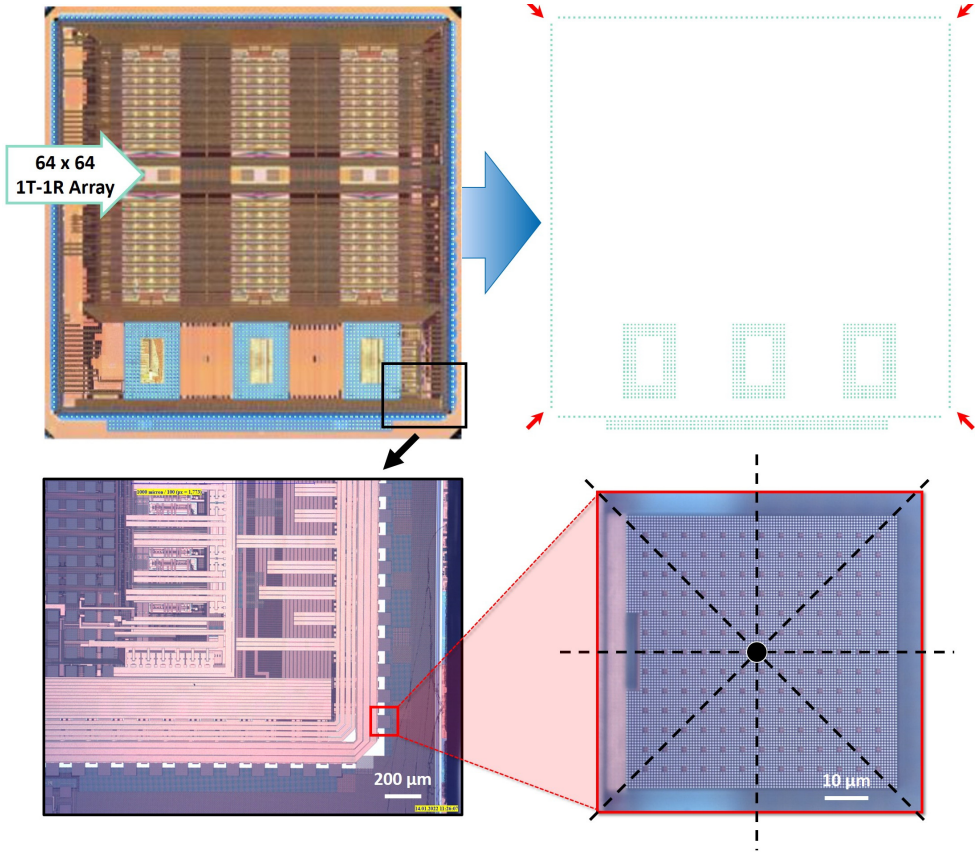

Figure 5.10: *Locations and alignment strategy for fabricating the I/O pads.*

The fabrication of I/O pads involves a UV photolithography, a thin-film deposition, and a lift-off process. Firstly, UV photolithography was conducted using the negative photoresist *AZ® nLOF™ 2020* and the maskless aligner for exposure, transferring the desired pattern to the photoresist as a hollow pattern. To remove the native oxide on the surface of W vias, a 5-second dip in 1% HF solution and an in-situ Ar sputtering were carried out prior to metal deposition. Next, a 5-nm-thick Ti adhesion and an 80-nm-thick Pt layer were deposited via thermal evaporation. Finally, the lift-off step was performed in the *TechniStrip® NI555*, a photoresist remover dedicated to the *AZ® nLOF™ 2020*, for 2 hours. The result is shown in Figure 5.11.

The use of an 80-nm-thick Pt layer is intended for wire bonding to a lead frame during the packaging stage, which enables chip-level operations on a customized operating hardware.

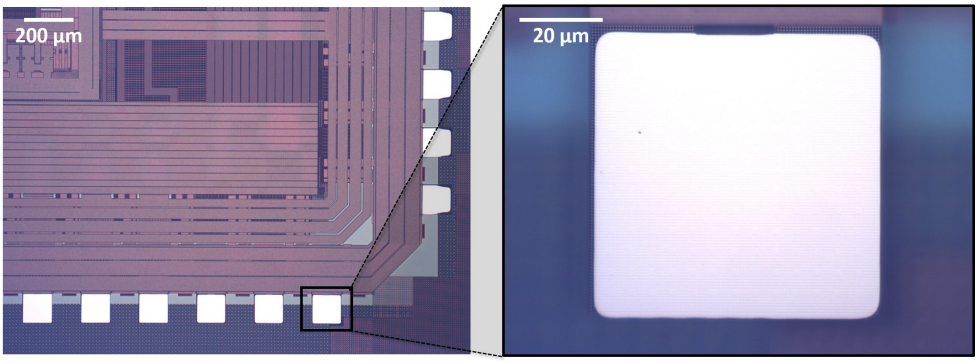

Figure 5.11: *OM images of fabricated I/O pads on the CMOS die.*

5.2.3 Fabrication of RRAM's bottom electrode and top electrode pad

Fabrication of nanoscale RRAM devices started after the I/O pad fabrication. As discussed in Subchapter 5.1, RRAM's BE is connected to one of two vias, while the TE to the other. In a single patterning step, larg metal pads were formed on both vias, and the RRAM's BE was established and connected to the corresponding metal pad, as shown in Figure 5.12. At this stage, the TE pad covers the via to avoid potential damages during subsequent processes, particularly during etching. The cross section perpendicular to the BE direction is illustrated on the right in Figure 5.12, while the cross section through both vias is provided on the left.

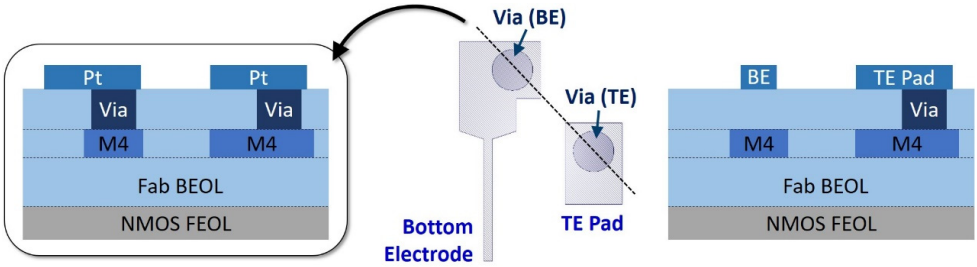

Figure 5.12: *Structure schematic of RRAM's BE and TE pad.*

The BE and TE pad were patterned together by either etching or lift-off. When the etching approach was adopted, a 5-second dip in 1% HF solution was performed at first to remove the native oxide on the surface of W vias, which was immediately followed by the deposition of a 5-nm-thick Ti adhesion layer and a 20-nm-thick Pt layer through in-situ DC sputtering. Next, EBL was conducted on the Pt surface using the negative e-beam resist *AZ® nLOF™ 2020* diluted with the *AZ® EBR Solvent* in a ratio of 1:2. The resist pattern obtained after development was transferred into the metal stack using RIBE with Ar. At last, the resist was stripped off in DMSO at 80 °C for 4 hours, and a moderate O_2 plasma etching followed after to ensure a clean Pt surface without organic residue. The overall patterning result by etching is shown in Figure 5.13a.

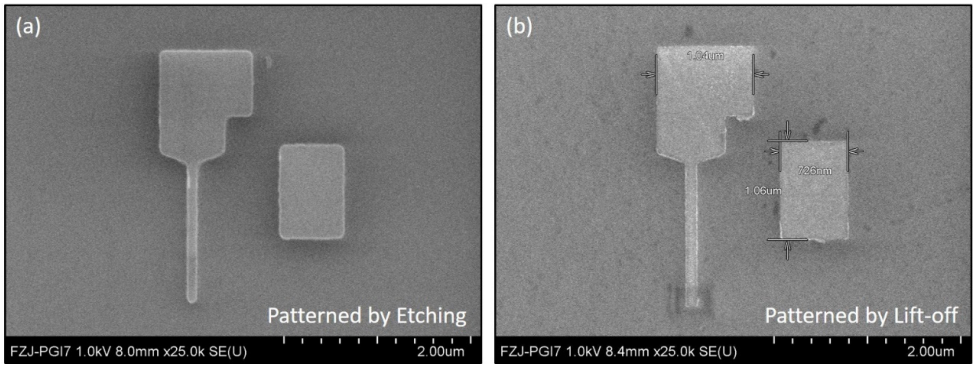

Figure 5.13: *SEM images of fabricated RRAM's BE and TE pad. There are two patterning techniques used alternatively in this fabrication step: (a) etching and (b) lift-off.*

Alternatively, a lift-off approach can be adopted for patterning the metal stack. Prior to the deposition, EBL was performed using the positive e-beam resist *CSAR AR-P 6200.04*. To mitigate the risk of resist residue affecting the lift-off result, a descum step by O_2 plasma etching was then conducted to removed approximately 5 nm of the resist, given the original thickness of around 105 nm. After a 5-second dip in 1% HF solution, deposition of the same metal stack was carried out. Finally, the desired pattern was transferred to the metal stack by a lift-off step in the resist remover *AR-P 600-71*, dedicated to the *AR-P* resists, for 30 minutes. The overall patterning result by lift-off is shown in Figure 5.13b.

Notably, the patterning by etching resulted in the yield below 50 %, which is attributed to a fencing issue during the Pt etching through Ar-based RIBE. Detailed discussion on the fencing issue can be found in Subchapter 5.3. As a result, it is recommended to use the lift-off approach for fabricating RRAM's BE and TE pad, which offers a process yield higher than 90 % and an advantage in process simplicity.

5.2.4 Fabrication of RRAM's top electrode

After RRAM's BE was patterned, the TE was patterned through the MO switching layer to form a RRAM cell at the crossing point of the two patterns. The physical layout design is illustrated from a top view and a cross-sectional view perpendicular to the BE direction, as depicted in Figure 5.14.

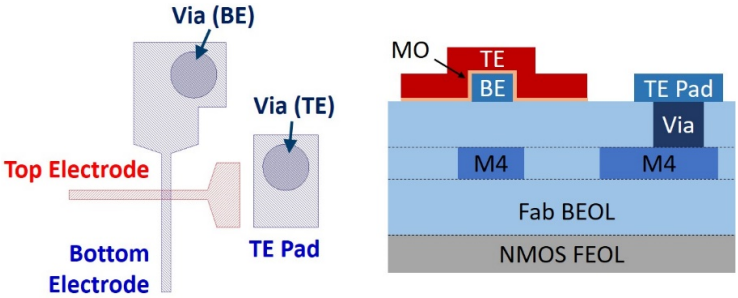

Figure 5.14: *Structure schematic of RRAM's TE.*

To pattern RRAM's TE by etching, the process began with the deposition of MO and TE layers. Depending on the chosen material stack, different deposition methods were employed. For the HfO_2/TiO_x (HOTO) RRAM stack, a 3-nm-thick HfO_2 layer and a 3-nm-thick TiO_x layer were deposited in-situ by ALD as a bilayer MO. Subsequently, a 10-nm-thick Ti layer and a 20-nm-thick Pt layer were grown using in-situ DC sputtering as the TE. In addition, for the TaO_x RRAM stack, a series of in-situ depositions started with a 10-nm-thick TaO_x layer reactively sputtered as a single-layer MO, followed by RF sputtering of a 15-nm-thick Ta layer and DC sputtering of a 15-nm-thick Pt layer. Following the deposition of the MO and TE layers for both RRAM stacks, EBL was performed on the Pt surface using the negative e-beam resist *AZ® nLOF™ 2020* diluted with the *AZ® EBR Solvent* in a ratio of 1:2. The resist pattern obtained after development was transferred into the entire deposited stack typically by RIBE, where Ar was used for the TE while CF_4 for the MO. At last, the resist was stripped off in DMSO at 80 °C for 4 hours. Taking the HOTO RRAM stack as an example, Figure 5.15 displays the fabrication result after resist stripping.

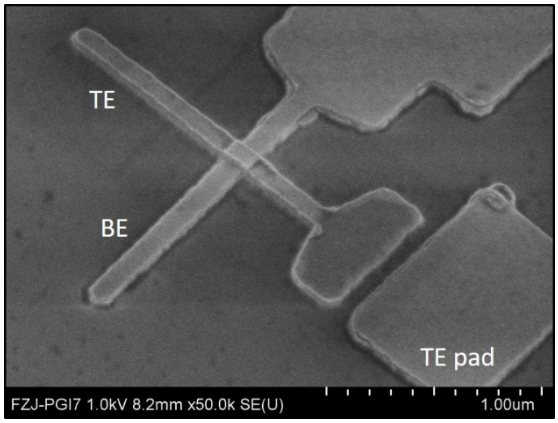

Figure 5.15: *SEM image of RRAM's TE fabrication result.*

Notably, similar to the previous fabrication step for RRAM's BE and TE pad, the Pt etching through Ar-based RIBE during the TE pattering caused a significant fencing issue, leading to aa process yield below 50 %. Detailed discussions can be referred to Subchapter 5.3.

5.2.5 Metal connection for RRAM's top electrode

In the last fabrication step, a rectangular pattern is used to bridge the gap between the TE and TE pad for enabling the metal connection. The physical layout and the cross section perpendicular to the BE direction are displayed in Figure 5.16.

To establish the TE connection by a lift-off approach, a hollow pattern was produced from EBL using the positive e-beam resist *CSAR AR-P 6200.04*. Next, a 5-nm-thick Ti adhesion layer and a 40-nm-thick Pt were deposited through thermal evaporation, and the pattern was then transferred into the metal stack by a lift-off step in the resist remover *AR-P 600-71*, dedicated to the *AR-P* resists, for 30 minutes. The lift-off result is shown in Figure 5.17, demonstrating the seamless metal connection bridging the gap between the TE and TE pad.

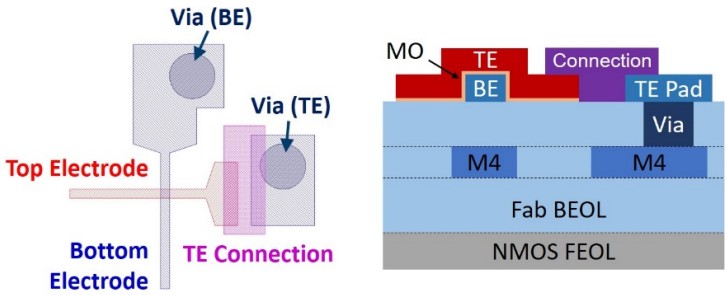

Figure 5.16: *Structure schematic of metal connection between RRAM's TE and TE pad.*

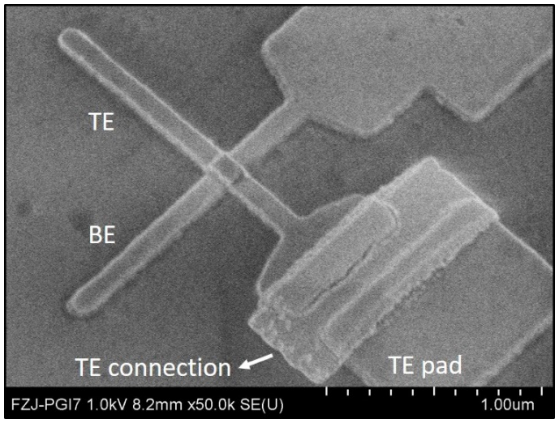

Figure 5.17: *SEM image of integrated RRAM after the TE connection established.*

5.3 Electrical characterization and nanoscale patterning process optimization

To characterize the integrated RRAMs on the CMOS die, there are 21 stand-along 1T-1R unit cells specifically designed for testing purposes before packaging, because the integrated 1T-1R arrays can only be characterized after packaging, when the customized hardware can be used to drive the connected peripheral circuits. As illustrated in Figure 5.18, the testing unit cell features four terminals connected to large metal pads for probing during electrical characterizations: the top electrode (TE), gate (G), source (S), and bulk (B). Additionally, to characterize the NMOS transistor, two testing unit cells among all have their drain (D) or bottom electrode (BE) terminals also connected to metal probing pads.

Initially, electrical characterization of the NMOS transistor was conducted. The transfer and output curves were obtained through DC voltage sweep measurements and are plotted in Figure 5.19. The transfer curve shows an exceptionally low drain current, restricted by the normally-off N-channel at $V_g = 0$. When V_g exceeds the V_{th}, approximately 1 V, the drain current approaches the mA range, which facilitates resistive switching and clear reading of the RRAM cell. Besides, by adjusting V_g within the range of 1 V to 2.5 V, for example, the current compliance applied from the transistor to the RRAM cell can be arbitrarily modulated between 40 µA to 960 µA, as indicated by the saturated drain currents in the output characteristic.

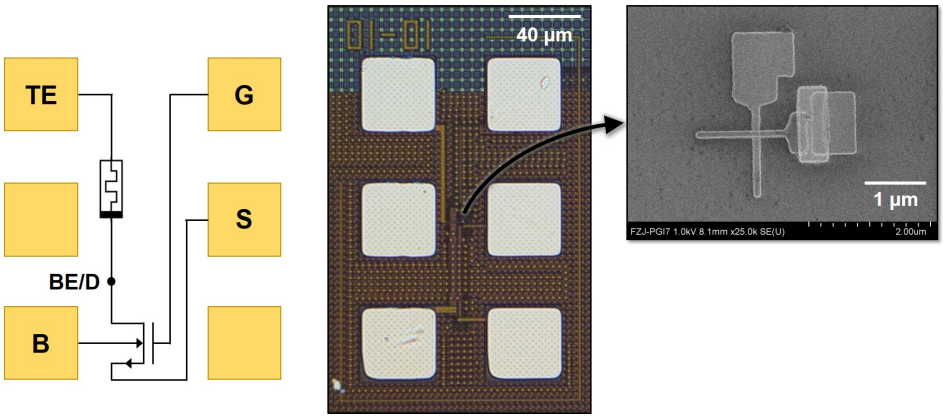

Figure 5.18: *Schematic diagram of stand-along 1T-1R unit cell for testing before packaging. The SEM image displays an integrated HOTO RRAM with the BE and TE both patterned by RIBE.*

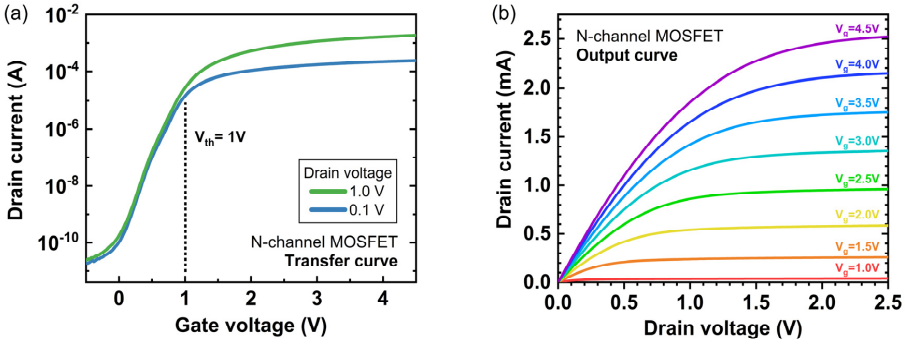

Figure 5.19: *Characteristics of foundry-manufactured N-channel MOSFET in the stand-along 1T-1R unit cell. (a) Transfer characteristic at two drain voltages, 0.1 V and 1 V. The extracted V_{th} is around 1 V. (b) Output characteristic at V_g ranging from 1 V to 4.5 V.*

Subsequently, an integrated HOTO RRAM with the BE and TE both patterned by RIBE was characterized in DC voltage sweep measurements. The I-V characteristic of 100 resistive switching cycles is depicted in Figure 5.20, where the positive voltage indicates to the V_{dd} applied at the TE terminal, while the negative voltage denotes the positive V_{dd} applied at the source terminal, indicated by adding a negative sign, for RESET operations. Despite the cycle-to-cycle (C2C) variability of $V_{SET,C}$ observed over 100 cycles, the mean $V_{SET,C}$ remains slightly below 1 V, and the maximum current during SET processes was precisely controlled at around 155 µA when $V_{g,SET}$ was fixed at 1.3 V. For the RESET processes, an ON/OFF ratio above 20 can be achieved with $V_{RESET,C}$ greater than 1 V, when $V_{g,RESET}$ and $|V_{RESET-stop}|$ were fixed at 4 V and 1.8 V respectively.

After the C2C study on one single 1T-1R unit cell, another four unit cells were investigated to disclose the D2D variability. By controlling HRS of all RRAM devices within a range of 160 kΩ to 240 kΩ, the D2D variability of $V_{SET,C}$ was studied through the cumulative distribution function (CDF) of each device, as shown in Figure 5.21a. The median $V_{SET,C}$ of

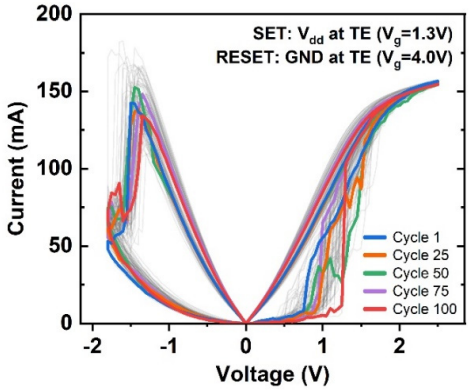

Figure 5.20: *I-V diagram of 100 switching cycles from the stand-along 1T-1R unit cell with a HOTO RRAM device. Two different testing schemes were used for the SET and RESET processes respectively. In the SET processes, the positive V_{dd} was applied on the TE of RRAM, and the source of NMOS transistor connected to the ground. Oppositely, in the RESET processes, the positive V_{dd} was applied on the source of NMOS transistor, and the TE of RRAM connected to the ground. The V_g was 1.3 V for the SET processes, and 4 V for the RESET processes. The $V_{g,SET}$ of 1.3 V is equivalent to a current compliance of around 155 μA.*

four devices center at 0.7 V, from which the other device deviates positively around 0.4 V with resulting $V_{SET,C}$ higher than 1V. Moreover, $V_{RESET,C}$ was studied with LRS controlled within a range of 10 kΩ to 20 kΩ. According to the CDF of $V_{RESET,C}$ plotted in Figure 5.21b, a significant D2D variability was observed with a maximum median difference of 0.9 V. Collectively considerably large D2D variabilities in SET and RESET processes were concluded from the 1T-1R unit cells for testing.

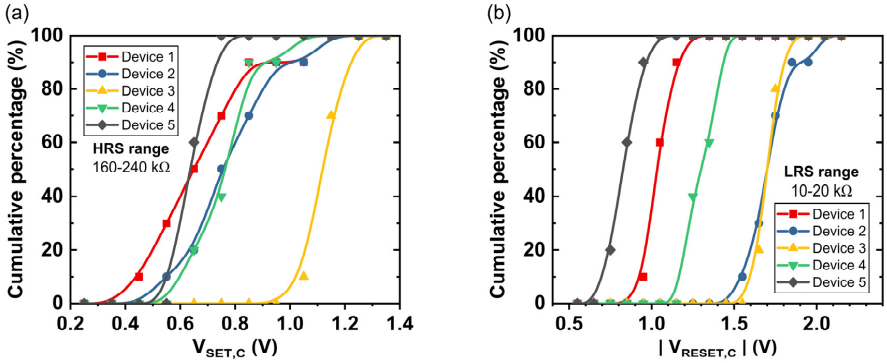

Figure 5.21: *D2D variability in five stand-along 1T-1R unit cells with a HOTO RRAM device. (a) D2D variability of $V_{SET,C}$ in the SET processes with the HRS fixed in a range of 160 kΩ to 240 kΩ. (b) D2D variability of $V_{RESET,C}$ in the RESET processes with the LRS fixed in a range of 10 kΩ to 20 kΩ. The statistics are based on ten SET and RESET processes respectively.*

The underlying cause of the large D2D variability links to the low yield of working RRAM devices. Merely 20 % of devices available for testing before packaging exhibit normal resistive switching after electro-forming. As monitored during the fabrication process of RRAM

integration, the device yield degradation originated from the process yield of nanoscale patterning. In-depth study on the root cause of low process yield and the feasible solutions will be discussed in the following sections.

5.3.1 Fencing during Ar-based RIBE process for Pt etching

During the fabrication process of RRAM integration, nanoscale patterning by RIBE often resulted in defective patterns owing to the critically downgraded effectiveness of resist removal performed in the resist stripper regardless of the duration. Taking the fabrication step for RRAM's BE and TE pad as an example, Figure 5.22 depicts the defective patterns observed after resist stripping in DMSO at 80 ℃ for 4 hours. Not only resist residue can be found on the Pt surface, but thin fences of unknown material can also be observed attaching at the edges of the desired patterns. Particularly for the thin feature serving as the actual BE of the RRAM, the resist entrapment caused by the surrounding fences notably affected the effectiveness of resist stripping. When one segment of the fence broke apart and detached from the edge, the effectiveness of resist stripping was improved. The detached segment may separate entirely from the remaining fence and either fall onto the nearby substrate surface or be carried away by the DMSO solvent during resist stripping. Alternatively, it may remain suspended when one end is left attaching to the remaining fence.

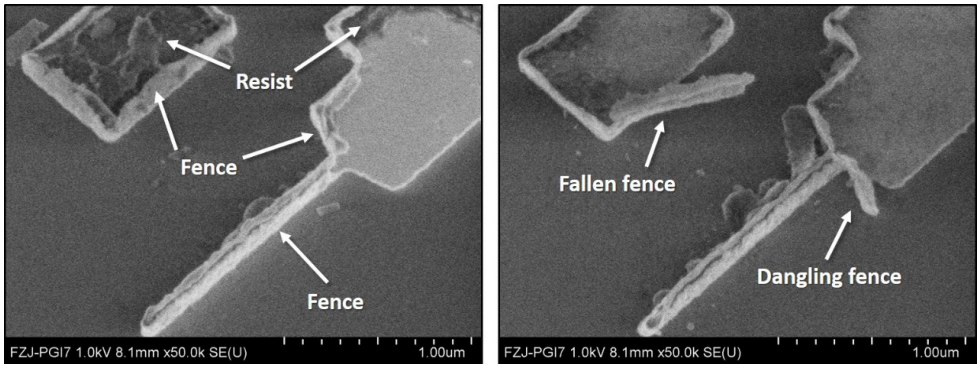

Figure 5.22: *SEM images of fabricated RRAM's BE and TE pad with fences observed after resist stripping.*

Even though additional O_2 plasma etching could remove the resist residue effectively by chemical reactions in the plasma, it did not solve the whole problem. The result can be seen in Figure 5.23. The resist residue is completely removed by O_2 plasma etching, but the fences fall nearby or stack on top of the desired pattern. Besides, it is nearly impossible to get rid of the fallen fences by ultrasonic cleaning in acetone possibly due to the Van der Waals force acting on the fallen fences. Overall, the fences exert a profound influence on diminishing the effectiveness of resist stripping, and their remaining part on the substrate even becomes contaminations further affecting the outcome of subsequent processes. To get rid of the fences is therefore the key to obtaining an appropriate patterning result.

To formulate a solution to the defective patterning result caused by the fences, a deeper comprehension of the fences is necessary. Taking one step back from resist stripping, the metal stack of 20-nm-thick Pt on 5-nm-thick Ti was etched through the Ar-based RIBE process with a mask of e-beam resist. The SEM inspection before and after the etching step, as displayed in Figure 5.24, disclosed a critical finding on fencing or fence formation. After the RIBE process, the substrate surface was darker because the metal stack was etched away and the dielectrics,

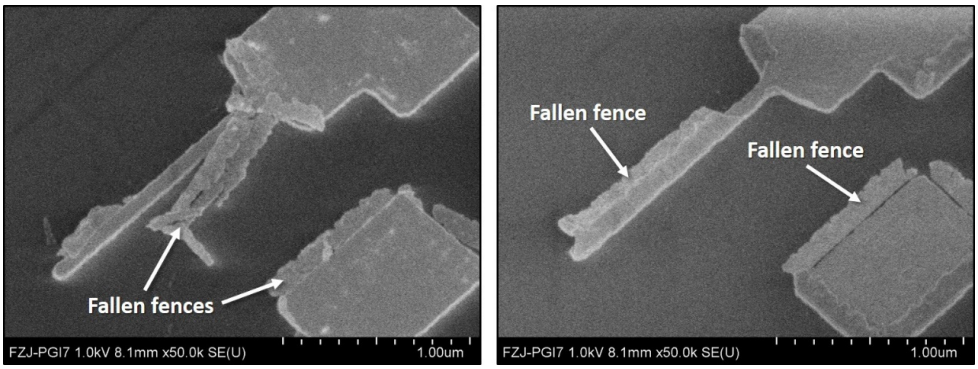

Figure 5.23: *SEM images of fences falling after additional O_2 plasma etching.*

SiO_2, was exposed. More importantly, a clear contrast difference was observed that the edge was brighter than the top surface of the resist pattern. As the resist exhibits a darker color under SEM, the resist pattern edge in a brighter color after etching indicates that a material other than the resist formed on the edge during etching, eventually resulting in the fences surrounding the resist pattern.

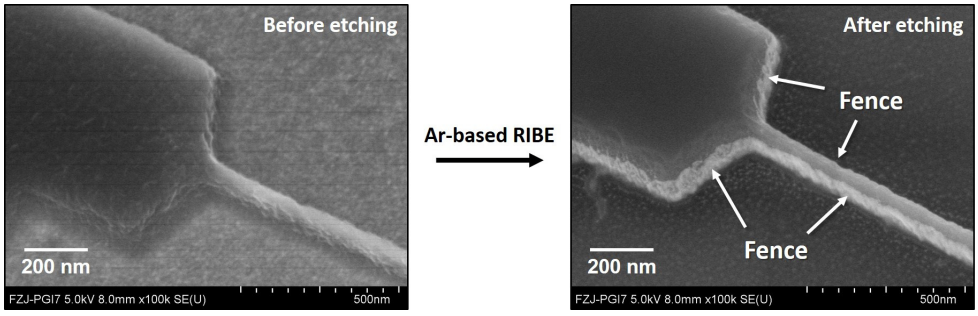

Figure 5.24: *Fencing resulting from Ar-based RIBE process using an e-beam resist mask.*

The presence of the fences was clearly shown in Figure 5.25 after removing the resist by O_2 plasma etching. The Pt surface was exposed with fences standing at the pattern edge. The fences are very complete instead of some randomly distributed pieces. It can be seen from the SEM image on the right that some parts of the fence were broken apart by mild stripping in an ultrasonic bath of DMSO, revealing the desired patterning result with an etching depth slightly larger than 25 nm. In comparison, the fence height is remarkably larger and analogous to the resist thickness before etching, i.e., close to 250 nm.

Based on the observations described above, sidewall redeposition during the Ar-based RIBE process has been deduced to be the primary cause of fencing [148-155]. Particularly for Pt etching to pattern the BE and TE, Ar ions was employed in RIBE because Pt cannot be reactively etched by any ion species. In the Ar-based RIBE process, material is etched through physical bombardment by Ar ion beam incident vertically onto the substrate surface, which is essentially sputtering of material from the substrate surface. The atoms sputtered from the surface are emit in all directions, thereby potentially deposited on the sidewalls of the resist pattern. The mechanism of sidewall reposition is illustrated in Figure 5.26 in a cross-sectional

view. Depending on the severity of sidewall redeposition, it can lead to tapering of the etched feature and even fencing, when the redeposit extensively covers the resist sidewalls.

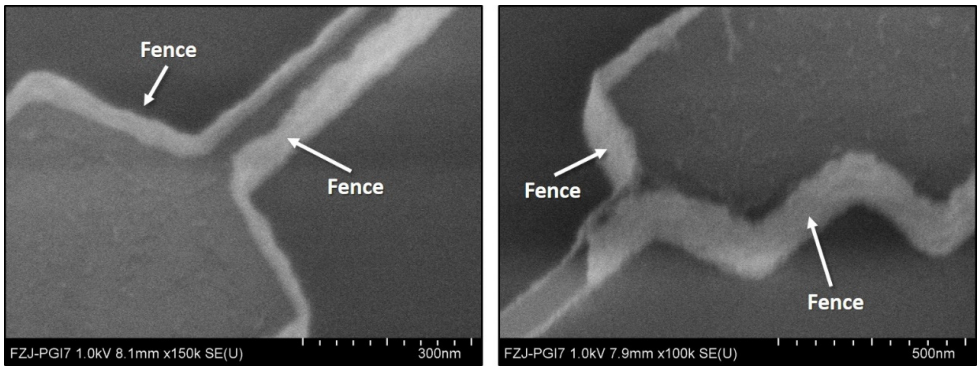

Figure 5.25: *Presence of fences standing at pattern edges after resist ashing by O_2 plasma etching.*

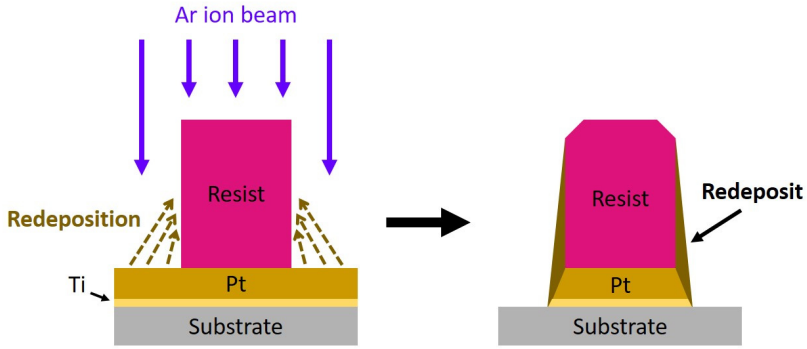

Figure 5.26: *Schematic diagram of sidewall redeposition during Ar-based RIBE process.*

Upon the formation of fences by sidewall redeposition, the resist stripping becomes more challenging. On the one hand, the contact area of resist to the solvent decreases substantially with the fences covering the edges. The wettability of the stripping solvent, on the other hand, limits the stripping effectiveness of the resist at the bottom or the corner regions near the fences, when a pattern width of 100 nm with a fence height of 250 nm is considered. Generally, there is a possibility of fence removal emerging during resist stripping in an ultrasonic bath of DMSO. Assisted with resist dissolution, the mechanical disturbance in the solvent could break the fences and cause fence detachment from the pattern edge. However, it is a random incident rather than a controllable process, which directly reflects on the yield that is lower than 50 %. A common stripping result is previously shown in Figure 5.22. To enhance the yield by excluding the process uncontrollability that comes along with the fences, a short-term solution based on post-processing and a long-term solution based on an alternative etching method will be discussed respectively in two following sections.

5.3.2 Fence removal by tilted-angle CF_4-based RIBE process

In order to eliminate the fences resulting from the Ar-based RIBE process, a post-processing step has been proposed using an additional RIBE process at a tilted angle, as illustrated in Figure 5.27. By adjusting the incident angle of the reactive ion beam to 80 degrees with respect to the surface normal, the fences can be uniformly etched from the side with substrate rotation. As the redeposit comprises not only Pt and other materials in the etching stack, but also organic products captured from resist etching, the reactive gas CF_4 has been chosen for the use in the RIBE to eliminate the fences [148].

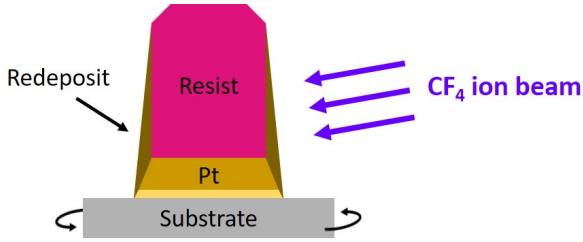

Figure 5.27: *Schematic diagram of tilted-angle CF_4-based RIBE process for removing the fences formed by the redeposit resulted from the Ar-based RIBE process.*

Utilizing the tilted-angle CF_4-based RIBE process to remove the fences resulted from TE patterning, which also involves an Ar-based RIBE process, the evolution over the etching time is displayed in Figure 5.28. It should be noted that a fallen fence that lies parallel to the BE was already present since the BE fabrication, and thus it is disregarded in the current discussion focusing on the TE fabrication. In the process of tilted-angle CF_4-based RIBE, no significant change was observed from the fence after 25 seconds. However, starting from 1 minute and 30 seconds, the fence began to undergo slight etching from the side, resulting in a small step that partially exposed the patterned TE underneath. When the etching lasted 3 minutes, the entire fence surrounding the TE pattern appeared to be roughly removed, and a noticeable decrease in the resist thickness can be identified, indicating the effect of resist etching. After the etching for 4 minutes and 30 seconds, the fence was completely removed, but the resist was harshly etched.

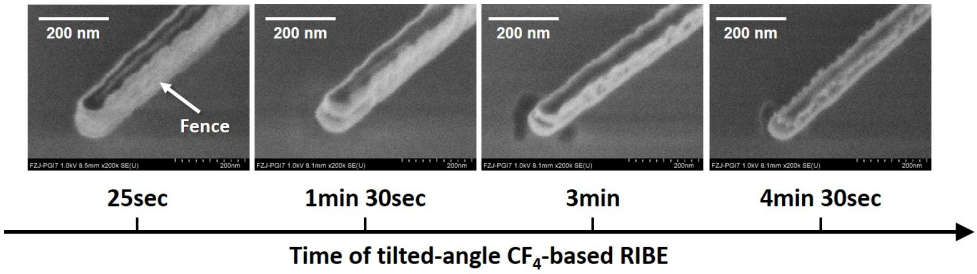

Figure 5.28: *Progress of fence removal over the etching time of tilted-angle CF_4-based RIBE process conducted in TE patterning.*

Based on the findings from various etching durations, 3 minutes and 30 seconds were determined to be the optimal duration of tilted-angle CF_4-based RIBE process for the fence removal. Notably, CF_4 is commonly utilized in RIBE to etch MO, implying that the fences can

be effectively removed from the side simultaneously with the MO layer etched from the top during the tilted-angle CF_4-based RIBE process conducted in TE pattering. To streamline the etching process and mitigate the risk of over-etching, the original CF_4-based RIBE step for the MO layer, typically lasting less than 30 seconds, was omitted. Consequently, in the TE pattering, an Ar-based RIBE process was directly followed by a tilted-angle CF_4-based RIBE process for 3 minutes and 30 seconds of tilted CF4 RIBE. The outcome is demonstrated in Figure 5.29 on the left, taking the TaO_x RRAM stack as an example. The fence was adequately etched, while the TaO_x layer was completely etched to expose the TE pad. It should be reiterated that fallen fences beneath the TE pattern were already present since the BE fabrication, and thus they are excluded from the current discussion. With the effective fence removal, a standard resist stripping step in DMSO at 80 °C for more than 4 hours successfully removed all the resist and eliminated any remaining broken fences, as depicted in Figure 5.29 on the right.

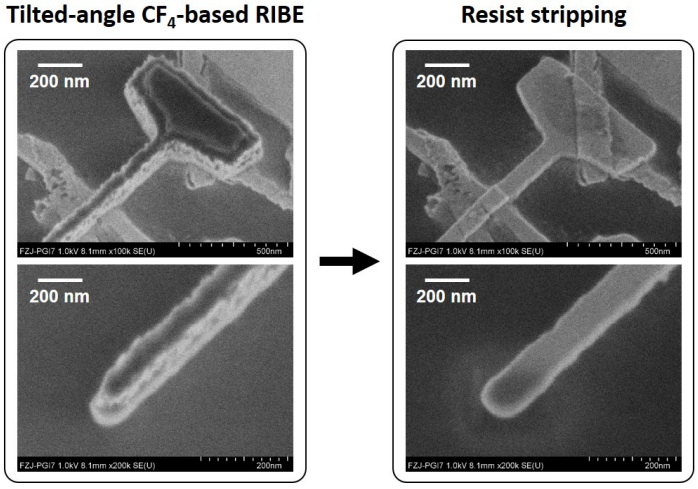

Figure 5.29: *Successful resist stripping result after tilted-angle CF_4-based RIBE.*

5.3.2.1 Electrical characterization of 1T-1R unit cells

By exclusively using the tilted-angle CF_4-based RIBE process for the fence removal during TE fabrication, the yield of functional TaO_x RRAM devices could surpass 90 %, as evaluated from the testing unit cells on chip. To study the D2D variability, five 1T-1R unit cells were electro-formed under $V_{g,FORM}$ of 1.5 V, which is equivalent to a current compliance of 300 µA. Controlling the HRS within a range of 90 kΩ to 140 kΩ, $V_{SET,C}$ of the five RRAM devices have the medians centering at 0.8 V with a maximum median difference of 0.25 V, as depicted in Figure 5.30a. Moreover, the narrow LRS distribution recorded from each RRAM device lies closely together in Figure 5.30b, implying a good LRS control in 1T-1R unit cells by fixing the $V_{g,SET}$ at 1 V or the equivalent current compliance at 40 µA. Based on the controlled LRS, $V_{RESET,C}$ exhibits a rather large C2C and D2D variabilities with a maximum D2D median difference of 0.39 V, as shown in Figure 5.30c. This suggests it is more stochastic to initiate the rupture than the growth or reconnection of a conductive filament. More precisely, it originates from the termination of previous change in the filament, where the reconnection of filament is more stochastic and abrupt than the rupture. Importantly, a much smaller D2D variability is found both in $V_{SET,C}$ and $V_{RESET,C}$, compared to the previous result shown Figure

5.21, obtained from the RRAM devices patterned without additional tilted-angle CF_4-based RIBE process for fence removal.

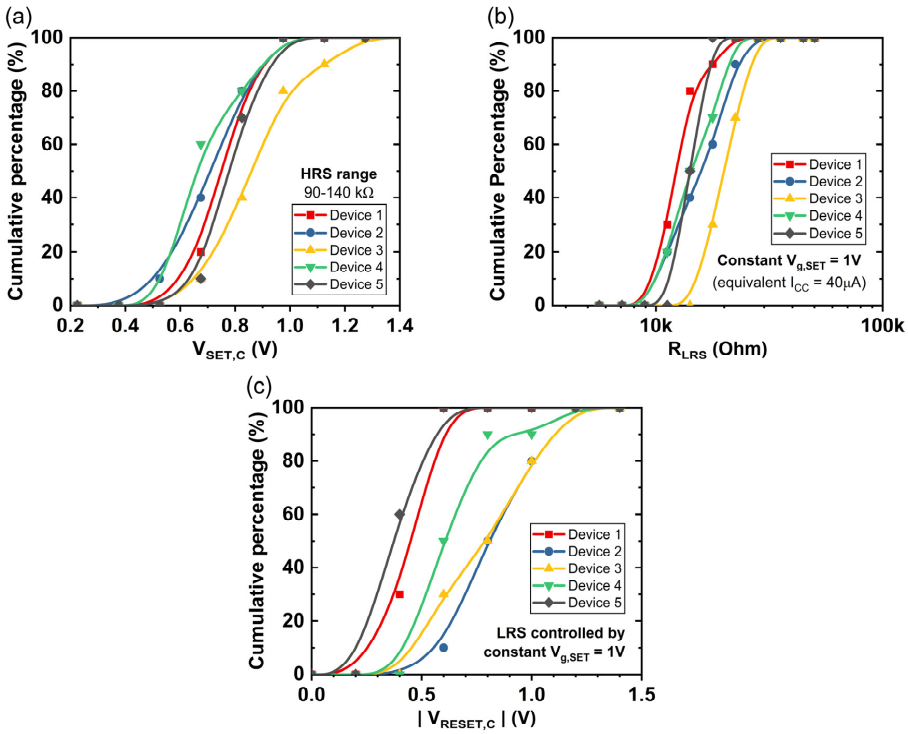

Figure 5.30: *D2D variability in five stand-along 1T-1R unit cells with a TaO_x RRAM device. Tilted-angle CF_4-based RIBE was applied in TE patterning for fence removal. (a) D2D variability of $V_{SET,C}$ in the SET processes with the HRS fixed in a range of 90 kΩ to 140 kΩ. (b) D2D variability of R_{LRS} corresponding to the SET processes with a constant $V_{g,SET}$ of 1 V. (c) D2D variability of $V_{RESET,C}$ in the RESET processes with the LRS controlled by a constant $V_{g,SET}$ of 1 V. The statistics are based on ten SET and RESET processes respectively.*

Focusing on one single 1T-1R unit cell, multi-level switching can be achieved by $V_{g,SET}$ modulation. Five levels determined by various $V_{g,SET}$ from 1.1 V to 1.5 V are plotted with corresponding RESET cycles under a constant $V_{g,RESET}$ of 4.5 V in Figure 5.31a. The higher equivalent current compliance in the SET process, given by the higher saturation current at a higher $V_{g,SET}$, fosters a more vigorous filament growth, which leads to a lower LRS, as observed from Figure 5.31b. With the lower LRS due to the stronger filament, a higher $V_{RESET,C}$ was found when the RESET operation was performed after the SET operation under a higher $V_{g,SET}$, as shown in Figure 5.31c [32].

In addition to $V_{g,SET}$, multi-level switching can also be achieved by $V_{RESET-stop}$ modulation. Fixing $V_{g,SET}$ and $V_{g,RESET}$ at 1.5 V and 4.5 V respectively, resistive switching of six different $|V_{RESET-stop}|$ from 1.7 V to 2.2 V is demonstrated and plotted in log scale in Figure 5.32a to bring out the low-current I-V characteristics of HRS. Extracting the HRS from Figure 5.32a, a trend of increasing R_{HRS} with increasing $|V_{RESET-stop}|$ can be recognized in Figure 5.32b [32,

156]. This trend coincides with the fact that a higher $V_{RESET-stop}$ causes a larger degree of filament rupture, resulting in a larger gap from the filament tip to the BE. Besides, the $|V_{RESET-stop}|$ of 2.2 V leads to a much larger R_{HRS} variation, indicating a decay of HRS controllability as the $|V_{RESET-stop}|$ increases. As a result, the HRS controllability decay appears to be the limit of $V_{RESET-stop}$ modulation. Based on the HRS corresponding to various $V_{RESET-stop}$, the $V_{SET,C}$ is plotted against the $V_{RESET-stop}$ in Figure 5.32c. A weaker increasing trend of $V_{SET,C}$ with increasing $|V_{RESET-stop}|$ was observed [32], compared to the distinct trend of $V_{RESET,C}$ in multi-level switching modulated by $V_{g,SET}$. In another word, $V_{SET,C}$ is less sensitive than $V_{RESET,C}$ when modulating $V_{RESET-stop}$ and $V_{g,SET}$ respectively for multi-level switching.

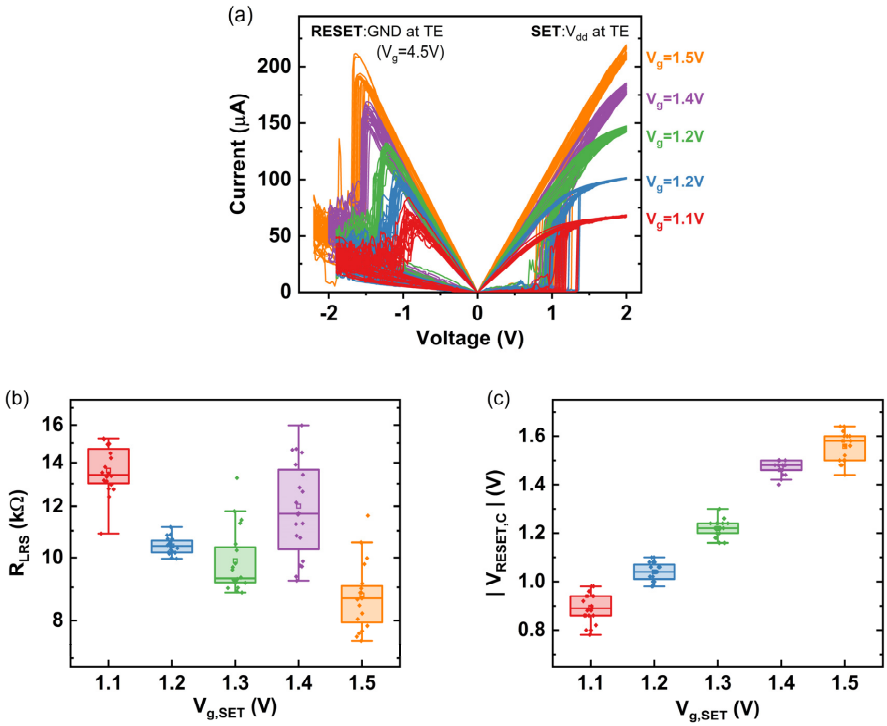

Figure 5.31: *Multi-level switching by $V_{g,SET}$ modulation in a stand-along 1T-1R unit cell with a TaO_x RRAM device. (a) I-V diagram of resistive switching with the applied $V_{g,SET}$ ranging from 1.1 V to 1.5 V. Two different testing schemes were used for the SET and RESET processes respectively. In the SET processes, the positive V_{dd} was applied on the TE of RRAM, and the source of NMOS transistor connected to the ground. Oppositely, in the RESET processes, the positive V_{dd} was applied on the source of NMOS transistor, and the TE of RRAM connected to the ground. The $V_{g,RESET}$ was fixed at 4.5 V. Effects of $V_{g,SET}$ modulation on (b) R_{LRS} and (c) $V_{RESET,C}$ are summarized from 20 switching cycles of each applied $V_{g,SET}$.*

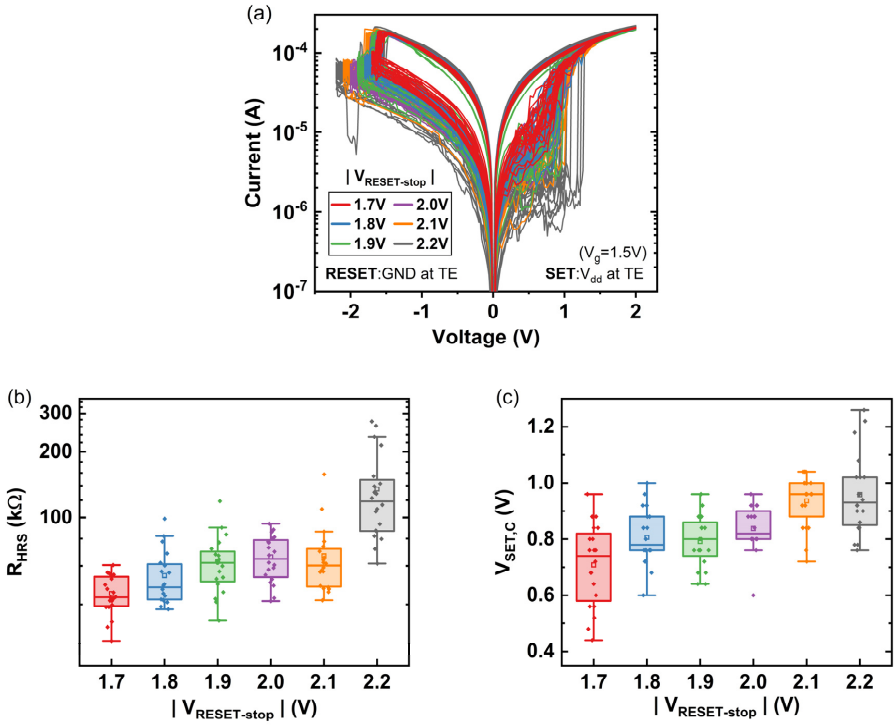

Figure 5.32: *Multi-level switching by $V_{RESET-stop}$ modulation in a stand-along 1T-1R unit cell with a TaO_x RRAM device. (a) I-V diagram of resistive switching with the applied $|V_{RESET-stop}|$ ranging from 1.7 V to 2.2 V. Two different testing schemes were used for the SET and RESET processes respectively. In the SET processes, the positive V_{dd} was applied on the TE of RRAM, and the source of NMOS transistor connected to the ground. Oppositely, in the RESET processes, the positive V_{dd} was applied on the source of NMOS transistor, and the TE of RRAM connected to the ground. The $V_{g,SET}$ and $V_{g,RESET}$ were fixed at 1.5 V and 4.5 V respectively. Effects of $V_{RESET-stop}$ modulation on (b) R_{HRS} and (c) $V_{SET,C}$ are summarized from 20 switching cycles of each applied $V_{RESET-stop}$.*

By applying different V_g during electro-forming in a 1T-1R unit cell, the effect of $V_{g,FORM}$ on resistive switching was investigated. For $V_{g,FORM}$ ranging from 1 V to 1.5 V, the LRS under a constant $V_{g,SET}$ of 1 V was evaluated from six relevant devices. A primary decreasing trend of R_{LRS} with increasing $V_{g,FORM}$ can be found in Figure 5.33a, where a remarkable R_{LRS} drop appears from 1 V to 1.1 V. Despite a number of oxygen vacancies initially existing in the pristine substoichiometric TaO_x layer, the concentration is substantially low, resulting in a poor initial electrical conductivity. Electro-forming is a critical process that brings up the conductivity in the first place by applying a V_{FORM}, much higher than V_{SET}, and puts the RRAM device into the working condition for resistive switching. Specifically, electro-forming involves redox reactions that introduce a critical amount of oxygen vacancies from the Ta layer to the TaO_x layer, and oxygen vacancy migration that forms the first complete filament, connecting the TE and the BE. Termination of electro-forming is practically determined by current compliance because the actual voltage applied across the device drops dramatically,

ceasing the electro-forming, when the loading current reaches the compliance value. The higher the current compliance, the higher the concentration of oxygen vacancies that can be introduced and the stronger the initial filament that can be formed in terms of structure geometry. Hence, a higher current compliance during electro-forming creates a highly-conducting working condition as the starting point for resistive switching, thereby increasing the conductance or decreasing the resistance of LRS. Moreover, the remarkable R_{LRS} drop implies the existence of a critical current compliance that dramatically changes the electro-formed condition.

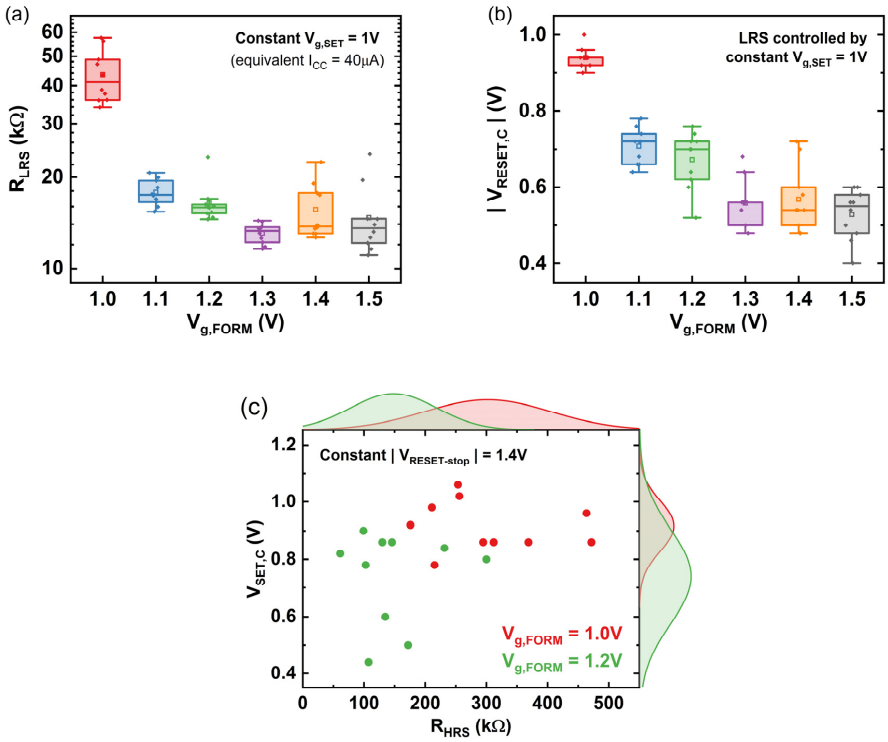

Figure 5.33: *Effect of $V_{g,FORM}$ on resistive switching properties. Six stand-along 1T-1R unit cells were electro-formed at different $V_{g,FORM}$ ranging from 1 V to 1.5V. The statistics are based on ten bipolar switching cycles respectively, when the $V_{g,SET}$ was fixed at 1 V. (a) Box plot of R_{LRS} versus $V_{g,FORM}$. (b) Box plot of $|V_{RESET,C}|$ versus $V_{g,FORM}$. (c) Scatter plot of $V_{SET,C}$ versus R_{HRS} with probability distribution curves displayed on the corresponding axes. The $V_{g,FORM}$ of 1 V and 1.2 V are compared with a constant $|V_{RESET-stop}|$ of 1.4 V.*

From the LRS under a constant $V_{g,SET}$ of 1 V, RESET operations were performed by fixing $V_{g,RESET}$ at 4.5 V. The $V_{RESET,C}$ and the corresponding $V_{g,FORM}$ are recorded in Figure 5.33b, where a major trend can be noticed that $|V_{RESET,C}|$ primarily decreases with increasing $V_{g,FORM}$. By comparing Figure 5.33a and Figure 5.33b, a lower R_{LRS} was found to correspond to a lower $V_{RESET,C}$. Interestingly, it is the opposite of the result previously concluded from Figure 5.31a and Figure 5.31b, which describes the case that the R_{LRS} of a single TaO_x RRAM is modulated by the $V_{g,SET}$. To understand the cause of this result disparity, it should be noticed that the current case describes the R_{LRS} of different TaO_x RRAMs corresponding to different $V_{g,FORM}$

applied during the electro-forming process. In other words, dynamics of conductive filament in separate systems are currently discussed. Here, the positive correlation between R_{LRS} and $|V_{RESET,C}|$ implies that the highly-conducting working condition introduced by a high $V_{g,FORM}$ allows the critical RESET condition to be fulfilled at a low $V_{RESET,C}$ even for a relatively low LRS.

Further fixing the $|V_{RESET-stop}|$ at 1.4 V, R_{HRS} of the devices corresponding to $V_{g,FORM}$ of 1 V and 1.2 V are marked in a coordinate system with $V_{SET,C}$ of the succeeding SET process, as shown in Figure 5.33c. According to the distributions of R_{HRS} and $V_{SET,C}$ displayed at the sides, the higher $V_{g,FORM}$ results in lower R_{HRS} and $V_{SET,C}$. As discussed previously, the higher $V_{g,FORM}$ leads to a lower LRS, from which a constant electrical stress in the RESET process can switch the device to a relatively low HRS. To enable the resistive switching back to the LRS, the required electrical stress to trigger the SET process is therefore relatively small, indicating a relatively low $V_{SET,C}$. As a conclusion to the role of $V_{g,FORM}$ in 1T-1R resistive switching, it determines the initial working condition for conductive filament after electro-forming, which exhibits direct impacts on resistive switching parameters including R_{LRS}, R_{HRS}, $V_{SET,C}$ and $V_{RESET,C}$.

5.3.3 Fence-free RIE process for Pt etching

To address the root cause of fencing during Ar-based RIBE, an alternative etching technique is required to serve as a long-term solution. In this regard, a RIE process using a gas mixture of Cl_2 and Ar has been proposed along with the use of a Cr hard mask in lieu of a polymer-based resist mask. Despite the inclusion of Cl_2 in the etching gas, physical sputtering remains as the dominant etching mechanism for Pt, leaving the undesired sidewall redeposition in place. Nonetheless, the redeposite solubility is significantly increased when a HCl-based solution is used for post-RIE wet etching [150-152, 154, 155, 157, 158]. Furthermore, the use of a hard mask has been found beneficial to fence-free Pt etching [159-161]. As the etching products of Cr hard mask are also captured in the process of sidewall redeposition, the redeposit composition changes accordingly. Therefore, a Cr etchant comprising a mixture of perchloric acid and ceric ammonium nitrate can react with the Cr-containing compounds in the redeposit simultaneously during Cr mask removal, which facilitates elimination of the fences. Moreover, the Cl_2-Ar RIE has the etching rates of Pt and Cr in a ratio of 5:4, resulting in a low etching selectivity. As considerable Cr etching occurs concurrently during the Cl_2-Ar RIE process, the protruding parts of fences are also etched, thereby reducing the fence height. In addition, reduction of the mask thickness lowers the height of sidewall, and therefore diminishes the fence height. As the thickness of a Cr mask is directly determined from the deposition of the Cr thin film, a thickness below a few tens of nm can be achieved. Overall, in the Cl_2-Ar RIE process with a Cr mask, the reduced mask thickness and low etching selectivity decrease the fence height, while the Cr etchant can eliminate the fences during Cr mask removal.

Utilizing the Cl_2-Ar RIE process with a Cr mask in TE patterning, which also involves an Ar-based RIBE process, the process flow and the results are displayed in Figure 5.34. To pattern the Cr mask by lift-off, it started with EBL using the positive e-beam resist *CSAR AR-P 6200.04* and followed by mild O_2 plasma etching as a descum step. Considering the pattern transfer though the TE of 15-nm-thick Pt on 15-nm-thick Ta and the switching layer of 10-nm-thick TaO_x, 20-nm-thick Cr was deposited next by thermal evaporation as the hard mask. After the lift-off step in the resist remover *AR-P 600-71* for 30 minutes, the complete Cr mask for TE could be observed in a dark color under SEM, as shown on the left in Figure 5.34. The pattern of Cr mask was afterwards transferred into the Pt/Ta/TaO_x stack by a two-step etching involving Cl_2-Ar RIE and CF_4 RIBE. From the result shown in the middle of Figure 5.34, a

thin layer of remaining Cr, around 5 nm, was observed with the edges outwardly tapered as the result of an aggressive Cr etching due to the low selectivity. Favorably, no fence was noticed at this stage. The Cr etchant *TechniEtch Cr01* was lastly used to remove the Cr mask and the potentially remaining fences. The final result shown on the right of Figure 5.34 concludes a successful demonstration of fence-free patterning on the Pt-layer-containing stack.

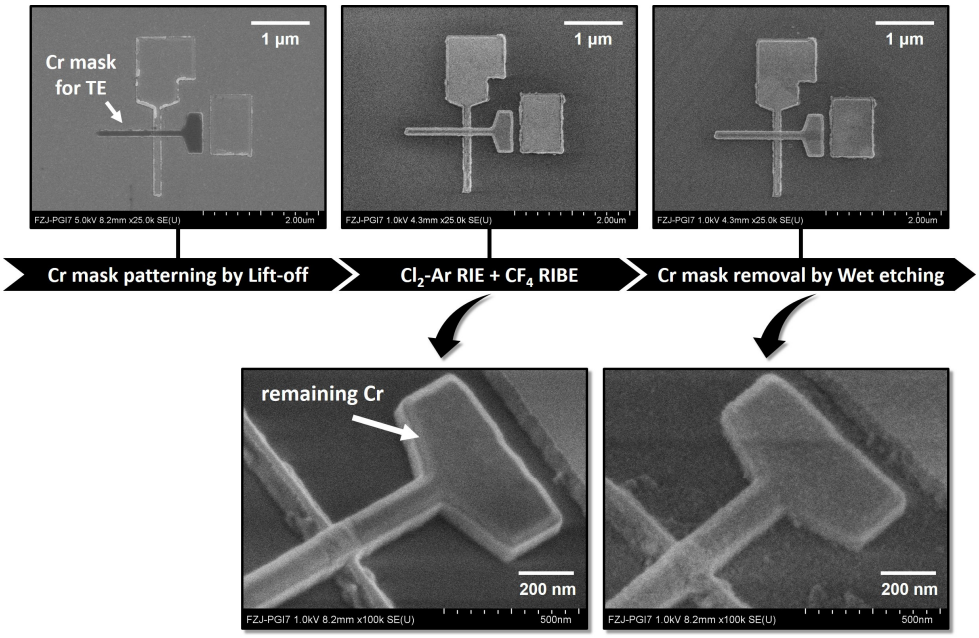

Figure 5.34: *Process flow of fence-free RIE patterning technique. The SEM images illustrate the actual application on RRAM's TE patterning.*

5.3.3.1 Electrical characterization of 1T-1R unit cells

With TE patterned by the fence-free RIE process with a Cr mask and BE patterned by a lift-off approach, the integrated RRAM devices benefit from the superior yields of both critical patterning steps. Resistive switching of 50 cycles was demonstrated from a 1T-1R unit cell in the on-die testing area and plotted in Figure 5.35, when $V_{g,SET}$, $V_{g,RESET}$, and $|V_{RESET-stop}|$ were controlled at 1.2 V, 4.5 V, and 1.2 V respectively. A superb switching properties are achieved, as the $V_{SET,C}$ and $|V_{RESET,C}|$ are both below 1 V and the ON/OFF ratio is close to 20. As a result, it can be concluded that the fence-free etching technique is able to not only provide a pronounced decrease of process defects in nanoscale patterning but also deliver well-functioning crossbar RRAM devices.

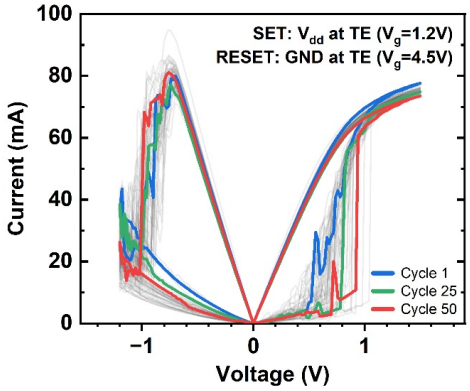

Figure 5.35: *I-V diagram of 50 switching cycles from the stand-along 1T-1R unit cell with a TaO_x RRAM device. Fence-free RIE was applied in TE fabrication. Two different testing schemes were used for the SET and RESET processes respectively. In the SET processes, the positive V_{dd} was applied on the TE of RRAM, and the source of NMOS transistor connected to the ground. Oppositely, in the RESET processes, the positive V_{dd} was applied on the source of NMOS transistor, and the TE of RRAM connected to the ground. The V_g was 1.2 V for the SET processes, and 4.5 V for the RESET processes.*

5.4 Chip packaging and verification for a customized operating hardware

To drive the CMOS die with monolithically integrated RRAM devices and taped-out circuits peripheral to the 1T-1R arrays, a customized operating hardware had been built on a printed circuit board (PCB) that communicates externally through a microcontroller [147]. The chip-operating hardware, as displayed in Figure 5.36, overall consists of a power plug-in area, a chip socket hosting a packaged chip, a number of DC analog reference circuits, and a microcontroller that executes operating commands to the chip.

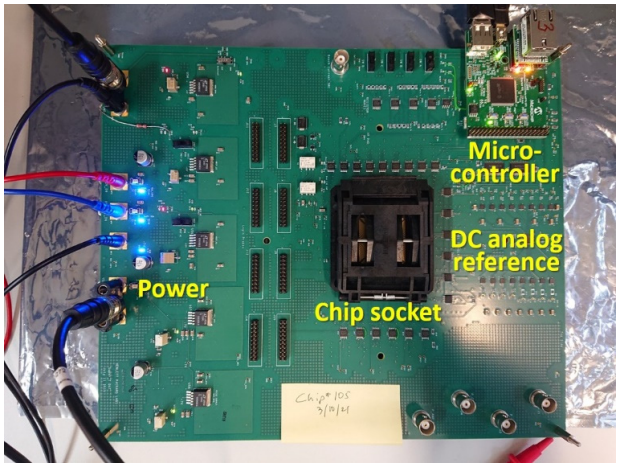

Figure 5.36: *Customized operating hardware of the RRAM-integrated CMOS chip.*

Until the 1T-1R arrays can be controlled by the hardware for any operation and further application, packaging of a known-good-die must be carried out according to the specification of the chip socket equipped on the PCB. When the packaged chip is mounted in the socket, a system verification can then be tested, where WRITE and READ operations of a specified RRAM cell would be demonstrated.

5.4.1 Chip packaging

As quad flat packaging with 304 leads was selected in the hardware design, a clamshell-socket and a ceramic lead frame of the CERQUAD® packages from KYOCERA were utilized accordingly. The chip packaging in this work, outsourced to Microelectronic Assembly Frankfurt (Oder) GmbH, started with adhesive bonding that fixed the CMOS die at the center of the lead frame on a surface coated with Au paste. Next, wire bonding connected the 304 I/O pads, previously formed on the die close to the edges, to the inner leads at four sides by thin gold wires. Lastly, outer leads were cut to match a specified dimension of the chip socket on the PCB. The packaging result is shown in Figure 5.37.

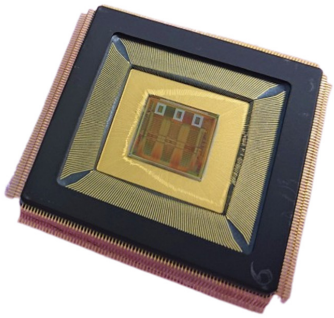

Figure 5.37: *Result of outsourced chip packaging on ceramic lead frame.*

5.4.2 Electrical verification of packaged chips

A packaged chip configured with TaO_x RRAMs was electrically tested on the customized operating hardware to verify the functionality. To this purpose, basic operations such as READ and WRITE were performed on integrated RRAMs in 64×64 1T-1R arrays. In the framework of a Python interface that had been built to drive the microcontroller, a READ operation is a command to return a calculated current value that refers to the loading current of a specified RRAM cell during a voltage pulse of 0.15 V. On the other hand, a WRITE operation is a command to apply a voltage pulse of a given amplitude (V_{WRITE}) and a given length to a specified RRAM cell without sensing. Depending on the biasing polarity, a WRITE operation can be intended to trigger either a SET or a RESET process in the RRAM cell.

To verify the functionality of the packaged chip in the hardware, bipolar resistive switching cycles were demonstrated on an integrated RRAM by programming WRITE operations with V_{WRITE} sweeps. Every WRITE operation was followed by a READ operation to obtain the resistance of RRAM (R_{READ}), which is derived from the current readout. The result showcasing 5 switching cycles is plotted in Figure 5.38. Concerning the WRITE operation for SET, the V_{WRITE} is indicated to be positive in the diagram. When the gate voltage ($V_{g,SET}$) and the pulse width (t_{SET}) were fixed at 1.5 V and 100 µs respectively, the SET processes can be found to be triggered by V_{WRITE} around 0.5 V. On the other hand, considering the WRITE operation for RESET, the V_{WRITE} is denoted to be negative in the diagram. With the gate voltage

($V_{g,RESET}$) and the pulse width (t_{RESET}) fixed at 5 V and 1 µs respectively, the RESET processes occurred at V_{WRITE} close to -1.5 V.

The resistive switching cycles demonstrated in Figure 5.38 clearly confirm the successful execution of programmed operations on the RRAM-integrated chip, which verifies the basic functionality of the hardware system and pave the way to energy-efficient AI applications by exploiting the DPEs.

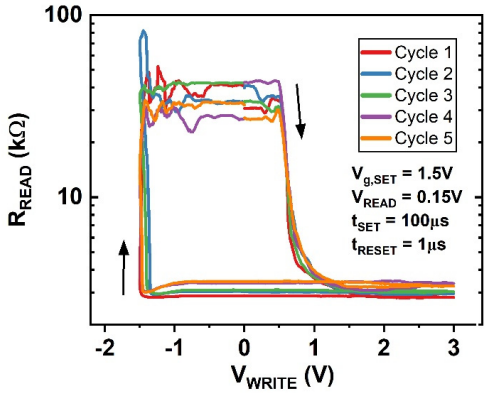

Figure 5.38: *Bipolar switching cycles of the integrated TaO_x RRAM in a 64×64 1T-1R array operated by the customized operating hardware. Both READ and WRITE operations were executed by a single square pulse. R_{READ} recorded after each WRITE operation is plotted against V_{WRITE}. In WRITE operations, the $V_{g,SET}$ and $V_{g,RESET}$ were fixed at 1.5 V and 5 V respectively, and the t_{SET} and t_{RESET} were fixed at 100 µs and 1 µs respectively. A constant V_{READ} of 0.15 V was used in READ operations.*

Chapter 6
Thermal crosstalk in passive nano-crossbar arrays of TaO_x RRAMs

In the pursuit of more powerful IC chips through more compact designs, the increase in device density has been one of the major pathways. However, every device in IC chips generates heat when operating. The heat generated per unit area rises, as the device density increases. As a result, the issue of heat dissipation becomes prominent [162-164]. A poor heat dissipation raises the working temperature that can possibly exceed the device working specification, leading to operation errors or irreversible physical damages. Similarly for memristive arrays, heat dissipation and relevant thermal issues have to be taken into account when memristors are downscaled and situated in a high-density crossbar array. In the past decade, several simulation studies have investigated the thermal crosstalk in RRAM arrays, which describes the thermal impact of active devices to the adjacent non-active devices regarding changes in their resistance states [22-26]. Practically during operations in an RRAM array, the heat caused by Joule heating of the active device is conducted to the adjacent devices though the metal connection in between, which serves also as a good thermal conductor. Among various RRAMs, the thermal crosstalk is expected to exert a greater influence on the ones that show a substantial temperature dependence in the switching mechanism, such as the VCM-type RRAMs, which are based on redox reaction and oxygen vacancy migration. In this context, passive nano-crossbar arrays of VCM-type TaO_x RRAMs with device areas of 100×100 nm^2 and the device spacing below 100 nm were fabricated in this work to experimentally disclose the thermal crosstalk effect in practice.

6.1 Layout design of nano-crossbar arrays

To study the thermal crosstalk, passive crossbar arrays were employed, instead of active 1T-1R arrays, due to the layout simplicity that enables a smaller device spacing and decreases the process complexity. The structure of passive crossbar array is adopted from a crossbar device, where the RRAM cell is formed by an MIM stack at the crossing point of a BE and a TE with a switching layer in between. A 4×4 crossbar array is depicted in Figure 6.1 as an example, where a cross-sectional schematic shows that the RRAM cell is built by patterning the BE first and then the TE directly through the switching layer.

Besides the 4×4 array, various line arrays with only one BE, or referred to as the word line, have been included in the layout design on a substrate of 2×2 cm^2, ranging from 1×2, 1×3, 1×4, up to 1×8 arrays, in order to provide simpler array configurations for the thermal crosstalk study. All the different arrays are arranged together in blocks in two formats, as shown in Figure 6.2. The blue patterns refer to the BEs, while the red ones stand for the TEs. For both formats, there are 64 probing pads arranged to form a certain square shape in order to match the probe card of an electrical characterization system, which enables high-throughput testing. To make

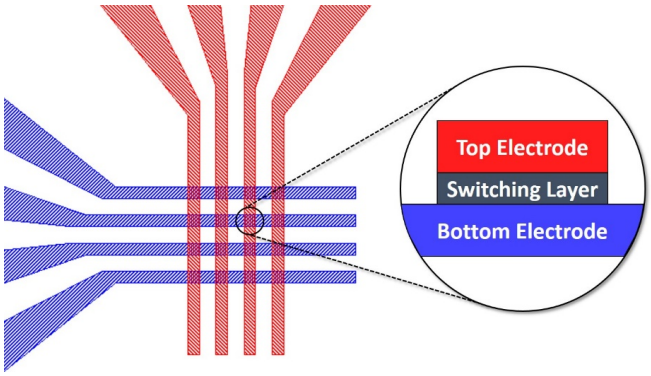

Figure 6.1: *Schematic diagram of passive RRAM crossbar array.*

use of the space, some 1×3 arrays or single devices have been arranged in the center of the block, which can be tested at a normal four-probe station without a probe card. In total on a substrate of 2×2 cm^2, the layout deign involves 10 sizes of different device spacing. The amounts of different arrays for every spacing size are listed in Table 6.1. Moreover, in order to enable auto alignments in EBL for the two pattern layers, global and local alignment marks have been designed according to e-beam writer specifications of the EBPG series from Raith GmbH, which was employed in this work.

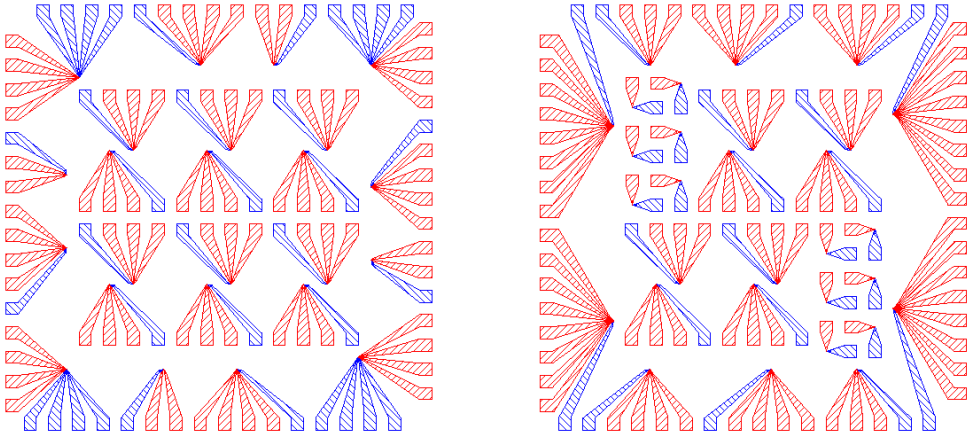

Figure 6.2: *Pattern layout of passive crossbar arrays arranged in blocks in two formats.*

Table 6.1: *Array amounts of different configurations designed on one substrate of 2×2 cm^2.*

Array configuration	Amount
1×2	4
1×3	17
1×4	6
1×8	2
4×4	4

6.2 Fabrication process of nano-crossbar arrays

Based on the layout design, the passive nano-crossbar arrays were fabricated in two nanoscale patterning steps respectively for the BE and the TE, which was patterned directly through the switching layer. To avoid the fencing issue during nanoscale patterning that was investigated in Section 5.3.1, the previously verified long-term solution, the Cl_2-Ar RIE with Cr masks, was involved both in the BE and the TE patterning. Basically, the process flows of patterning the BE and the TE are the same despite differences in detail regarding the deposition and dry etching, as shown in Figure 6.3.

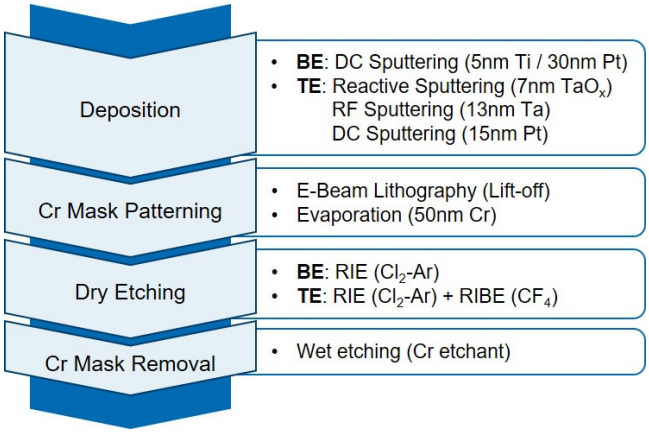

Figure 6.3: *Process flow of BE and TE in passive TaO_x RRAM nano-crossbar arrays.*

Starting from the BE fabrication, the BE stack was deposited first by in-situ DC sputtering that a 5-nm-thick Ti adhesion layer was followed by a 30-nm-thick Pt layer. Next, EBL was performed by using the positive e-beam resist *CSAR AR-P 6200.04* in order to obtain a Cr mask pattern for BE patterning through a lift-off approach. After a descum step by mild O_2 plasma etching to remove the residual resist possibly remained in the developed pattern, Cr was evaporated for 50 nm and a lift-off step was then carried out in the resist remover *AR-P 600-71*, dedicated to the *AR-P* resists, for 30 minutes. Subsequently, the BE stack was etched by the Cl_2-Ar RIE with the 50-nm-thick Cr mask, which was removed afterwards by wet etching in the Cr etchant *TechniEtch Cr01*. The TE fabrication, like the BE, began with the in-situ deposition of the TE stack, which involved reactive sputtering of 7-nm-thick TaO_x, RF sputtering of 13-nm-thick Ta, and DC sputtering of 15-nm-thick Pt. By using the same lift-off

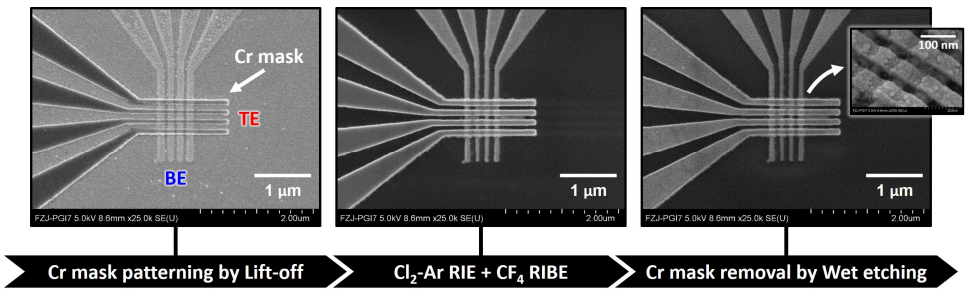

Figure 6.4: *Pattering process with the use of Cr hard mask.*

approach, the Cr mask was formed for the subsequent dry etching step, which started with the Cl_2-Ar RIE and ended with RIBE using CF_4 gas. At last, the remaining Cr mask was removed similarly by the wet etching. The results monitored during TE pattering were shown in Figure 6.4, taking the 4×4 array as an example.

6.3 Process issues of device spacing reduction

To bring out the thermal crosstalk effect, which has a strong spatial dependence, the spacing of RRAM devices must approach the feasible minimum based on the process selection. When exploring the limit of the nanoscale patterning by EBL and the dry etching with Cr masks, some process issues were found affecting the outcome and the limit of actual spacing. In this subchapter, the process issues of major concerns for minimizing the device spacing will be discussed.

6.3.1 E-beam lithography and lift-off

In the process development stage, a preliminary test was conducted for the lift-off process using EBL. Following the process flow previously mentioned in the Subchapter 6.2, the lift-off result of a metal stack consisting of 5-nm-thick Ti and 20-nm-thick Pt is shown in Figure 6.5, where the line width was designed to be 100 nm. When the exposure was executed directly according to the original line width, i.e., 100 nm, the stand-alone metal line had the line width increased, for instance, to 122 nm, as shown in the SEM image on the left. Besides, the increase in line width can also be found in the metal lines with the spacing designed equal to the width. As shown in the SEM image on the right for instance, the line width increased from 100 nm to 115 nm, while the line spacing decreased from 100 nm to 78 nm.

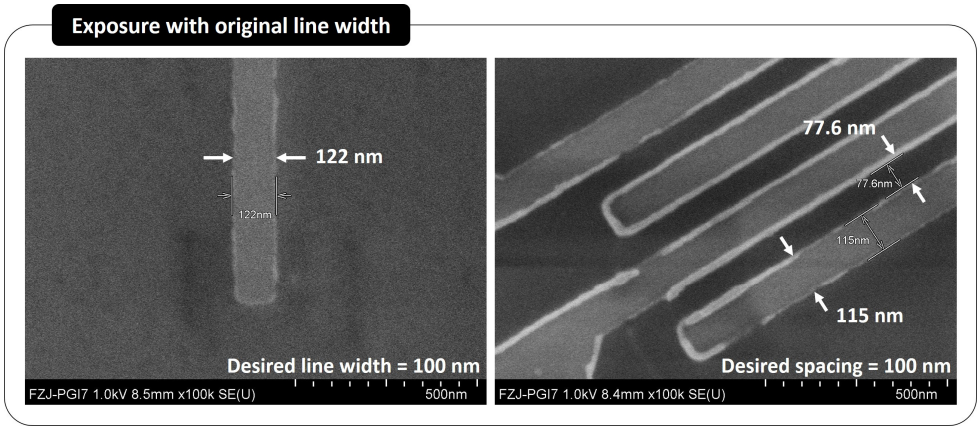

Figure 6.5: *Lift-off patterning result of e-beam exposure with original line width.*

The line width increase, accompanied with the spacing reduction, can be explained by the pattern enlargement due to the proximity effect of e-beam writing and the non-ideal undercut profile of e-beam resist. The proximity effect comes from interactions of electrons with the resist, substrate, and nearby structures during the exposure by e-beam writing, as illustrated in Figure 6.6 [165-171]. The scattered electrons and secondary electrons cause unintentional exposure outside the region originally addressed by the e-beam, leading to pattern fidelity loss

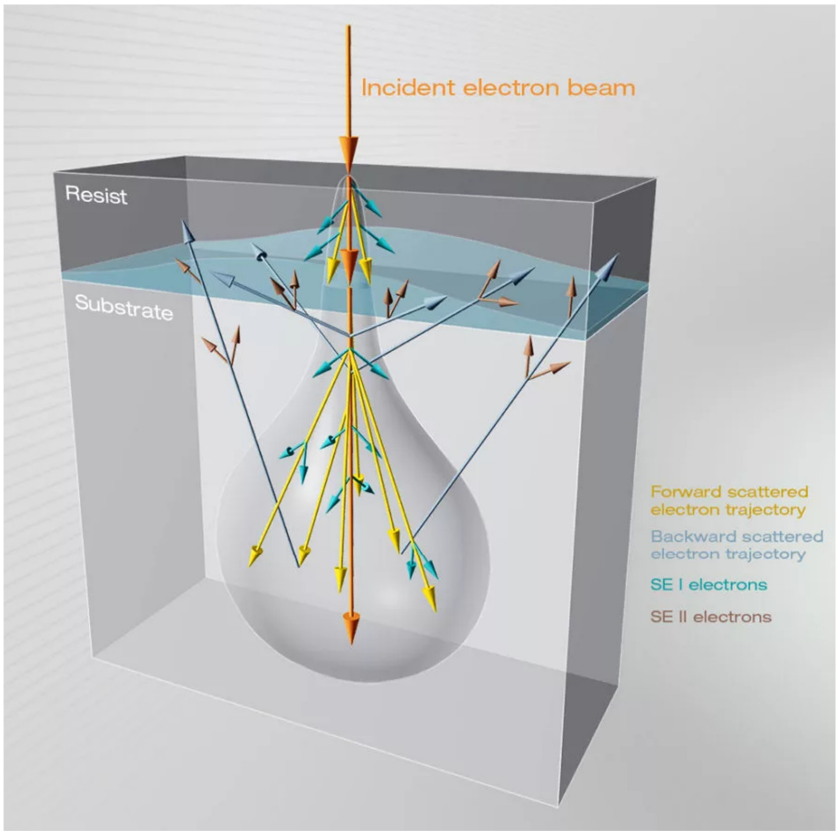

Figure 6.6: *Proximity effect caused by e-beam interaction with resist and substrate [167].*

such as blurring, distortion, and magnification. On the other hand, the pattern enlargement can also be attributed to the resist undercut profile, dedicated to the lift-off process. As shown in Figure 6.7, the undercut is a typical profile that can favorably prevent the deposition on the resist sidewalls. A clear exposure of the resist sidewalls can provide an advantage for lift-off, which relies on dissolving the resist by a solvent. However, the undercut also creates a room for the deposited feature to expand during deposition. As a result, a non-ideal undercut profile can cause a discrepancy of feature dimension between the outcome and the design, among which feature enlargement is commonly found. Possibly, the feature enlargement can be further enhanced, when the undercut profile is altered by the proximity effect during e-beam writing and the mild O_2 plasma etching in the descum step right before the deposition.

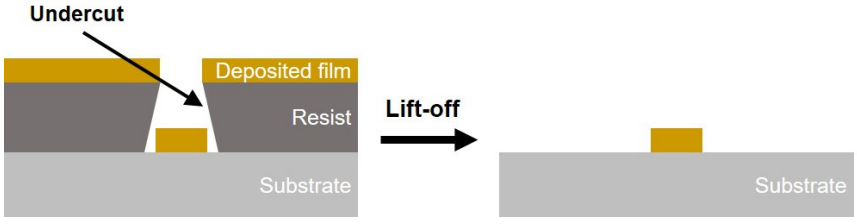

Figure 6.7: *Schematic diagram of undercut resist profile for lift-off patterning.*

To compensate the pattern enlargement caused by the proximity effect and the non-ideal undercut profile, the exposure was corrected by a pattern shrinkage in the layout design. Specifically, it was implemented by a line width reduction that the exposed line width decreased from 100 nm to 80 nm. The lift-off result is shown in Figure 6.8, where the actual line width has a better match to the desired value of 100 nm. The line width of a stand-alone metal line was found to be 102 nm, and the metal lines with the designed spacing of 50 nm were accurately fabricated that the line width and the spacing were found to be 97 nm and 51 nm respectively.

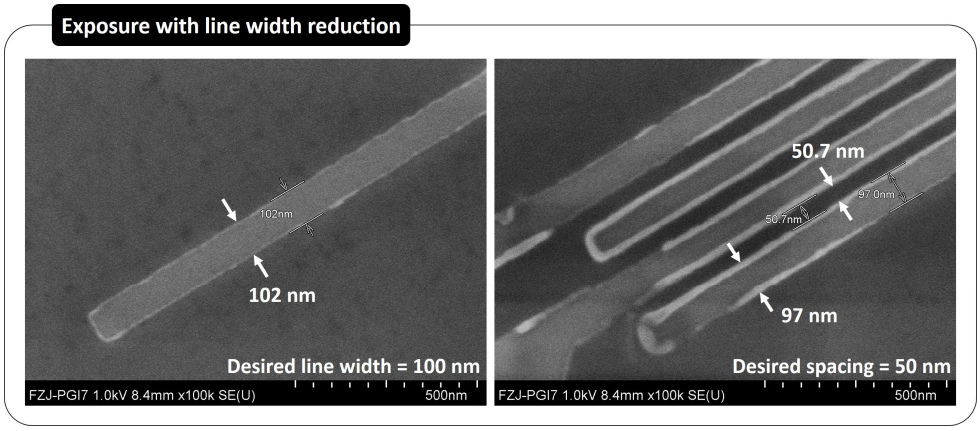

Figure 6.8: *Lift-off patterning result of e-beam exposure with line width reduction.*

6.3.2 Dry etching for top electrode patterning

Despite the accurate line width obtained for the Cr mask, an increase in line width was observed after the dry etching for TE patterning. Taking the 4×4 array designed with the line width of 100 nm and the line spacing of 70 nm as an example, Figure 6.9 depicts the actual situations before and after the dry etching, involving Cl_2-Ar RIE and CF_4 RIBE. Before the dry etching for TE patterning, the line width and line spacing of the Cr mask were found to be 108 nm and 65 nm respectively. After the dry etching step patterned the TE stack through the switching layer, the line width increased by 18 nm and the line spacing decreased by 14 nm. Compared to the layout design, the actual TE spacing decreased, in total, by approximately 20

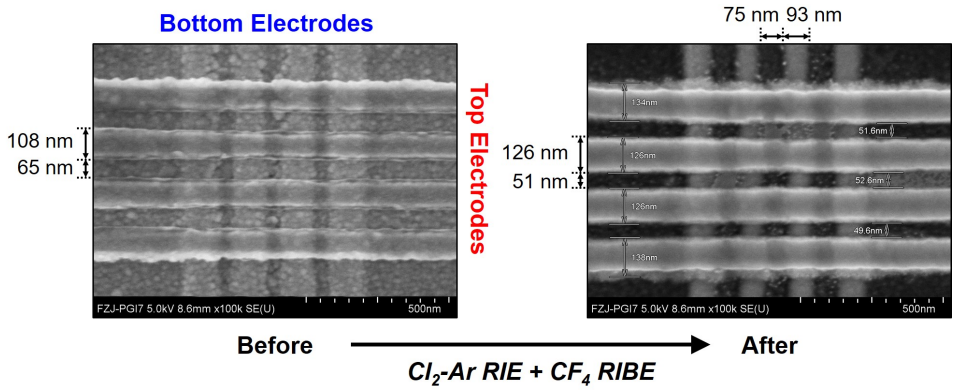

Figure 6.9: *Line width increase after dry etching in TE patterning.*

nm, which is a considerable deviation especially when targeting the spacing below 100 nm. Meanwhile, it can be noticed that the fabricated BEs have the line width of 93 nm and the line spacing of 75 nm, which does not show pattern enlargement effect. Excluding the Cl_2-Ar RIE, which was used for both BE and TE patterning, the CF_4 RIBE was concluded as the cause of pattern enlargement possibly originating from the tapering effect during the etching.

6.3.3 Process defects and within-wafer variability

Besides the pattern enlargement effect emerging along the fabrication process, any process defect can affect the outcome and functionality of fabricated structures. Taking the 4×4 arrays designed with the line width of 100 nm and the line spacing of 30 nm as an example, the result of BE patterning is shown in Figure 6.10 for all four arrays on one substrate. Generally, the desired line spacing was resolved by the patterning process. However, there are pattern defects, pointed out by arrows in Figure 6.10, bridging the adjacent patterns except for the array shown in Figure 6.10c. Due to the fact that the patterned structures involve metal layers, these pattern defects create undesired metal connections, causing short circuits of the adjacent BEs, suggesting the fact that even a tiny defect coming from the process variability within wafer can already lead to malfunction of the entire array. As the line spacing is downscaled to the point when the within-wafer variability of the patterning process cannot be properly tolerated in the designed structure, the process yield can be critically affected.

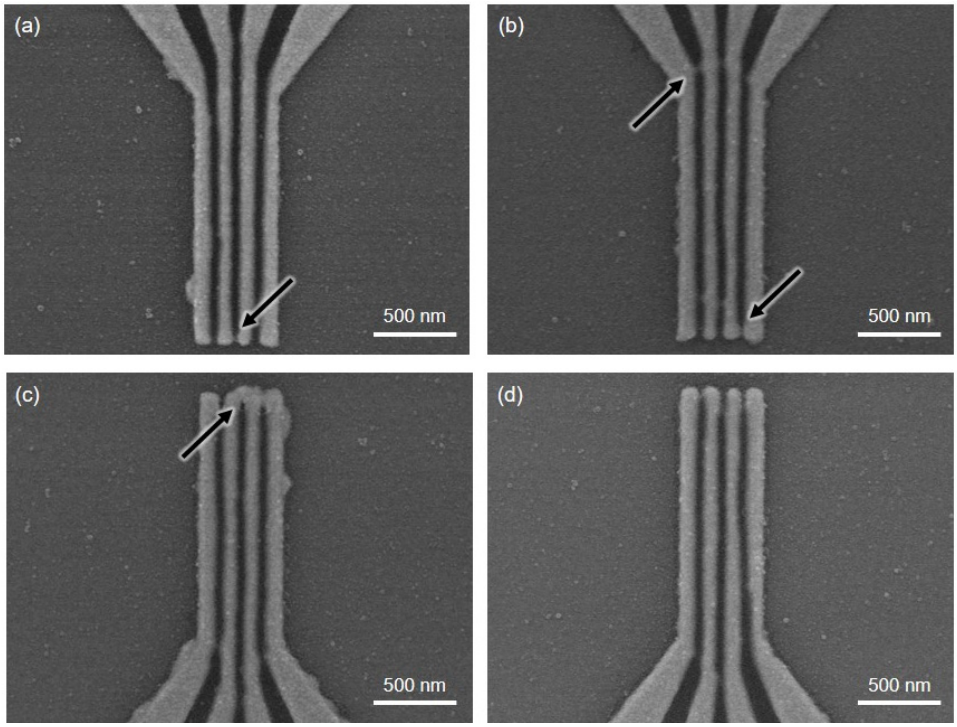

Figure 6.10: *Pattern defects and within-wafer variability in BE patterning. The result of all four arrays of the same kind fabricated on a substrate are shown individually in SEM images. (a)(b)(c) Pattern defects between adjacent BEs causing short circuits. (d) Proper patterning result with all BEs separate from each other.*

For the TE patterning, the pattern enlargement effect caused by the dry etching enhances the within-wafer variability. Taking the 4×4 array designed with the line width of 100 nm and the line spacing of 70 nm as an example, the result of TE patterning is shown in Figure 6.11. As pointed out by arrows, large defects can be found connecting the adjacent TEs, when the actual line spacing was already reduced to 50 nm because of the pattern enlargement effect. Meanwhile, the BEs in the array were fully resolved, and the actual spacing was found to be 70 nm without a considerable reduction. As a result, the more pronounced within-wafer variability of TE patterning turned out to be the limiting factor, when minimizing the device spacing in the crossbar array.

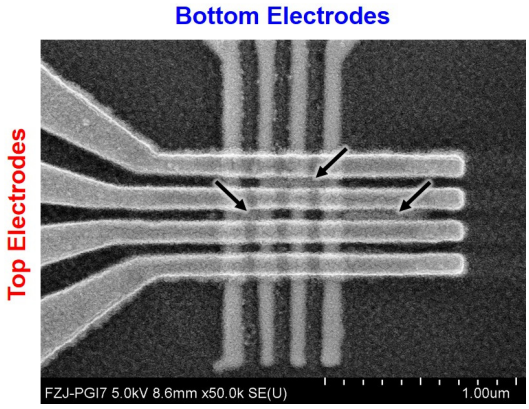

Figure 6.11: *Pattern defects in TE patterning result.*

6.4 Electrical characterization of RRAM device

To investigate the thermal crosstalk in the fabricated arrays, the general functionality of RRAM devices must be tested. Before the resistive switching behavior can be characterized, the RRAM device needs to be electro-formed. An actual electro-forming step performed on a stand-along device is shown in Figure 6.12, when the DC voltage sweep was applied with the I_{CC} set at 10 µA. As the voltage raised above 4 V, a rapid current increase was observed, indicating the electro-forming was triggered and the device switched from the initial state to the LRS. However, a substantial current drop closely followed, implying the resistance switched to an extremely high value. As shown in the SEM images taken before and after the electro-forming step, the current drop originating from a physical damage on the TE, which led to disconnection and even destruction of the device. This physical damage was attributed to Joule heating. As the line width is downscaled for the nanoscale crossbar devices, the parasitic resistances from the metal lines, formed as the BEs and TEs, are increased. When current overshoot occurs during resistive switching, the huge current spike can lead to an extraordinary high temperature caused by Joule heating, which prompts the melting of metal lines. As the electro-forming is triggered at a higher voltage than a normal SET process, it encounters a larger current overshoot accompanied with more drastic Joule heating. Generally, the current overshoot is due to the instrumental time delay of sourcemeter for the current compliance to be actually in place since the sensed current reached the compliance value. As shown in Figure 6.12, the current drop occurred actually far before the current reached the I_{CC} of 10 µA, suggesting that the sudden current increase during current overshoot caused the melting of metal line even before the sourcemeter could properly sense the current and trigger the current

compliance mechanism. Thus, the sourcemeter is incapable of providing sufficient control on the current overshoot.

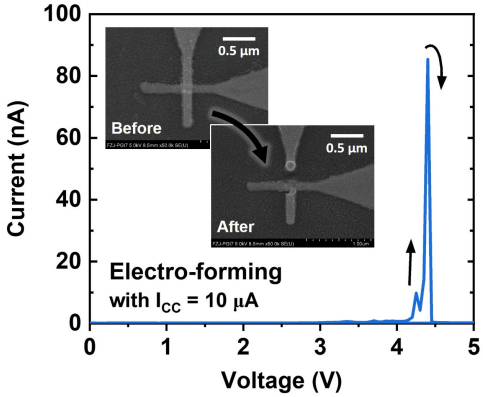

Figure 6.12: *Electro-forming failure due to electrode melting. A current compliance set by the sourcemeter was applied at 10 µA. The inserted SEM images show a structure destruction of the TE after the electro-forming.*

Alternatively, the electro-forming step was performed with an external series resistor to inherently prevent the large current overshoot. When the series resistor has the resistance (R_{series}) much larger than the R_{LRS}, it can immediately take over the voltage drop from the RRAM device during the resistive switching to the LRS, appropriately suppressing the loading current over the RRAM device. With a series resistor of 100 kΩ, the actual electro-forming performed by DC voltage sweep is shown in Figure 6.13a. A huge current jump was found at 4.1 V, where the RRAM device switched from the initial state to the LRS. The current was then limited just below the linear curve of $R = 100\text{k}\Omega$, indicating the current overshoot protection was properly implemented owing to the series resistor dominating the total resistance. After the electro-forming, the series resistor was removed to conduct resistive switching over 120 cycles by DC voltage sweep fixing I_{CC} at 100 µA during the SET operations. A consistent bipolar switching was observed without any considerable trend over 120 cycles despite the inherent C2C variability, according to the result plotted in Figure 6.13b.

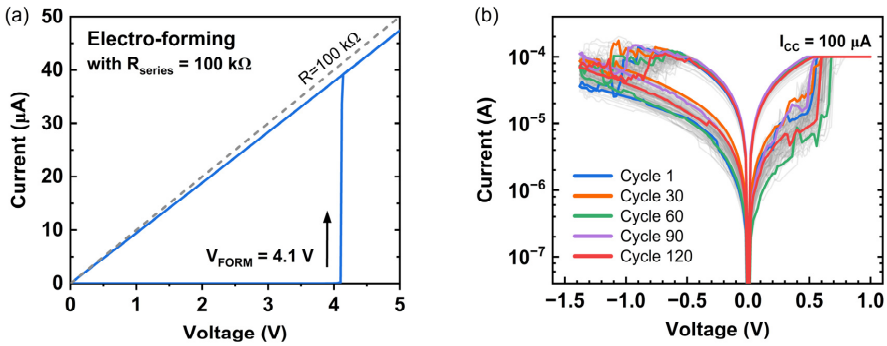

Figure 6.13: *Resistive switching properties of nanoscale TaO_x RRAM electro-formed with a series resistor. (a) I-V diagram of the electro-forming with R_{series} of 100 kΩ. The I-V relationship of $R = 100k\Omega$ is displayed as a baseline by a dashed line. (b) I-V diagram of 120 bipolar switching cycles with I_{CC} at 100 µA in the SET processes.*

To find the minimum of effective device spacing that allows every RRAM device in the array function appropriately, the fabricated arrays were examined by SEM inspection and electrical characterization. Eventually, 70 nm was found to be the valid minimum of line spacing and device spacing, of which the fabricated 1×3 array is shown in Figure 6.14 as an example. All the three RRAM devices were electro-formed with a series resistor of 100 kΩ, and the resistive switching was carried out individually on each device over 15 cycles, when I_{CC} of 100 µA was applied during SET operations. Apart from the D2D switching variability that can be distinguished fairly from the median switching cycles, which are marked in the dark color, all the three devices are functioning with a self-consistent switching behavior, implying a proper working array.

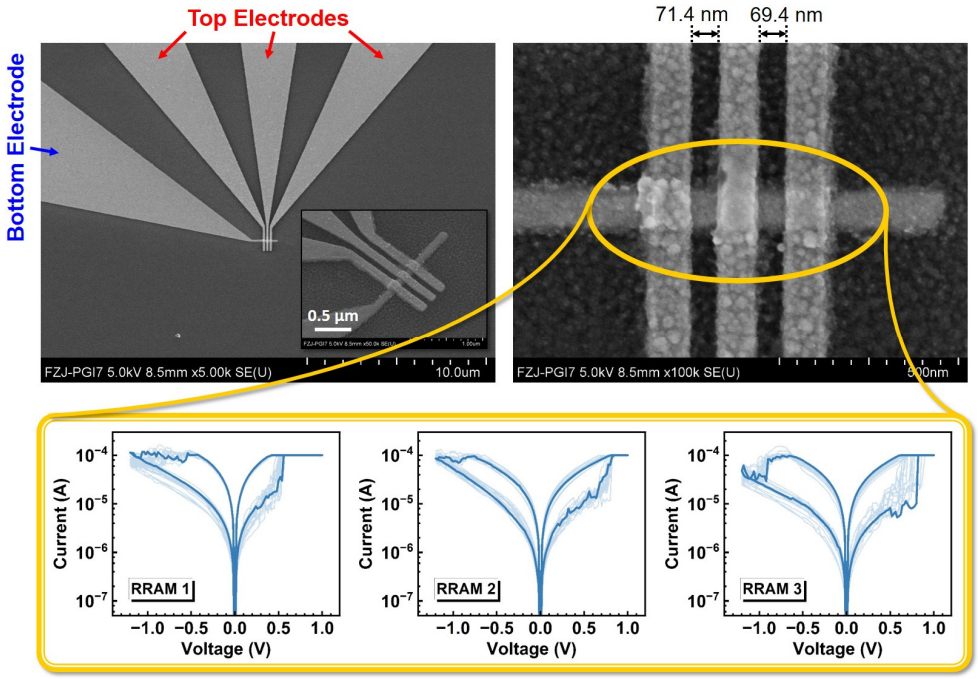

Figure 6.14: *Resistive switching characteristics of TaO_x RRAMs in a 1×3 crossbar array with electrode spacing of 70 nm. The 1×3 array has three TEs with a common BE. All RRAMs were electro-formed initially with a series resistor of 100 kΩ. With I_{CC} of 100 µA applied in the SET processes, 15 switching cycles of each RRAM are plotted in an I-V diagram, where the median cycle is marked out explicitly.*

6.5 Investigation of thermal crosstalk

As the thermal crosstalk in crossbar arrays focuses on the impact from a selected cell to the adjacent cells, 1×3 arrays are a relatively simple case for analyzing the thermal crosstalk effect. With three cells involved in total, the study of 1×3 array can be simply conducted by operating the middle cell as the selected cell and monitoring the only two adjacent cells, intrinsically disabling any uncontrollable interventions possibly from other cells that have to be considered when a larger array is used.

Using the V/2 scheme for array operation, the selected word line is grounded, and a WRITE voltage (V_{WRITE}) is applied on the selected bit line, while the $V_{WRITE}/2$ is applied on all the unselected bit lines and word lines. The assessment of thermal crosstalk was focused on the WRITE operation from the HRS to the LRS, which involves a SET process triggered on the fully-selected (active) cell. Thus, the V_{WRITE} should be chosen that $V_{WRITE} \geq V_{SET,C} > 0$. Since every half-selected cell is positively biased by $V_{WRITE}/2$, which is principally smaller than $V_{SET,C}$, it is impossible to trigger a SET process. However, the thermal crosstalk from the active cell can possibly foster and even trigger a SET process on the adjacent half-selected cell. The bit flip from the HRS to the LRS can, therefore, be monitored to evaluate the thermal crosstalk effect, when the adjacent half-selected cells are programmed to the HRS beforehand. Moreover, in order to simplify the testing condition and procedure, the fully-selected cell can be programmed to the LRS in advance such that the switching process from the HRS to the LRS is excluded from the WRITE operation. When applying the V_{WRITE} for the WRITE operation, the fully-selected cell in the LRS can still result in a high current that can induce the thermal crosstalk effect. In summary, the testing setup is illustrated in Figure 6.15 on the left that the cell 2 in the center of 1×3 array is fully-selected under the V/2 scheme, when it is in the LRS whereas the other cells in the HRS. To bring out the impact of thermal crosstalk, this exact scenario is compared to a reference scenario, where the heat source is removed by not applying the V_{WRITE} to the cell 2, and the cell 1 and 3 stay half-selected, as depicted in Figure 6.15 on the right.

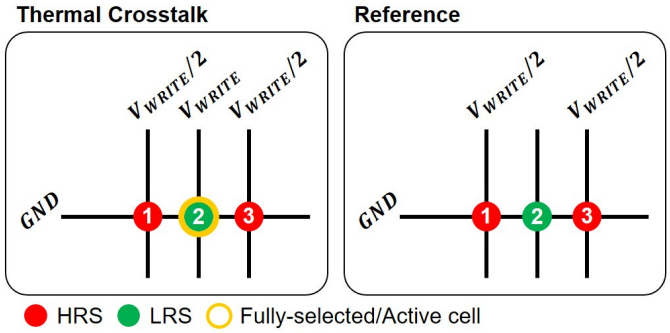

Figure 6.15: *Testing schemes of a 1×3 crossbar array in thermal crosstalk and reference scenarios based on the V/2 scheme. The serial numbers of RRAM cells are labeled, and the designated resistance states are marked by colors. The yellow circle indicates the fully-selected or active cell, on which the full* V_{WRITE} *is applied.*

For actual operations on the RRAM array, voltage is applied in the form of square pulses. As the key to any in-depth experimental study of thermal crosstalk, verification of its actual impact in the array is of the largest interest. To this purpose, a long single pulse was used rather than short multiple pulses to strengthen the thermal effect generated from the active cell. As illustrated in Figure 6.16, the long single pulse can keep the active cell generating heat such that the thermal crosstalk continuously remains in effect. On the other hand, the time delay between short pulses provides a cooling time for the entire system, when there is no active cell contributing as the heat source. As a result, if no impact of thermal crosstalk is observed by using the long single pulse, there is absolutely no observable impact when the short multiple pulses are used. To verify the impact of thermal crosstalk, a long single pulse is applied to a 1×3 array with the pulse amplitude following the V/2 scheme, while resistances of the half-selected cells were checked before and after the long pulse to identify the possible bit flips.

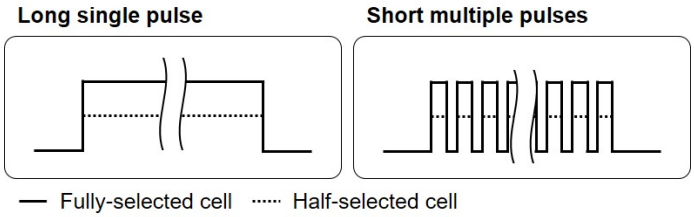

Figure 6.16: *Pulse scheme schematics of long single pulse and short multiple pulses.*

To perform the testing efficiently on the 1×3 array, a programmed measurement has been designed as described by the flow chart shown in Figure 6.17. After resistances of the cell 1 and 3 (R_1 and R_3) are obtained by an initial READ operation, they are programmed to a defined HRS range by alternative loops of SET and RESET operations respectively with incremental V_{SET} or V_{RESET}. Following the V/2 scheme, a long single pulse is then applied to each cell to execute a WRITE operation on the cell 2, when it is already programmed to the LRS in advance. Subsequently, a READ operation is carried out on the cell 1 and 3 to determine whether the bit flips occur. The whole process is then repeated until the number of total runs hit the defined value. Similarly, the same programmed measurement flow can be used for the reference scenario with the only change that no voltage is applied on the cell 2 when applying the long single pulse to the array after the cell 1 and 3 are programmed to the defined HRS range.

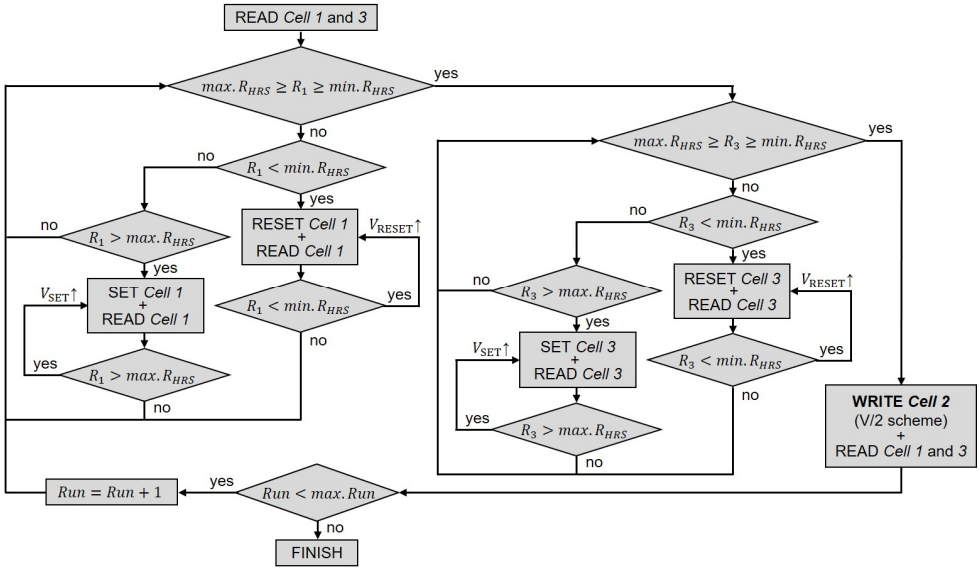

Figure 6.17: *Programmed measurement flow chart of bit-flip assessment for thermal crosstalk study. The measurement script is available for both thermal crosstalk and reference scenarios with the only difference on whether or not the V_{WRITE} is applied on cell 2 during the WRITE operation after resistance state verifications on cell 1 and 3.*

Given the defined HRS range from 30 kΩ to 50 kΩ, the V_{WRITE} of 1 V, and the pulse length of 10 seconds, the programmed measurements were performed on the fabricated 1×3 array of TaO_x RRAMs with the actual device spacing of 70 nm. As mentioned in the testing strategy,

the cell 2 was programmed, in advance, to the LRS of around 1.8 kΩ, corresponding to the loading current of 0.56 mA in average when applying the V_{WRITE} at 1 V. After applying the long single pulse, the resistances of the cell 1 and 3 are described as probability density functions in Figure 6.18, where 25 samples were taken respectively for the reference scenario, shown in blue, and the scenario with thermal crosstalk, shown in red. When the resistance falls below the defined HRS after applying the long single pulse, then it is recognized as a bit flip, which is marked by the grey regions in Figure 6.18.

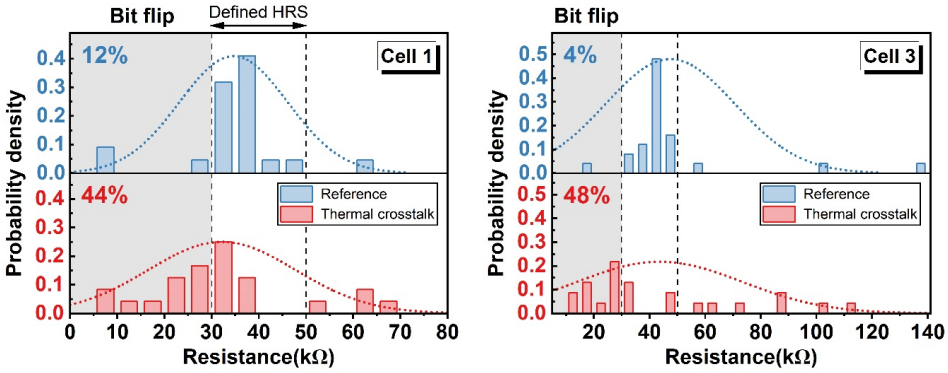

Figure 6.18: *Probability density of resistances monitored from cells 1 and 3 after WRITE operations of cell 2 in thermal crosstalk and reference scenarios. The measurement was carried out from a 1×3 crossbar array with electrode spacing of 70 nm. The HRS was defined in a range of 30 kΩ to 50 kΩ. With V_{WRITE} set at 1 V, the write pulse had a length of 10 seconds. For both scenarios, the statistics based on 25 samples.*

For the reference scenario, the cell 1 and 3 have only 12 % and 4 % of cases respectively showing the bit flips, which indicates the long single pulse of $V_{WRITE}/2$ can only switch the device rarely due to the C2C variability of $V_{SET,C}$. The cell 1 showed a higher bit-flip probability because it has the $V_{SET,C}$ generally lower than the cell 3, according to the previous characterization shown in Figure 6.14. On the other hand, few cases were found with the resistances ending up higher than the defined HRS, which can be explained by the unstable HRS that has the conductive filament favorably changing towards the geometry showing a higher resistance, when the $V_{WRITE}/2$ is applied.

For the scenario with thermal crosstalk, the bit-flip probability increases dramatically above 40 % for both the cell 1 and 3, which clearly points out the impact of thermal crosstalk arising from the cell 2. Hypothetically, the high current loading on the cell 2 generates heat by Joule heating, and the word line formed by the bottom electrode conducts the heat to the adjacent cells over a spacing of 70 nm with a heat loss. In the adjacent cell, the temperature increase enhances the mobility of oxygen vacancy inside the TaO_x layer and promotes the redox reaction at the Ta/TaO_x interface for the formation and annihilation of oxygen vacancies. As a result, the $V_{WRITE}/2$ over the adjacent half-selected cell can switch the cell to a lower resistive state. Besides the bit-flip cases, the cell 1 and 3 both showed some cases with the resistances ending up higher than the defined HRS, of which the probabilities were higher, compared to the reference scenario. This can also be explained by the thermal crosstalk effect. When the temperature increase due to the thermal crosstalk cannot make the $V_{WRITE}/2$ sufficient to switch the adjacent half-selected cell to a lower resistance state, the enhanced mobility of oxygen vacancy at the elevated temperature still makes the current HRS relatively unstable. Compared to the reference scenario, more cases were, therefore, found with the unstable HRS

that has the conductive filament favorably changing towards the geometry showing a higher resistance, when the $V_{WRITE}/2$ is applied.

Overall, the probability of resistance falling into each region is summed up and summarized in Table 6.2 to further discuss the probability change due to the thermal crosstalk. In principle, the cell 1 and 3 were half-selected in the 1×3 array, so they were supposed to be kept in the HRS after a WRITE operation on the cell 2, implying their resistances should stay within the defined HRS range. For the reference scenario, the probability of resistance retaining within the defined HRS range is above 80 % for both the cell 1 and 3. However, it is decreased remarkably below 40 % and even down to 20 % for the cell 3, when the thermal crosstalk comes into effect by applying the V_{WRITE} on the cell 2. This is already a retention failure rate that cannot be tolerated in the array operation. Moreover, the probabilities of resistance falling below and above the defined HRS range are both increased owing to the thermal crosstalk effect. More specifically, the resistance has a higher probability to fall below the defined HRS range than above it, because the $V_{WRITE}/2$ applied across them provides a positive electric field leading to a resistive switching tendency towards a lower resistance state, facilitates bit-flip errors during the WRITE operations.

Table 6.2: *Percentage distribution of resistance in different intervals monitored from cells 1 and 3 after WRITE operations of cell 2 in thermal crosstalk and reference scenarios.*

Scenario	**Cell 1**			**Cell 3**		
	below	defined HRS	above	below	defined HRS	above
Reference	12 %	84 %	4 %	4 %	84 %	12 %
Thermal crosstalk	44 %	36 %	20 %	48 %	20 %	32 %

To conclude the thermal crosstalk effect in the 1×3 array with the device spacing of 70 nm, the WRITE operation by 10-second pulses under the V/2 scheme has shown a significant retention failure rate of 72 % in average on the adjacent half-selected cells. Depending on the WRITE operation intending to trigger either a SET or RESET process on the fully-selected cell, the retention failure could have higher probability resulting in either a lower or a higher resistance state because of the bias polarity over the half-selected cell. Accordingly, it has been shown in this work that the positively biased half-selected cells originally in the HRS showed the average bit-flip probability of 46 % among the average retention failure rate of 72 %. The thermal crosstalk effect can vary along with changes of the device spacing and the pulse length. The delay time between pulses also needs to be considered when the temporal thermal effect arises at shorter pulse lengths. Nevertheless, this work has highlighted the potential thermal crosstalk effect in nanoscale RRAM crossbar arrays in terms of retention failure, which can be a challenge for implementing high-density RRAM arrays.

Chapter 7
Conclusion and outlook

The scope of this work focuses on the BEOL integration of active RRAM arrays for applications in CIM and NC. Two distinct technology platforms of active RRAM arrays were developed respectively on micro- and nanoscale at HNF in Research Center Jülich. The microscale technology platform was built on SOI substrates and enabled by UV photolithography. In collaboration with the PGI-9, planar HKMG MOSFETs were fabricated in the FEOL with the channel length of 2 μm. The plug-type TaO_x RRAMs were integrated in the BEOL to realize active RRAM arrays up to 16×16 in 1T-1R and 1T-nR configurations. Following the electrical characterization of 1T-1R unit cells, CIM was implemented in a 1T-nR line array as an IMP stateful logic gate, where the D2D variability of RRAM devices were also studied. In collaborations with the PGI-14 and Hewlett Packard Labs, the nanoscale technology platform is based on monolithic integration of RRAMs, enabled by EBL, with CMOS circuitry taped out with TSMC 180 nm technology node. With crossbar RRAM devices (100×100 nm^2) integrated in the BEOL, the 64×64 1T-1R array was designed to realize a DPE, a brain-inspired energy-efficient AI accelerator, together with on-chip signal amplifiers and driving/sensing circuitry. After investigating the bipolar resistive switching characteristics of integrated RRAMs, chip packaging was followed by verification on a customized operating hardware. In addition to the active RRAM array platforms, nanoscale passive crossbar arrays were designed and fabricated to study the thermal crosstalk in high-density TaO_x RRAM arrays. In summary, this chapter provides an overview of research outcomes and discusses future prospects for this work.

7.1 Process technology development

The planar HKMG SOI MOSFETs showed a significant improvement in the transfer characteristic after post-passivation annealing in the ambient of forming gas at 380°C for 10 minutes. The annealing process prompts defect annihilation in the gate dielectric, which passivates the charge-trapping defects. Consequently, the built-in potential is mitigated by the reduction of trapped charges in the gate dielectric. Due to the decrease in the built-in potential, the OFF current is suppressed, and the overall subthreshold behavior is improved.

In the integration of plug-type RRAMs and HKMG SOI MOSFETs, patterning contact holes with an area of 2×2 μm^2 proved to be a significant challenge in contact lithography using the i-line (365 nm) of Hg lamp. Owing to light interference under the photomask during exposure, distortion of pattern corners were found inevitable. To avoid this inherent drawback inherent drawback associated with photomasks, a maskless aligner was introduced to perform the exposures through a direct writing approach, which ensures a high fidelity of the contact hole pattern. Nonetheless, when transferring the photoresist pattern to the substrate, a reduced etching rate in the contact holes was observed due to the microloading effect in RIE using CHF_3. To ensure clean exposure of NiSi drain contacts, the RIE duration was prolonged to

compensate the reduced etching rate, and a post-RIE wet etching step in 1% HF solution, was introduced.

When EBL and RIBE were utilized for patterning in nanoscale RRAM fabrication, the presence of thin fences surrounding the patterns were found to severely affect the device yield. This fencing issue stemmed from sidewall redeposition during Pt etching via the Ar-based RIBE process. Resist stripping in DMSO and resist ashing in O_2 plasma proved ineffective in removing the fences because of their composition, comprising inorganic and organic products from the etching of material stack and resist respectively. Accordingly, effective fence removal was demonstrated by introducing an additional CF_4-based RIBE process at a tilted angle prior to the resist stripping. In addition, a fence-free RIE process with Cr hard masks using a gas mixture of Cl_2 and Ar was developed to avoid significant fencing during the Ar-based RIBE process. Substituting resist masks with Cr masks reduced mask thickness and etching selectivity, which lowers fence height during the etching process. Besides, the use of Cr masks and the inclusion of Cl_2 in the process gas both altered the redeposit composition, which enhances the solubility in a Cr etchant comprising a mixture of perchloric acid and ceric ammonium nitrate. Therefore, the fences can be eliminated simultaneously during the Cr mask removal by using the Cr etchant.

When downscaling device spacing to achieve high-density RRAM arrays, pattern enlargement effect due to non-ideal undercut in resist profile for lift-off patterning and the proximity effect in EBL becomes prominent. To counter this effect, the actual e-beam exposure area was shrunk accordingly. In the patterning process with a Cr hard mask, it was observed that RIBE of Ta and TaO_x using CF_4 exhibited a more pronounced profile tapering effect than RIE of Pt using Cl_2 and Ar. Moreover, shrinking device spacing led to reduced tolerance to within-wafer process variability. Ultimately, device spacing of 70 nm was achieved in passive TaO_x RRAM arrays, with device sizes approximately 100×100 nm^2, by using the EBL alongside the fence-free etching technique.

7.2 Integration of active RRAM arrays on SOI substrates

While NiSi is a material extensively used for enabling ohmic contacts on the S/D terminals of transistors, it was deemed unsuitable as the BE material for TaO_x RRAMs in comparison to Pt, which is widely recognized for its efficacy as the high-work-function electrode of VCM-type RRAMs. The high reactivity of NiSi with oxygen, owing to its relatively low activation energy, results in the formation of a thin interfacial oxide layer, approximately 2.5 nm thick, between the TaO_x layer and the NiSi layer in a TaO_x RRAM device with NiSi BE. Consequently, abrupt RESET processes and large variabilities of $V_{SET,C}$, $V_{RESET,C}$, R_{LRS} and R_{HRS} were observed in the resistive switching characteristics with a lack of consistency. To ensure superior resistive switching performance when integrating TaO_x RRAMs onto NiSi drain contacts, maintaining Pt as the BE material is recommended over directly employing NiSi.

When integrating SOI MOSFETs and RRAMs for active RRAM arrays, the amount of layers to be stacked in the fabrication increased. Potential overlay issues emerged consequently were addressed. Concerning alignment strategy, the use of a constant baseline is advantageous, although relatively large misalignment tolerances must be taken into account. In addition, consideration of pattern size disparity between the physical layout and the fabrication result is critical in the overlay design. Notably, incorporating overlay tolerance into the layout design can limit the pitch size minimum of certain patterns.

7.3 Resistive switching characteristics of integrated RRAMs

By integrating a TaO_x RRAM in a 1T-1R unit cell, the bipolar resistive switching exhibits a smaller variability compared to the same RRAM device in 1R configuration. In the 1T-1R unit cell, the output characteristic of transistor internally offers a fast and effective control of output current, which makes the current increase in a SET process more gradual and consequently alleviates the current overshoot. In addition, the reduced current overshoot shows a low variability, which reflects on the R_{LRS} and R_{HRS} variabilities. Overall, in comparison to the 1R configuration, the transistors employed in active RRAM arrays not only mitigate the sneak path issue, but also improve the current overshoot control, which enables the reliable and consistent resistive switching property.

Multi-level resistive switching was demonstrated by TaO_x RRAMs integrated in 1T-1R unit cells through $V_{g,SET}$ and $V_{RESET-stop}$ modulations respectively. For the $V_{g,SET}$ modulation, the R_{LRS} decreases with increasing $V_{g,SET}$, because the enhanced saturation current of transistor increase the equivalent current compliance, allowing a thicker conductive filament to be formed during the SET process. To initiate the subsequent RESET process, a high $V_{RESET,C}$ can be expected from the relatively thick filament. For the $V_{RESET-stop}$ modulation, the R_{HRS} can be raised by increasing $|V_{RESET-stop}|$, which leads to a larger maximum electric stress applied for the filament rupture in a RESET process. As a result, a larger spatial gap left between the filament tip and the BE causes a higher R_{HRS}, and a larger $V_{SET,C}$ was found in the succeeding SET process. By comparing the multi-level switching modulated by $V_{g,SET}$ and $V_{RESET-stop}$, the elevation of $V_{RESET,C}$ when increasing $V_{g,SET}$ is more sensitive to that of $V_{SET,C}$ when increasing $|V_{RESET-stop}|$. Nonetheless, a larger R_{HRS} variability was observed at a higher $V_{RESET-stop}$, which appears to be the constraint of $V_{RESET-stop}$ modulation on multi-level resistive switching.

Lastly, the effect of $V_{g,FORM}$ on the resistive switching characteristics was studied from TaO_x RRAMs integrated in 1T-1R unit cells. Under a constant $V_{g,SET}$, a primary decreasing trend of R_{LRS} with increasing $V_{g,FORM}$ was determined, which is explained by the highly-conducting working condition for the conductive filament as a result of electro-forming at a high $V_{g,FORM}$. The electro-forming process involves not only redox reactions that introduce a critical amount of oxygen vacancies in the TaO_x layer, but also oxygen vacancy migration that forms the first complete filament. By applying a high $V_{g,FORM}$, a high current compliance during the electro-forming process allows an extensive introduction of oxygen vacancies, and enables the formation of a strong filament in terms of structure geometry. Notably, the electro-formed condition changes dramatically once the current compliance exceeds a critical value, according to the observation of an abrupt R_{LRS} drop when increasing $V_{g,FORM}$. Following the negative correlation of R_{LRS} and $V_{g,FORM}$, $|V_{RESET,C}|$ also shows a primary decreasing trend with increasing $V_{g,FORM}$. This suggests that the critical RESET condition of a relatively low LRS can be fulfilled at a low $V_{RESET,C}$, when a highly-conducting working condition for the conductive filament is introduced by a high $V_{g,FORM}$. Furthermore, when the $V_{RESET-stop}$ is fixed, the relatively low LRS resulted from a high $V_{g,FORM}$ is switched to a relatively low HRS under the constant electrical stress in the RESET process, and a low $V_{SET,C}$ was found in the succeeding SET process. Overall, $V_{g,FORM}$ is concluded to be a decisive factor on the resistive switching property, because it can determine the oxygen vacancy concentration and the conductive filament geometry after the electro-forming process, which is the initial working condition for the subsequent RESET and SET processes.

7.4 Applications in computing-in-memory and neuromorphic computing

To achieve logic operations in a CIM approach, an IMP stateful logic gate was experimentally demonstrated by a 1T-nR line array based on TaO_x RRAMs. The transistor replacing the fixed resistor in a stateful logic gate provides flexible tuning of the series resistance by V_g, and improves current overshoot control during SET processes in the RRAM cells when performing logic operations. By designing the operating voltages, the functionality of IMP stateful logic gate was successfully verified for all four input cases. Furthermore, as the IMP stateful logic gate is classified as a SET-based stateful logic gate, the impact of D2D variability in the logic operations was investigated with the focus on the SET process. In principle, the bit flip after an IMP logic operation only appears in the input case (p/q=0/0), where the q bit switches from the HRS to the LRS through a SET process. Additionally, assigning the RRAM cell with the lower $V_{SET,C}$ as the q bit can facilitate the bit flip in the q bit, while impeding the bit flip in the p bit. Therefore, this arrangement results in the success rate close to 70 %, while the opposite case shows the success rate of 0 %, as demonstrated by experiments. This suggests that the inherent D2D variability in the RRAM array can be exploited to enhance the success rate in the IMP stateful logic operations.

To realize NC by employing the RRAM-integrated CMOS die, it was packaged on a ceramic lead frame after monolithic integration of TaO_x RRAMs in the BEOL. To drive the packaged chip with on-chip signal amplifiers and driving/sensing circuitry, it was connected to a customized operating hardware that had been built on a PCB that communicates externally through a microcontroller [147]. By executing computer commands via a Python interface, the hardware successfully demonstrated bipolar resistive switching on the packaged chip, which verifies the functionality and paves the way to realizing NC applications.

7.5 Crosstalk in RRAM arrays

For the investigation of thermal crosstalk effect in passive nano-crossbar arrays of TaO_x RRAMs with the device spacing of 70 nm, the experiment was designed to emulate the scenario of WRITE operations under the V/2 biasing scheme. The statistics from programmed electrical measurements indicate the thermal crosstalk causes a retention failure rate over 70 % in average on the half-selected RRAM cells adjacent to the fully-selected cell, where the SET process takes place. As a high loading current on the fully-selected cell during the WRITE operation induces Joule heating, the generated heat can be transmitted through the metal interconnect to the adjacent cells. The consequent temperature increase in the adjacent cells prompts the critical retention failure. Furthermore, the average bit-flip probability of 46 % was identified from the average retention failure rate of 72 %, when the half-selected cells were originally in the HRS. This suggests a tendency of the retention failure leading to a lower resistance state, which can be explained by the positive bias applied on the half-selected cells. As a result, the retention failure on the half-selected cells can be concluded that the tendency for the resistance to increase or decrease hinges on whether the WRITE operation intended to trigger a SET or a RESET process on the fully-selected cell.

7.6 Outlook

To improve the process reliability in the microscale RRAM technology platform, a RIE process with negligible microloading effect can be developed to etch the contact holes.

Considering the process gas, a low ratio of O_2 can be introduced, and the C/F ratio can be lowered by mixing the original process gas, CHF_3, with CF_4, for example. Besides, reducing the process temperature and pressure can favorably alleviate the polymerization and release the desorption limitation of volatile etching products during the RIE process respectively.

To apply the integrated 1T-1R arrays in spiking neural networks for NC applications, biological synaptic plasticity should be demonstrated on the RRAMs in the 1T-1R configuration. As $V_{g,FORM}$ is concluded to be a decisive factor on the bipolar resistive switching property, it can be investigated to tailor the electro-forming process for obtaining the optimum properties of multi-level switching and synaptic plasticity.

For the NC hardware configured with the RRAM-integrated chip, the resistance programmability of the integrated RRAM devices need be demonstrated to spontaneously tune the synaptic weight in a neural network that are stored as the RRAM conductance in an active RRAM array. Therefore, supervised learning by backpropagation in an artificial neural network can be realized for NC in AI applications such as image recognition, handwritten digit classification, solving optimization problems, etc.

For the 1T-2R IMP stateful logic gate enabled by the microscale RRAM technology platform, different logic operations can be explored by using the 1T-nR line arrays. For instance, two RRAM cells or more can be employed simultaneously or sequentially, and the series resistance can also be tuned accordingly by the V_g applied on the transistor.

Concerning the study of thermal crosstalk in high-density RRAM arrays, the device spacing can be reduced below 70 nm potentially by exploiting multiple patterning, where a single layer of pattern is formed by patterning multiple times. In addition, the impact from the thermal conductivity of metal electrode can be investigated. For instance, a highly-thermal-conducting layer, such as an Al layer, can be stacked on top of the Pt layer of TE, to improve the heat dissipation, which can potentially mitigate the thermal crosstalk effect. Lastly, an investigation of temporal effect in the thermal crosstalk can be carried out by using multiple short pulses in the experiments, where the delay time between pulses would introduce a cooling effect.

Appendix
Development of $SiTe_x$ binary chalcogenides for ovonic threshold switching selectors

To realize passive RRAM arrays in 1S-1R configuration, a preliminary work on developing OTS selectors based on $SiTe_x$ binary chalcogenides was carried out. Study on thin-film deposition of $SiTe_x$ is presented first in this appendix, and the device fabrication process is then introduced with an encountered process issue. Electrical characterizations ensues to demonstrate and study OTS behaviors of the fabricated $SiTe_x$ selectors.

A.1 Review of chalcogenide OTS selectors

Passive crossbar arrays are the simplest configuration of RRAM arrays. To address the sneak path issue in the passive crossbar arrays, 1S-1R array configuration is a promising solution. By connecting a two-terminal selector device to each RRAM in series, the electrical access to each RRAM can be better controlled. Among several types of selectors, OTS selector stands out mainly in two aspects: high non-linearity and high current.

The OTS behavior was initially observed in alloys of chalcogenide semiconductors and reported by S. R. Ovshinsky in 1968 [172]. He found reversible electrical switching phenomena in the I-V characteristic shown in Figure A.1. Between a threshold voltage (V_{th}) and a holding voltage (V_h), a rapid and unsteady behavior of negative differential resistance (NDR) leads to two steady resistance states, namely OFF (high-resistance) and ON (low-resistance) states. Originally, with electrical biasing below the V_{th}, the resistance is relatively high, indicating the OFF state. When a voltage exceeding the V_{th} is applied, the resistance state switches from OFF to ON, allowing high driving current. When the applied voltage is descending below the V_h, a transition from the ON state to the OFF state takes place, resulting in a fully reversible process. The OTS selector has a large resistance ratio of OFF and ON state, which can eliminate the sneaking current in the OFF state and supply a high current for write operation in the ON state.

Considerable research efforts have been dedicated to elucidating the OTS behavior in chalcogenide thin films. Besides some thermal-effect-oriented theories [173-175], the nature of threshold switching was predominantly recognized to be electronic. In 1970s, Alder et al. brought out a field-induced carrier generation model as one of the first electronic models for OTS, which agrees with the Ovshinsky's original point of view [176, 177]. Pirovano and Redaelli et al. later investigated further with experiments and simulations to support and improve the model [178, 179]. In addition, Ielmini et al. proposed an analytical electrical transport model based on trap-limited conduction, which can be applied on both subthreshold conduction and threshold switching in chalcogenide memory devices [180, 181]. Apart from discussions on electronic processes, Karpov et al. proposed a unified field-induced nucleation

model to explain both the PCM and OTS behaviors in chalcogenide materials by describing the electrical switching in terms of nucleation [182-185].

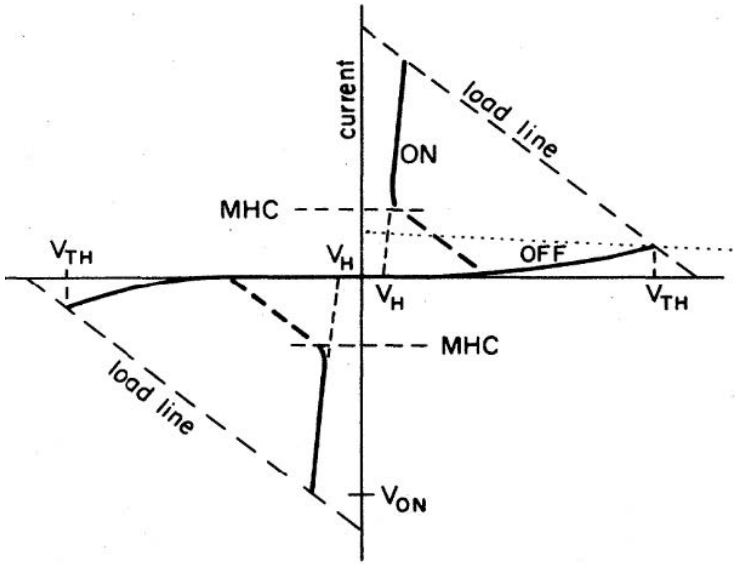

Figure A.1: *I-V characteristic of OTS behavior [172].*

Amorphous chalcogenide has been acknowledged as a prominent OTS material system, well-suited for implementing selector devices. Alongside the volatile characteristic of OTS, it also exhibits a nonvolatile switching characteristic, extensively utilized in PCM, owing to localized crystallization induced by Joule heating. Hence, realizing an OTS selector necessitates suppression of the crystallization effect in the amorphous chalcogenide during threshold switching, particularly when operating under high currents in the ON state. Since the quantity of crosslinks and their bonding energies in a chalcogenide material system can differentiate between the PCM and the OTS devices, the material composition plays a crucial role [186, 187]. The amorphous chalcogenides containing robust and numerous crosslinking atoms such as Si and Ge exhibit higher crystallization temperatures and, consequently, greater thermal stability [186, 188]. The resulting structural resistance to crystallization in chalcogenide materials tends to impede the nonvolatile PCM behavior while retaining the OTS behavior. As several Te-based chalcogenides reported with good potentials in implementing OTS selectors, the element Te was targeted in this work [186, 189-197]. In addition, experiments [198] and calculations [199] have suggested that $SiTe_x$, in comparison to $GeTe_x$, shows better selector properties in terms of low OFF-state current and optimum V_{th} due to several advantages of Si over Ge such as process compatibility, thermal stability, large bandgap, high resistivity, and stable bonding characteristics with Te [200, 201]. Consequently, the $SiTe_x$ binary chalcogenides were targeted in this work as the material system of interest for implementing OTS selectors. As the material composition significantly affects physical and electrical properties of the $SiTe_x$, the composition tunability through thin film deposition process was investigated, and later applied on selector device fabrication for performance assessment.

A.2 Deposition of $SiTe_x$ thin films

Prior to selector device fabrication, deposition of $SiTe_x$ thin films was studied in detail. This subchapter will first explain the deposition method employed in this work. Then, investigations of deposition rate and film roughness will be discussed before verification of $SiTe_x$ composition tuning by RF power of sputter deposition.

A.2.1 Deposition method

The $SiTe_x$ thin film deposition was carried out by RF sputtering, which is a physical vapor deposition relying on field-induced plasma generated by RF voltage biasing. The sputtering chamber is configured with a single sputtering target. To deposit $SiTe_x$ thin films by a single target, a 2-inch compound target from EVOCHEM Advanced Materials GmbH was used. Specifically, it was produced in a composite form of $SiTe_6$, implying 14 % of Si and 86 % of Te in molar percentage.

In the typical RF sputtering procedure, the sputtering chamber is initially supplied with Ar gas at a constant mass flow of 60 sccm for 1 minute, serving as the standard gas inlet time to stabilize chamber pressure. Subsequently, the plasma is ignited and maintained at room temperature using an RF power generator and control electronics. Initially, a shutter positioned between the target and the substrate remains closed to prevent deposition on the substrate, and facilitate a pre-sputtering phase lasting 1 minute for plasma stabilization. The deposition process begins only when the shutter is opened, as depicted in Figure A.2, and terminates after the desired deposition duration, achieved by deactivating the RF power and closing the shutter simultaneously.

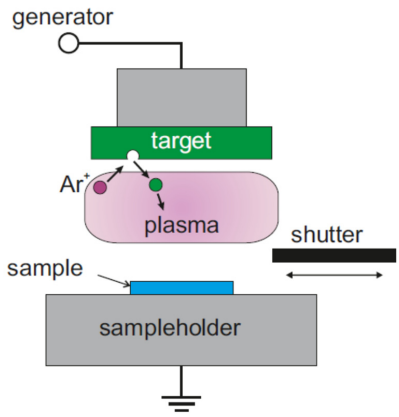

Figure A.2: *Schematic diagram of a sputter deposition chamber configured with a shutter [148].*

The sputter deposition system encompasses several sputtering chambers, which offers a range of sputtering materials including $SiTe_6$ and electrode metals. Thus, electrode metals like W can be deposited in-situ onto a $SiTe_x$ film as deposited recently. Additionally, deposition of a capping layer, such as Pt, can follow in-situ by choice to passivate chemically-active electrodes. In essence, the versatility and adaptability of this sputter deposition system pose significant advantages practically on the fabrication process of OTS selectors.

A.2.2 Deposition rate and film roughness

Preliminary efforts are necessary to establish a fundamental understanding of the SiTex deposition for controlling film thickness and roughness. Initially, a series of depositions were performed by varying only the deposition time. The pressure was constantly kept at 2.4×10^{-2} mbar with a gas inlet of Ar set at 60 sccm, and the RF power was fixed at 20 % of the total available power, i.e., 116 W. Following the deposition of $SiTe_x$ thin film onto a bare Si substrate, film thickness was analyzed using X-ray reflectometry. Figure A.3 illustrates the relationship between film thickness and deposition time, spanning from 10 to 200 seconds. Upon a closer examination, a non-linear relationship can be recognized, indicating that the deposition rate varies during the deposition in practice.

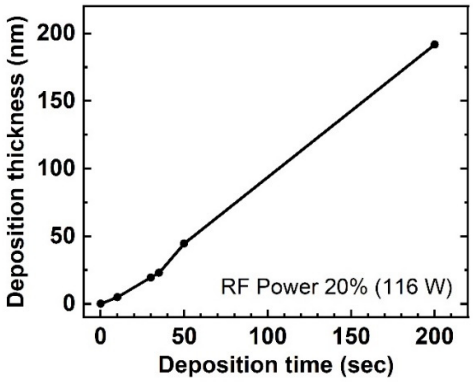

Figure A.3: *Diagram of deposition thickness of $SiTe_x$ film versus deposition time.*

A time dependence of deposition rate is revealed in Figure A.4a, where the average deposition rate is plotted against the deposition time. The average deposition rate prominently escalates approximately in the initial 50 seconds and gradually saturates at around 55 nm/min, which is almost twice as much as 28.8 nm/min at 10 seconds. Even though the average deposition rate does not represent the transient behavior, a similar time dependence can be generally deduced for the transient deposition rate. A rapid increase followed by a gradual saturation strongly relates to an increasing growth with exponential decay, which is depicted in Figure A.4b. The simplest mathematic expression of the transient deposition rate (R_{dep}) as a function of the deposition time (t) can be given by

$$R_{dep}(t) = Ae^{-\frac{t}{t_0}} + R_{dep,sat}\ , \tag{AP.1}$$

where $R_{dep,sat}$ indicates the saturated deposition rate, and both A and t_0 are coefficients. In Figure A.4b, a typical increasing function $R_{dep}(t)$ is shown, when $A < 0$ and $t_0 > 0$.

The time-dependent deposition rate can be attributed to a plasma stabilization process during the deposition process. As detailed in the Section A.2.1, the typical deposition procedure includes a pre-sputtering step to stabilize the plasma before the shutter opens for the actual deposition on the substrate. However, there exists a spatial gap between the shutter and the substrate. Once the shutter is opened, the plasma must fill this gap space, requiring a certain amount of time for stabilization depending on the gap size. As a result, the actual deposition begins in a non-equilibrium state before entering a steady state, which results in the deposition rate increasing from the start and eventually reaching a saturation as illustrated in Figure A.4b.

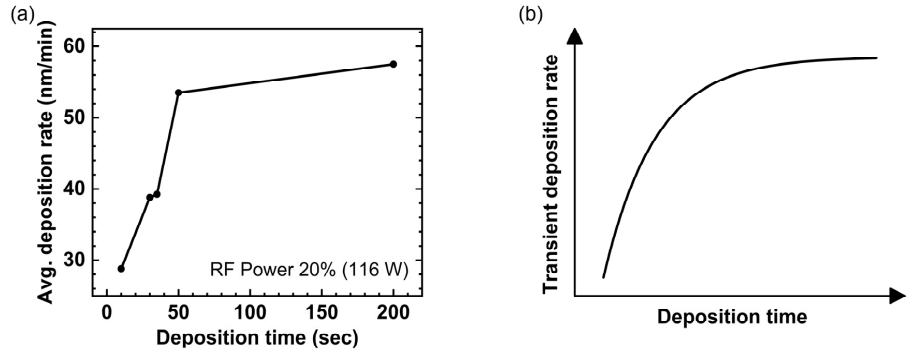

Figure A.4: *Change of deposition rate over deposition time. (a) Experimental data of average deposition rate versus deposition time. (b) Illustrative diagram of transient deposition rate increasing over deposition time with exponential decay.*

Considering the overall deposition result, the impact of this initializing state depends on the total deposition time and the time required to reach the saturated deposition rate. Taking the example shown in Figure A.4a, the average deposition rate starts to saturate approximately at 50 seconds, when the impact of the initializing state becomes more and more negligible. Thus, only when the deposition time is longer than 50 seconds, the deposition process is dominated by a more stable condition, which ensures reproducible deposition results. In addition, the average deposition rates are comparable and can be used to estimate the deposition time of a desired thickness. As a consequence, there is a minimum thickness (d_{min}) by which controllable and reliable results can be ensured. As shown in Figure A.4a, d_{min} is approximately 44.6 nm corresponding to the deposition time of 50 seconds. For the deposition with a higher saturated deposition rate, a larger d_{min} is expected. If a longer saturation time also comes along, the d_{min} would be even larger. As the film thickness of interest was set to be 40 nm in this work, any deposition with deposition rate higher than the current case is not suitable for study, when changing the deposition parameters.

For a fixed film thickness of 40 nm, a deposition study on RF power was conducted by keeping other deposition parameters constant (e.g. gas species, temperature, pressure, pre-sputtering time, etc.). Varying the RF power from 5 % to 20 %, the average deposition rate and

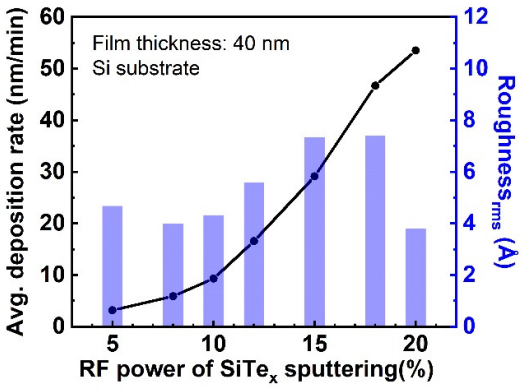

Figure A.5: *Diagram of average deposition rate and root-mean-square film roughness versus RF power for sputter deposition of 40nm-thick SiTe$_x$ film.*

film roughness are summarized in Figure A.5. The average deposition rate shows a strict increase from 3.2 nm/min to 53.5 nm/min as the RF power goes from 5 % up to 20 %. However, no discernible correlation was observed between the RF power and film roughness based on AFM measurements. With the root-mean-square roughness smaller than 7.5 Å, all deposited films exhibit low roughness relative to the thickness of 40 nm.

In typical selector device fabrication, the $SiTe_x$ thin film needs to be positioned between two electrodes, a BE and a TE, by vertical material stacking, implying that the $SiTe_x$ thin film must be deposited onto the BE. Therefore, the roughness of $SiTe_x$ film deposited on difference electrode materials such as Pt and W was investigated with a constant RF power of 10 % while other deposition parameters kept fixed. Figure A.6 illustrates the roughness disparities before and after depositing 40-nm-thick $SiTe_x$ films on different substrates. The Pt substrate comprises 30-nm-thick Pt atop a 5-nm-thick Ti adhesion layer, both in-situ sputtered onto a prepared SiO_2 substrate. The W substrate, on the other hand, features 30-nm-thick W directly sputtered onto a similar SiO_2 substrate. The SiO_2 substrate is prepared by thermal oxidation, producing a 430-nm-thick SiO_2 layer on top of a Si wafer. Before the $SiTe_x$ deposition, the Pt and W substrates exhibit slightly higher roughness than the Si substrate. Interestingly, regardless of the substrate used, the roughness decreases after the deposition. Overall, the investigation on film roughness suggests a smooth $SiTe_x$ film surface, which ensures the quality of material stacking during device fabrication.

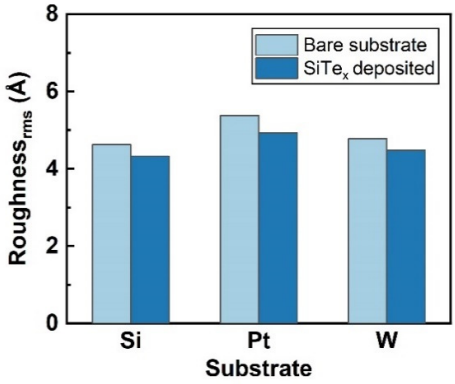

Figure A.6: *Roughness comparison before and after sputter deposition of 40-nm-thick $SiTe_x$ film on Si, Pt, and W substrates.*

A.2.3 Composition tuning by RF power of sputter deposition

Concerning RF sputtering with the $SiTe_6$ compound target, which has a fixed composition, this work aims to tune the composition of deposited film by RF power variation. Utilizing Rutherford backscattering spectrometry (RBS), a composition analysis was conducted on 40-nm-thick $SiTe_x$ films deposited under three different RF powers. All pertinent parameters for sample preparation are detailed in Table A.1. Notably, signals detected from the substrate can influence the analysis of the deposited $SiTe_x$ film, particularly when a Si-containing substrate is used. To mitigate this interference, an intermediate isolating layer was introduced between the $SiTe_x$ film and the substrate. In addition to the thickness that can sufficiently separate the $SiTe_x$ film away from the substrate, the material selection is crucial to have the energy peaks in the RBS spectrum distinct from both Si and Te peaks for ensuring accurate data analysis on the $SiTe_x$ film composition. Given the capabilities of the sputter deposition system used in this work, a Ti film with a thickness around 50 nm was chosen for this purpose. Conclusively, the

Table A.1: *Specifications of sample preparation for RBS analysis.*

Sputter deposition parameters	RF Power % (W)	5 % (26W)	10 % (58W)	20 % (116W)
	Deposition Time (sec)	780	220	50
	$SiTe_x$ Thickness (nm)	41.0	37.9	44.6
	Temperature	Room temperature		
	Pressure (mbar)	2.40×10^{-2}	2.41×10^{-2}	2.41×10^{-2}
	Ar Flow (sccm)	60		
Intermediate isolating layer		51.7 nm Ti		
Substrate		430 nm SiO_2 on Si wafer		

RBS analysis results depicted in Figure A.7 confirm the manipulating ability of RF power on the film composition. The Te concentration increases from 73 at.% to 90 at.% as the RF power goes from 5% up to 20%. This observation suggests that the RF power varies the sputtering ratio between Si and Te elements from a fixed $SiTe_6$ compound target. Specifically, higher RF powers yield higher Te ratios.

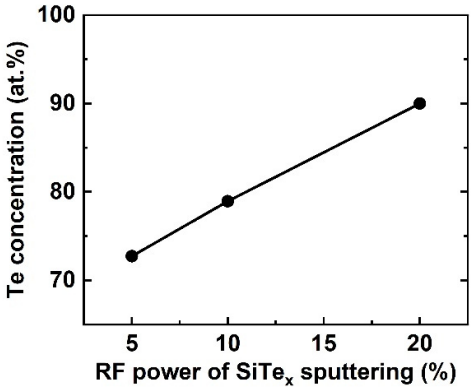

Figure A.7: *RBS analysis result on Te concentration of $SiTe_x$ film versus RF power of sputter deposition. The analysis from measurement to data fitting was supported by Dr. Jürgen Schubert from PGI-9 at Research Center Jülich.*

A.3 Micro crossbar devices of $SiTe_x$ selectors

Following the thin film deposition study on the $SiTe_x$, selector device fabrication proceeded for OTS behavior characterization. This subchapter will first illustrate the device fabrication method and discuss the results of electrical characterizations with comparisons between different $SiTe_x$ compositions achieved by the RF power variation.

A.3.1 Device fabrication

As a two-terminal device, OTS selector shares the same MIM stacking structure with RRAM, where an insulating chalcogenide layer sandwiched between two metal layers serving as electrodes. In this work, the $SiTe_x$ selector device was configured within a crossbar structure

consisting of two patterns, namely a BE and a TE, as displayed in Figure A.8. The BE is patterned first, followed by TE patterning through the $SiTe_x$ layer. Consequently, a functional device is established at the crossing point of the two patterns, where the complete MIM stack is formed. Using W electrodes for both BE and TE, the selector results in a device with symmetric electrodes. As the crossing area determines the actual device area, an electrode width of 2 µm yields a device area of 2×2 µm^2.

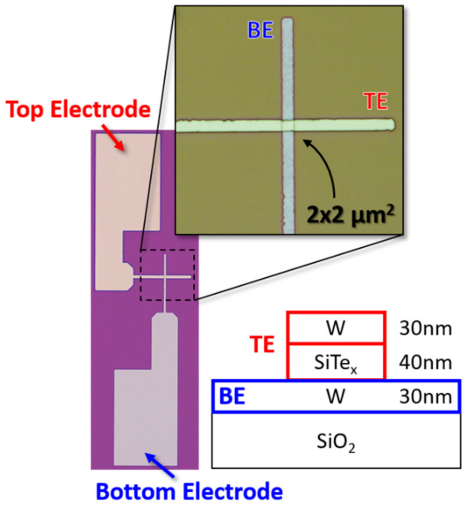

Figure A.8: *Schematic diagram of $SiTe_x$ selector in a crossbar device structure. A 30-nm-thick W layer was patterned first as the BE, and the TE pattern was transferred through the 40-nm-thick $SiTe_x$ from a 30-nm-thick W on the top. The device area of 2×2 µm^2 was determined by the electrode width.*

The device fabrication process began with a thermally oxidized Si wafer with 430-nm-thick SiO_2 on the top. Initially, a 30-nm-thick W layer was deposited via RF sputtering, followed by a photolithography step using the *AZ® MiR™ 701 14cps* positive photoresist to define the BE pattern. Next, RIBE using Ar was employed to transfer the BE pattern to the deposited W thin film. Resist stripping in DMSO at 80 °C ensued for a minimum of 3 hours to remove the photoresist after the RIBE. Subsequently, the TE fabrication deviated depending on the selected patterning technique via either RIBE or lift-off process. Nonetheless, the deposition step stayed constant that was a 40-nm-thick SiTex layer and a 30-nm-thick W layer were in-situ deposited via RF sputtering in separate sputtering chambers.

For the TE patterning via RIBE, the same photolithography process was repeated after the RF sputtering of $SiTe_x$ and W. Following RIBE of the $SiTe_x$ and W layers with Ar, an issue arose during the resist stripping in DMSO, after which severe damages to the TE pattern were observed, as shown in Figure A.9a. A material compatibility test was, therefore, conducted such that a sample with a deposited $SiTe_x$ film was partially immersed in DMSO at 80 °C for 4 hours. The results presented in Figure A.9b revealed a clear contrast between the immersed and pristine surfaces, confirming chemical reactions occurring during the DMSO immersion. In other words, the deposited $SiTe_x$ film exhibited chemical reactivity with DMSO. Conclusively, the compatibility test result suggests that DMSO chemically affected the SiTex layer during the resist stripping, leading to the lifting off of the W layer atop it.

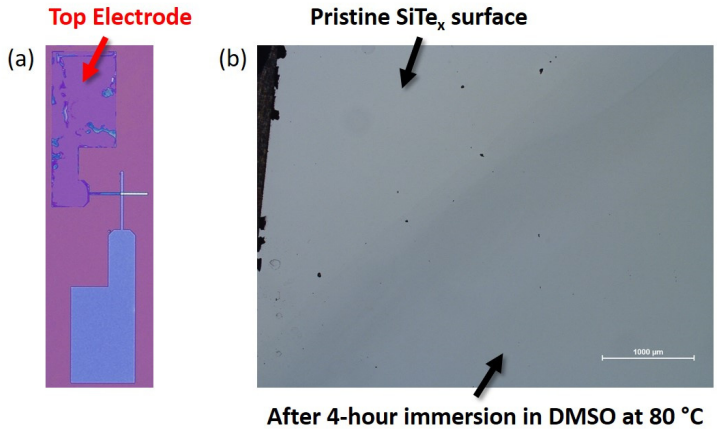

Figure A.9: *Compatibility issue of SiTe$_x$ with DMSO during resist stripping in TE patternig. (a) Severe TE pattern destruction found after resist stripping in DMSO. (b) Result of compatibility test using a SiTe$_x$ sample. Half of the sample was immersed in DMSO at 80 °C for 4 hours after SiTe$_x$ deposition on a Si substrate.*

Avoiding the use of DMSO, acetone was used instead to remove the photoresist after RIBE in TE patterning. Due to the low boiling point and the low stripping rate, 24 hours at room temperature were needed to ensure an appropriate stripping result, which is shown in Figure A.10a. Alternatively, lift-off pattering can be used to substantially avoid the need of resist stripping. In this case, a photolithography step using the *AZ® 5214 E* image reversal photoresist was performed to obtain the TE pattern prior to the RF sputtering of SiTe$_x$ and W layers. The TE pattern was then transferred to the deposited stack by a lift-off step using acetone. The result of TE patterning by lift-off is shown in Figure A.10b.

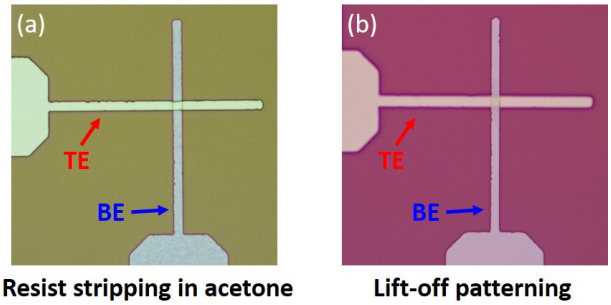

Figure A.10: *Fabrication results of SiTe$_x$ selector devices using different techniques in TE patterning. (a) Device with TE patterned by RIBE with resist stripping in acetone. (b) Device with TE patterned by lift-off.*

A.3.2 Electrical characterization

To investigate the influence of SiTe$_x$ composition, electrical characterization was conducted on three distinct types of OTS selectors, corresponding to 5 %, 10 %, and 20 % of RF sputtering power utilized during the SiTe$_x$ deposition. For devices fabricated using 5 % and 20 % of RF power, the TE was patterned using the RIBE followed by the resist stripping in acetone.

Conversely, for devices fabricated using 10 % of RF power, the lift-off patterning was employed for the TE patterning.

First, the initial resistances (R_{ini}) of pristine devices were measured at a READ voltage (V_{READ}) of 0.5 V and plotted in Figure A.11 versus RF power. R_{ini} can be found increasing from 3 kΩ to 65 kΩ when RF power goes up from 5 % to 20 %. By adding the Te concentration measured previously into the diagram, a higher Te concentration can be concluded with correspondence of a higher R_{ini} [186]. Normally, the ON state of RRAM has resistance in a range of 1 kΩ to10 kΩ. Thus, the matching selector should have an OFF state much higher than this range to effectively prevent the sneak current passing through the unselected RRAM cell in its ON state. For an OTS selector, the initial state is also the OFF state. As 5 % and 10 % of RF power resulted in low R_{ini} of 3 kΩ and 9 kΩ respectively, it can be concluded already at this stage that they cannot match the criteria as selectors in 1S-1R arrays.

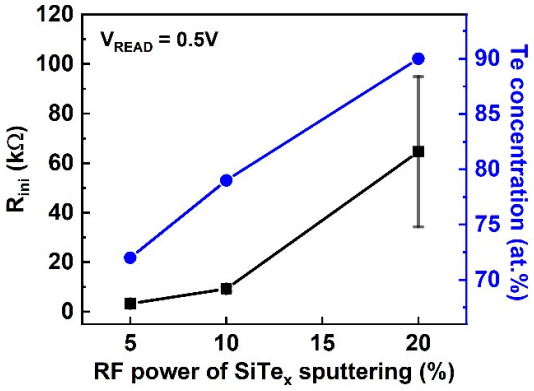

Figure A.11: *Diagram of R_{ini} and Te concentration versus RF power of SiTe$_x$ sputtering. With the R_{ini} monitored by a V_{READ} of 0.5 V, the statistic was based on 10 devices for each RF power.*

Subsequently, the I-V characteristics were investigated with DC voltage sweeping that a forward sweep was directly followed by a backward sweep. For 5 % and 10 % of RF power, linear curves were found in the I-V diagrams shown in Figure A.12a and Figure A.12b, indicating they are essentially normal resistor devices without showing the OTS behavior. In contrast, for 20 % of RF power, the OTS behavior was clearly observed in Figure A.12c. The current showed a sharp increase when the applied voltage reached the V_{th} at 2.4 V. During the backward voltage sweep, a sharp current decrease appeared when the applied voltage hit the V_h at 1.8 V. The resistance successfully switched between the OFF state resistance (R_{OFF}) of 50 kΩ and the ON state resistance (R_{ON}) of 5 kΩ, demonstrating an ON/OFF ratio of 10.

As the typical OTS behavior was found in the device corresponding to 20 % of RF power, cycling of identical DC voltage sweeps was executed to study the C2C characteristic evolution. Figure A.13a illustrates 30 unipolar switching cycles with each cycle drawn in different colors as per the color scale. A considerable trend among the cycles can be observed that the V_{th} and V_h, where the threshold switching took place, shift towards the right of diagram, indicating the operating voltages increase when the switching cycle appends. By extracting the V_{th} and V_h from each cycle, Figure A.13b depicts more clearly that both V_{th} and V_h show rising trends with progressing cycles. However, the relation between them varied such that the V_{th} was not consistently larger than the V_h. Considering the distributions plotted in Figure A.13c, a general observation of $V_{th} > V_h$ can still be concluded with the mean V_{th} of 1.8 V and the mean V_h of

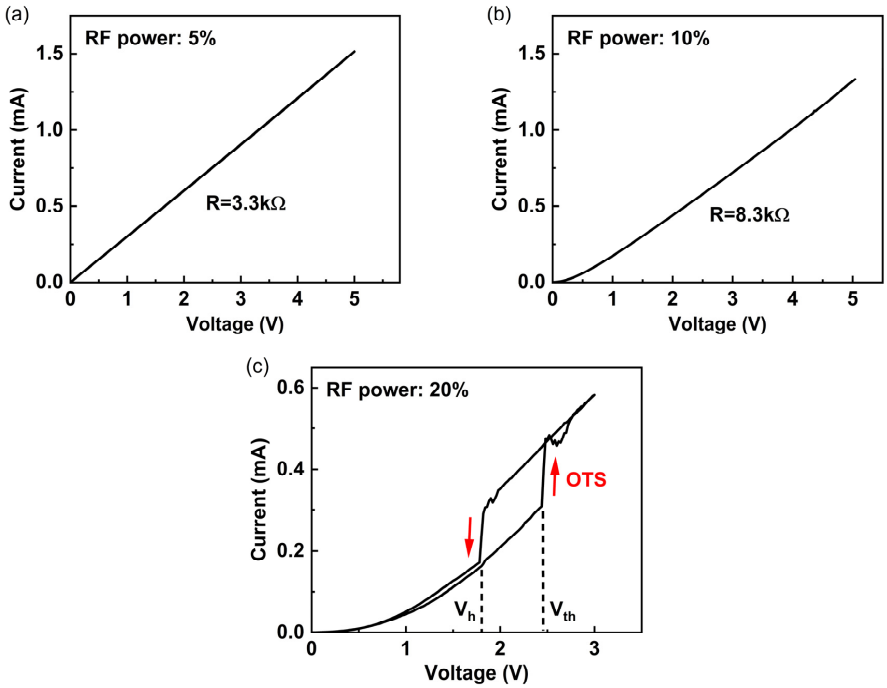

Figure A.12: *I-V characteristics of $SiTe_x$ selectors corresponding to different RF powers. (a)(b) Linear I-V curves showing no OTS behavior for 5 % and 10 % of RF power. (c) I-V characteristic showing the typical OTS behavior for 20 % of RF power. The V_{th} and the V_h are 2.4 V and 1.8 V respectively.*

1.7 V, despite a large distribution overlapping. In addition to the operating voltages, the R_{OFF} and R_{ON} were monitored respectively at 1 V and 2.5 V over 30 cycles. Unlike the stable ON state of around 5 kΩ, the OFF state deviated a lot in the first 10 cycles and then slightly fluctuated at around 30 kΩ, as presented in Figure A.13d. Considering the relatively stable OFF state, it can be concluded with an ON/OFF ratio of around 6.

Following the discussion on C2C variability of a single device, D2D variability was studied that five selector devices on the same sample corresponding to 20 % of RF power were characterized. Figure A.14a visualizes the statistics of V_{th} and V_h in a box plot. For all devices, the distributions of V_{th} and V_h overlap to a massive extent. The device 3 even has extremely similar distributions for V_{th} and V_h, where the median difference can be barely distinguished. Except for the device 3, the median V_{th} is larger than the median V_h, implying a V_{th} higher than a V_h could be generally expected. Furthermore, the medians of V_{th} and V_h vary among the five devices with the maximum differences of 0.6 V and 0.4 V respectively. A higher median V_{th} commonly accompanies with a higher median V_h, indicating the operating voltages affect each other with a positive correlation.

Besides the operating voltages, D2D variabilities of R_{OFF} and R_{ON} were investigated by visualizing the statistics in cumulative distribution functions plotted in Figure A.14b. Favorably, R_{ON} of all the devices exhibit narrow distributions, which stand closely to each other with mean values centered around 5.2 kΩ. In contrast, broader distributions were commonly found in R_{OFF}. Since the ON state is rather stable, a large variation in the OFF state can only be attributed to switching from the ON state to the OFF state, which appears to be

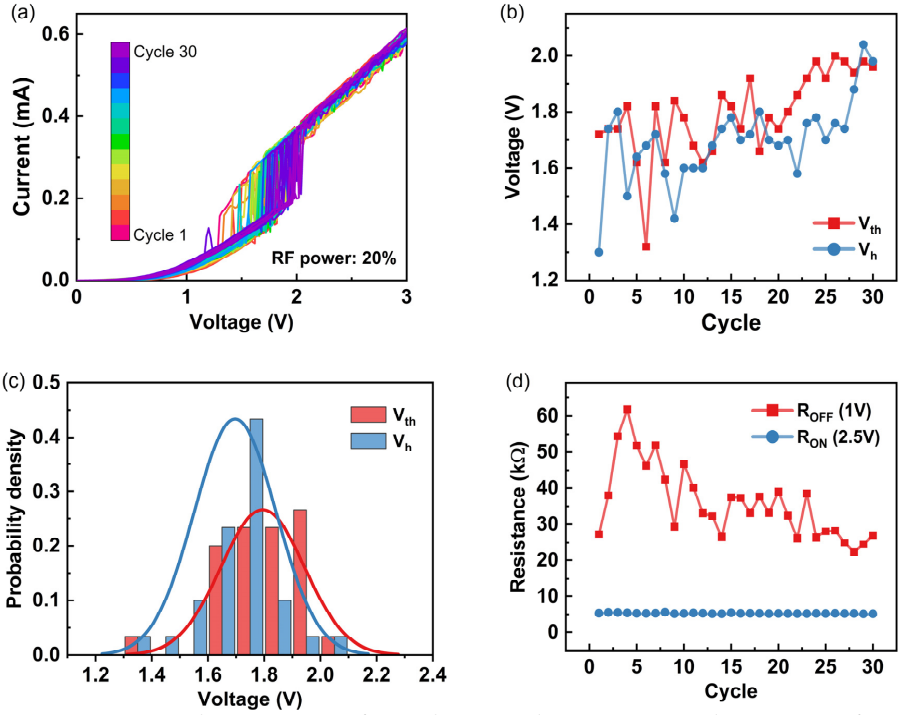

Figure A.13: *OTS characteristics of a single SiTe$_x$ selector corresponding to 20 % of RF power over 30 unipolar switching cycles. (a) C2C evolution of I-V characteristic. (b) Diagram of V_{th} and V_h versus number of cycles. (c) Distributions of V_{th} and V_h as probability density functions. (d) Diagram of R_{OFF} and R_{ON} versus number of cycles.*

more challenging in the control of OTS behavior. For the devices 1, 3, and 5, the R_{OFF} distributions are relatively narrow with the mean values centered around 22 kΩ. Considering the means of R_{OFF} and R_{ON}, an ON/OFF ratio of 4.2 is concluded. With higher R_{OFF}, the devices 2 and 4 have higher ON/OFF ratios of 9.1 and 6.9 respectively.

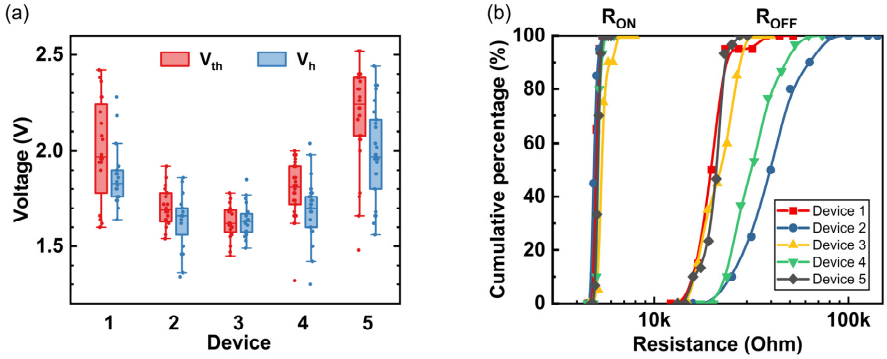

Figure A.14: *D2D variabilities among five SiTe$_x$ selectors corresponding to 20 % of RF power. The statistics are based on 20 unipolar switching cycles of each device. (a) Box plot of V_{th} and V_h. (b) D2D variabilities of R_{OFF} and R_{ON}.*

By comparing Figure A.14a and Figure A.14b, the C2C variability of R_{OFF} does not clearly correlate to the C2C variabilities V_{th} and V_h in a single device. As the triggering point to switch on the device from the OFF state, the V_{th} is not affected by the R_{OFF}. For instance, the device 1 has a low R_{OFF} variability but a high V_{th} variability. Additionally, there is no clear effect of the V_h on the R_{OFF}, even though the device resistance ends up at the R_{OFF} after the device is switched off at the V_h. For example, the device 5 has a high V_h variability but a low R_{OFF} variability.

In conclusion, the OTS behavior was observed only from the devices corresponding to 20 % of RF power in $SiTe_x$ sputter deposition, which results in Te concentration of 90 at.%. Despite C2C and D2D variabilities, both V_{th} and V_h were found in a range of 1.5 V to 2.5 V, which is favorably comparable to RRAM's typical operating voltages. The major challenge towards applications in 1S-1R arrays remains to be the non-linearity. The highest ON/OFF ratio was found merely 9.1, while the practical requirement is in a range of 10^4 to 10^6 depending on the operating voltages and the desired margins for READ and WRITE. To increase the ON/OFF ratio by 3 to 5 orders of magnitude, R_{OFF} enhancement is more of the concern in comparison to R_{ON} reduction, because the R_{ON} reported at around 5.2 kΩ is already comparable to RRAM's typical R_{LRS}, which is in a range of 1 kΩ to 10 kΩ. When the selector is switched on, not only a considerable portion of voltage drop can be passed to the RRAM, but also a high current can be delivered in series for enabling a RESET process in 1S-1R configuration. Consequently, the aim of further developments on the $SiTe_x$ selectors can be mainly focused on a substantial R_{OFF} increase above 100 MΩ possibly by increasing Te concentration or film thickness.

Bibliography

[1] A. Boroumand *et al.*, "Google workloads for consumer devices: Mitigating data movement bottlenecks," in *Proceedings of the Twenty-Third International Conference on Architectural Support for Programming Languages and Operating Systems*, 2018, pp. 316-331.

[2] KLA Corporation. "KLA Celebrates the 75th Anniversary of the Transistor." https://www.kla.com/advance/innovation/kla-celebrates-the-75th-anniversary-of-the-transistor (accessed March 3, 2024).

[3] G. W. Burr *et al.*, "Neuromorphic computing using non-volatile memory," *Advances in Physics: X,* vol. 2, no. 1, pp. 89-124, 2017.

[4] D. V. Christensen *et al.*, "2022 roadmap on neuromorphic computing and engineering," *Neuromorphic Computing and Engineering,* vol. 2, no. 2, p. 022501, 2022.

[5] S. Kumar, X. Wang, J. P. Strachan, Y. Yang, and W. D. Lu, "Dynamical memristors for higher-complexity neuromorphic computing," *Nature Reviews Materials,* vol. 7, no. 7, pp. 575-591, 2022.

[6] W. Wang *et al.*, "Integration and co-design of memristive devices and algorithms for artificial intelligence," *iScience,* vol. 23, no. 12, p. 101809, 2020.

[7] D. Ielmini, "Brain-inspired computing with resistive switching memory (RRAM): Devices, synapses and neural networks," *Microelectronic Engineering,* vol. 190, 2018.

[8] A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, and E. Eleftheriou, "Memory devices and applications for in-memory computing," *Nat Nanotechnol,* vol. 15, no. 9, pp. 812-812, 2020.

[9] L. Chua, "Memristor-the missing circuit element," *IEEE Transactions on circuit theory,* vol. 18, no. 5, pp. 507-519, 1971.

[10] L. Chua, "Resistance switching memories are memristors," *Handbook of memristor networks,* pp. 197-230, 2019.

[11] A. Amirsoleimani *et al.*, "In-memory vector-matrix multiplication in monolithic complementary metal–oxide–semiconductor-memristor integrated circuits: design choices, challenges, and perspectives," *Advanced Intelligent Systems,* vol. 2, no. 11, p. 2000115, 2020.

[12] C. Bengel *et al.*, "Utilizing the switching stochasticity of HfO_2/TiO_x-based ReRAM devices and the concept of multiple device synapses for the classification of overlapping and noisy patterns," *Frontiers in neuroscience,* vol. 15, p. 661856, 2021.

[13] W.-H. Chen *et al.*, "CMOS-integrated memristive non-volatile computing-in-memory for AI edge processors," *Nature Electronics,* vol. 2, no. 9, pp. 420-428, 2019.

[14] M. Le Gallo *et al.*, "A 64-core mixed-signal in-memory compute chip based on phase-change memory for deep neural network inference," *Nature Electronics,* vol. 6, no. 9, pp. 680-693, 2023.

[15] C. Li *et al.*, "Efficient and self-adaptive in-situ learning in multilayer memristor neural networks," *Nature communications,* vol. 9, no. 1, p. 2385, 2018.

[16] C. Li *et al.*, "Long short-term memory networks in memristor crossbar arrays," *Nature Machine Intelligence,* vol. 1, no. 1, pp. 49-57, 2019.
[17] V. Milo *et al.*, "Multilevel HfO_2-based RRAM devices for low-power neuromorphic networks," *APL materials,* vol. 7, no. 8, 2019.
[18] W. Wan *et al.*, "A compute-in-memory chip based on resistive random-access memory," *Nature,* vol. 608, no. 7923, pp. 504-512, 2022.
[19] W. Wang *et al.*, "Computing of temporal information in spiking neural networks with ReRAM synapses," *Faraday discussions,* vol. 213, pp. 453-469, 2019.
[20] S. Yu, Y. Wu, and H.-S. P. Wong, "Investigating the switching dynamics and multilevel capability of bipolar metal oxide resistive switching memory," *Applied Physics Letters,* vol. 98, no. 10, 2011.
[21] Q. Xia and J. J. Yang, "Memristive crossbar arrays for brain-inspired computing," *Nature materials,* vol. 18, no. 4, pp. 309-323, 2019.
[22] Y. Cai *et al.*, "Technology-array-algorithm co-optimization of RRAM for storage and neuromorphic computing: Device non-idealities and thermal cross-talk," in *2020 IEEE International Electron Devices Meeting (IEDM)*, 2020: IEEE, pp. 13.4. 1-13.4. 4.
[23] D. Schön and S. Menzel, "Spatio-Temporal Correlations in Memristive Crossbar Arrays due to Thermal Effects," *Advanced functional materials,* p. 2213943, 2023.
[24] F. Staudigl *et al.*, "NeuroHammer: Inducing bit-flips in memristive crossbar memories," in *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2022: IEEE, pp. 1181-1184.
[25] R. Sun, H. Chen, G. Wang, C. Wang, and L. Hao, "Reduction of thermal disturbances in 3D 1S1R RRAM crossbar arrays for neuromorphic computing," *Semiconductor Science and Technology,* vol. 34, no. 11, p. 115023, 2019.
[26] F. Staudigl *et al.*, "It's getting hot in here: Hardware security implications of thermal crosstalk on ReRAMs," *IEEE Transactions on Reliability,* 2024.
[27] R. Dittmann, S. Menzel, and R. Waser, "Nanoionic memristive phenomena in metal oxides: the valence change mechanism," *Advances in physics,* vol. 70, no. 2, pp. 155-349, 2021.
[28] B. Choi *et al.*, "Resistive switching mechanism of TiO_2 thin films grown by atomic-layer deposition," *Journal of applied physics,* vol. 98, no. 3, 2005.
[29] H. Zhang *et al.*, "Understanding the coexistence of two bipolar resistive switching modes with opposite polarity in Pt/TiO_2/Ti/Pt nanosized ReRAM devices," *ACS applied materials & interfaces,* vol. 10, no. 35, pp. 29766-29778, 2018.
[30] Z. Yong *et al.*, "Tuning oxygen vacancies and resistive switching properties in ultra-thin HfO_2 RRAM via TiN bottom electrode and interface engineering," *Applied Surface Science,* vol. 551, p. 149386, 2021.
[31] H. Y. Lee *et al.*, "Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO_2 based RRAM," in *2008 IEEE International Electron Devices Meeting*, 2008, pp. 1-4.
[32] W. Kim, "Investigation of switching mechanism in Ta_2O_5-based ReRAM devices," PhD, RWTH Aachen University, 2017.
[33] A. Kindsmüller, "The role of interface and exchange processes in forming and switching of ZrO_2 based ReRAM devices," PhD, RWTH Aachen University, 2020.
[34] C.-Y. Lin, C.-Y. Wu, C.-Y. Wu, C. Hu, and T.-Y. Tseng, "Bistable resistive switching in Al_2O_3 memory thin films," *Journal of the Electrochemical Society,* vol. 154, no. 9, p. G189, 2007.
[35] N. Aslam, "Resistive switching memory devices fromatomic layer deposited binary and ternaryoxide thin films," PhD, RWTH Aachen University, 2017.

[36] K. Szot, R. Dittmann, W. Speier, and R. Waser, "Nanoscale resistive switching in SrTiO3 thin films," *physica status solidi (RRL)–Rapid Research Letters,* vol. 1, no. 2, pp. R86-R88, 2007.

[37] A. Hardtdegen, C. La Torre, F. Cüppers, S. Menzel, R. Waser, and S. Hoffmann-Eifert, "Improved switching stability and the effect of an internal series resistor in HfO_2/TiO_x bilayer ReRAM cells," *IEEE Trans. Electron Devices,* vol. 65, no. 8, pp. 3229-3236, 2018.

[38] I. Kärkkänen *et al.*, "Study of atomic layer deposited ZrO_2 and ZrO_2/TiO_2 films for resistive switching application," *physica status solidi (a),* vol. 211, no. 2, pp. 301-309, 2014.

[39] M.-J. Lee *et al.*, "A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structures," *Nature materials,* vol. 10, no. 8, pp. 625-630, 2011.

[40] G.-S. Park *et al.*, "In situ observation of filamentary conducting channels in an asymmetric $Ta2O_{5-x}/TaO_{2-x}$ bilayer structure," *Nature communications,* vol. 4, no. 1, p. 2382, 2013.

[41] Y.-S. Chen *et al.*, "An ultrathin forming-free HfO_x resistance memory with excellent electrical performance," *IEEE Electron Device Letters,* vol. 31, no. 12, pp. 1473-1475, 2010.

[42] Y. Jiang *et al.*, "Forming-Free TaO_x based RRAM device with low operating voltage and high on/off characteristics," *ECS Journal of Solid State Science and Technology,* vol. 4, no. 12, p. N137, 2015.

[43] W. Kim *et al.*, "Forming-free metal-oxide ReRAM by oxygen ion implantation process," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016: IEEE, pp. 4.4. 1-4.4. 4.

[44] W. Kim *et al.*, "Forming-free nitrogen-doped AlO_x RRAM with sub-μA programming current," in *2011 Symposium on VLSI Technology-Digest of Technical Papers*, 2011: IEEE, pp. 22-23.

[45] Z. Wang *et al.*, "Highly uniform, self-compliance, and forming-free ALD HfO_2-based RRAM with Ge doping," *IEEE Trans. Electron Devices,* vol. 59, no. 4, pp. 1203-1208, 2012.

[46] K.-C. Chang *et al.*, "Characteristics and mechanisms of silicon-oxide-based resistance random access memory," *IEEE Electron Device Letters,* vol. 34, no. 3, pp. 399-401, 2013.

[47] R. Waser, R. Bruchhaus, and S. Menzel, "Redox-based resistive switching memories," in *Nanoelectronics and Information Technology*, R. Waser Ed., 3rd ed. Berlin: Wiley-VCH, 2012, p. 683.

[48] A. Chen, "Analysis of partial bias schemes for the writing of crossbar memory arrays," *IEEE Trans. Electron Devices,* vol. 62, no. 9, pp. 2845-2849, 2015.

[49] W. Sun, S. Choi, and H. Shin, "A new bias scheme for a low power consumption ReRAM crossbar array," *Semiconductor Science and Technology,* vol. 31, no. 8, p. 085009, 2016.

[50] G. W. Burr *et al.*, "Access devices for 3D crosspoint memory," *Journal of Vacuum Science & Technology B,* vol. 32, no. 4, 2014.

[51] A. Chen, "Nonlinearity and asymmetry for device selection in cross-bar memory arrays," *IEEE Trans. Electron Devices,* vol. 62, no. 9, pp. 2857-2864, 2015.

[52] A. N. Youssef, A. L. Jagath, N. K. Thulasiraman, and H. A. F. Almurib, "Effect of sneak path current in TiO_x/HfO_x based 1S1R RRAM crossbar memory array," in *2021 IEEE 19th Student Conference on Research and Development (SCOReD)*, 2021: IEEE, pp. 267-272.

[53] M. Alayan *et al.*, "In-depth investigation of programming and reading operations in RRAM cells integrated with Ovonic Threshold Switching (OTS) selectors," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017: IEEE, pp. 2.3. 1-2.3. 4.

[54] Y. Kim *et al.*, "Novel selector-induced current-limiting effect through asymmetry control for high-density one-selector–one-resistor crossbar arrays," *Advanced Electronic Materials,* vol. 5, no. 7, p. 1800806, 2019.

[55] D. A. Robayo *et al.*, "Integration of OTS based back-end selector with HfO 2 OxRAM for crossbar arrays," in *2019 IEEE 11th International Memory Workshop (IMW)*, 2019: IEEE, pp. 1-4.

[56] N. K. Upadhyay *et al.*, "A memristor with low switching current and voltage for 1S1R integration and array operation," *Advanced Electronic Materials,* vol. 6, no. 5, p. 1901411, 2020.

[57] H. C. Woo, J. Kim, S. Lee, H. J. Kim, and C. S. Hwang, "Stacked one-selector-one-resistive memory crossbar array with high nonlinearity and on-current density for the neuromorphic applications," *Advanced Electronic Materials,* vol. 8, no. 12, p. 2200656, 2022.

[58] D. Depla, S. Mahieu, and J. E. Greene, "Chapter 5 - Sputter Deposition Processes," in *Handbook of Deposition Technologies for Films and Coatings (Third Edition)*, P. M. Martin Ed. Boston: William Andrew Publishing, 2010, pp. 253-296.

[59] Matsusada Precision Inc. "Sputtering." https://www.matsusada.com/application/ps/sputtering/ (accessed February 16, 2025).

[60] K. Wasa, M. Kitabatake, and H. Adachi, "2 - Thin Film Processes," in *Thin Film Materials Technology*, K. Wasa, M. Kitabatake, and H. Adachi Eds. Norwich, NY: William Andrew Publishing, 2004, pp. 17-69.

[61] M. Huff, "Recent Advances in Reactive Ion Etching and Applications of High-Aspect-Ratio Microfabrication," *Micromachines,* vol. 12, no. 8, p. 991, 2021. [Online]. Available: https://www.mdpi.com/2072-666X/12/8/991.

[62] V. Kohli. "Reactive Ion Etching: A Comprehensive Guide." Wevolver. https://www.wevolver.com/article/reactive-ion-etching-a-comprehensive-guide (accessed March 7, 2024).

[63] S. Hönl, H. Hahn, Y. Baumgartner, L. Czornomaz, and P. Seidler, "Highly selective dry etching of GaP in the presence of $Al_xGa_{1-x}P$ with a SiCl4/SF6 plasma," *Journal of Physics D: Applied Physics,* vol. 51, no. 18, p. 185203, 2018.

[64] T. Uesugi *et al.*, "Novel ArF photoresist polymer to suppress the formation of roughness in plasma etching processes," *Journal of Vacuum Science & Technology A,* vol. 31, no. 6, 2013.

[65] MKS Instruments. "Isotropic Radical Etching." https://www.mks.com/n/isotropic-radical-etching (accessed March 8, 2024).

[66] S. J. Hwang *et al.*, "Dry etching of polydimethylsiloxane using microwave plasma," *Journal of Micromechanics and Microengineering,* vol. 19, no. 9, p. 095010, 2009.

[67] Y.-M. Lee, "Introduction to optical lithography," *Efficient Extreme Ultraviolet Mirror Design*: IOP Publishing, 2021, pp. 1-1-1-16.

[68] E. Abbe, "The Relation of Aperture and Power in the Microscope (continued)," *Journal of the Royal Microscopical Society,* vol. 2, no. 4, pp. 460-473, 1882.

[69] L. Rayleigh, "XXXI. Investigations in optics, with special reference to the spectroscope," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* vol. 8, no. 49, pp. 261-274, 1879.

[70] A. Yen, "Rayleigh or Abbe? Origin and naming of the resolution formula of microlithography," *Journal of Micro/Nanolithography, MEMS, and MOEMS,* vol. 19, no. 4, pp. 040501-040501, 2020.

[71] K. Jain, C. G. Willson, and B. J. Lin, "Ultrafast deep UV Lithography with excimer lasers," *IEEE Electron Device Letters,* vol. 3, no. 3, pp. 53-55, 1982.
[72] G. Yeap *et al.*, "5nm cmos production technology platform featuring full-fledged euv, and high mobility channel finfets with densest 0.021 μm 2 sram cells for mobile soc and high performance computing applications," in *2019 IEEE International Electron Devices Meeting (IEDM)*, 2019: IEEE, pp. 36.7. 1-36.7. 4.
[73] Heidelberg Instruments Mikrotechnik GmbH, "The Raster Scan Mode," in *Fact sheet μMLA V5.0*.
[74] J. Meszaros, "Large area zone plate exposure by fixed beam moving stage lithography," Master of Science, KTH - Royal Institute of Technology, Stockholm, 2011.
[75] P. E. Batson, N. Dellby, and O. L. Krivanek, "Sub-ångstrom resolution using aberration corrected electron optics," *Nature,* vol. 418, no. 6898, pp. 617-620, 2002.
[76] M. A. O'Keefe, "Sub-Ångstrom Resolution with Aberration-Corrected TEM: Present and Future," *Microscopy and Microanalysis,* vol. 10, no. S03, pp. 68-69, 2004.
[77] M. A. O'Keefe and L. F. Allard, "A standard for sub-Ångstrom metrology of resolution in aberration-corrected transmission electron microscopes," *Microscopy and Microanalysis,* vol. 10, no. S02, pp. 1002-1003, 2004.
[78] M. A. O'Keefe and L. F. Allard, "Sub-Angstrom electron microscopy for sub-Angstrom nano-metrology," Lawrence Berkeley National Lab (LBNL), 2004.
[79] M. A. O'Keefe, L. F. Allard, and D. A. Blom, "Resolution quality and atom positions in sub-Ångström electron microscopy," *Microscopy and Microanalysis,* vol. 11, no. S02, pp. 540-541, 2005.
[80] E. Ruska, "The development of the electron microscope and of electron microscopy," *Reviews of modern physics,* vol. 59, no. 3, p. 627, 1987.
[81] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy*, 2nd ed. (A Textbook for Materials Science). New York, NY: Springer, 2009.
[82] Y. Liao, *Practical electron microscopy and database*, 2nd ed., 2006. [Online]. Available: https://www.globalsino.com/EM/.
[83] Nanoscience Instruments. "What is Transmission Electron Microscopy?" https://www.nanoscience.com/techniques/transmission-electron-microscopy/ (accessed March 8, 2024).
[84] M. Knoll, "Aufladepotentiel und sekundäremission elektronenbestrahlter körper," *Zeitschrift für Technische Physik,* vol. 16, p. 467, 1935.
[85] D. McMullan, "Scanning electron microscopy 1928–1965," *Scanning,* vol. 17, no. 3, pp. 175-185, 1995.
[86] M. Von Ardenne, "Das elektronen-rastermikroskop. Praktische ausführung," *Zeitschrift für technische Physik,* vol. 19, no. 11, pp. 407-416, 1938.
[87] M. Von Ardenne, "Das Elektronen-Rastermikroskop: Theoretische Grundlagen," *Zeitschrift für Physik,* vol. 109, no. 9-10, pp. 553-572, 1938.
[88] V. K. Zworykin, "The scanning electron microscope," *Scientific American,* vol. 167, no. 3, pp. 111-113, 1942.
[89] V. K. Zworykin and J. Hillier, "Electronic Microscopy," *The Scientific Monthly,* vol. 59, no. 3, pp. 165-179, 1944.
[90] C. Oatley, D. McMullan, and K. Smith, "The development of the scanning electron microscope," in *The Beginngs of Electron Microscopy*. Amsterdam, Netherlands: Elsevier, 1985, pp. 443-482.
[91] C. W. Oatley, "The early history of the scanning electron microscope," *Journal of Applied Physics,* vol. 53, no. 2, pp. R1-R13, 1982.

[92] Nanoscience Instruments. "What is Scanning Electron Microscopy?" https://www.nanoscience.com/techniques/scanning-electron-microscopy/ (accessed March 11, 2024).
[93] W. Zhou, R. Apkarian, Z. L. Wang, and D. Joy, "Fundamentals of Scanning Electron Microscopy (SEM)," in *Scanning Microscopy for Nanotechnology: Techniques and Applications*, W. Zhou and Z. L. Wang Eds. New York, NY: Springer, 2007, pp. 1-40.
[94] Ponor, "Electron-matter interaction volume and various types of signal generated," Wikimedia Commons, 2020.
[95] Central Microscopy Research Facility at The University of Iowa. "Scanning Electron Microscopy." https://cmrf.research.uiowa.edu/scanning-electron-microscopy (accessed March 11, 2024).
[96] A. Nanakoudis. "EDX Analysis with SEM: How Does it Work?" Thermo Fisher Scientific Inc. https://www.thermofisher.com/blog/materials/edx-analysis-with-sem-how-does-it-work/ (accessed February 16, 2025).
[97] C. Suryanarayana, "Microstructure: An Introduction," in *Aerospace Materials and Material Technologies*, N. E. Prasad and R. J. H. Wanhill Eds. Singapore: Springer, 2017, pp. 105-123.
[98] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, "Tunneling through a controllable vacuum gap," *Applied Physics Letters,* vol. 40, no. 2, pp. 178-180, 1982.
[99] G. Binnig, C. F. Quate, and C. Gerber, "Atomic Force Microscope," *Phys Rev Lett,* vol. 56, no. 9, pp. 930-933, 1986.
[100] O. Marti, "AFM instrumentation and tips," in *Handbook of Micro/Nano Tribology*, B. Bhushan Ed., 2nd ed.: CRC Press, 1999, pp. 81-144.
[101] NanoAndMore USA Corp. "What is Atomic Force Microscopy (AFM)." https://www.nanoandmore.com/what-is-atomic-force-microscopy (accessed March 11, 2024).
[102] Nanoscience Instruments. "What is Atomic Force Microscopy?" https://www.nanoscience.com/techniques/atomic-force-microscopy/ (accessed March 11, 2024).
[103] A. Vilalta-Clemente, K. Gloystein, and N. Frangis, "Principles of atomic force microscopy (AFM)," *Proceedings of physics of advanced materials Winter School,* pp. 1-8, 2008.
[104] Tektronix, "4200A-SCS Parameter Analyzer Datasheet." [Online]. Available: https://download.tek.com/datasheet/1KW-60780-61KW-60780-6_4200A-SCS_Parameter_Analyzer_Datasheet_061523.pdf.
[105] D. Chi, D. Mangelinck, A. Zuruzi, A. Wong, and S. Lahiri, "Nickel silicide as a contact material for submicron CMOS devices," *Journal of electronic materials,* vol. 30, pp. 1483-1488, 2001.
[106] J. Gambino and E. Colgan, "Silicides and ohmic contacts," *Materials chemistry and physics,* vol. 52, no. 2, pp. 99-146, 1998.
[107] T. Morimoto *et al.*, "Self-aligned nickel-mono-silicide technology for high-speed deep submicrometer logic CMOS ULSI," *IEEE Trans. Electron Devices,* vol. 42, no. 5, pp. 915-922, 1995.
[108] S.-L. Zhang and U. Smith, "Self-aligned silicides for Ohmic contacts in complementary metal-oxide-semiconductor technology: $TiSi_2$, $CoSi_2$, and NiSi," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 22, no. 4, pp. 1361-1370, 2004.
[109] Z. Chen, Z. Fang, X. Wang, G.-Q. Lo, D.-L. Kwong, and Y. Wu, "Switching Behaviors of $HfO_2/NiSi_x$ Based RRAM," *International Journal of Electronics and Communication Engineering,* vol. 7, no. 9, pp. 1230-1232, 2013.

[110] Z. Chen *et al.*, "Impact of Ni concentration on the performance of Ni silicide/HfO_2/TiN resistive RAM (RRAM) cells," *Journal of electronic materials,* vol. 43, pp. 4193-4198, 2014.
[111] T. Kawanago, "A study on high-k/metal gate stack MOSFETs with rare earth oxides," PhD, Tokyo Institute of Technology, 2011.
[112] T. Kawanago *et al.*, "Optimized oxygen annealing process for V_{th} tuning of p-MOSFET with high-k/metal gate stacks," in *2010 Proceedings of the European Solid State Device Research Conference*, 2010: IEEE, pp. 301-304.
[113] J. Robertson and R. M. Wallace, "High-K materials and metal gates for CMOS applications," *Materials Science and Engineering: R: Reports,* vol. 88, pp. 1-41, 2015.
[114] M. Takayanagi, "Challenge for High-k/Metal Gate CMOSFETs in 32 nm Generation and Beyond," *ECS Transactions,* vol. 6, no. 1, p. 157, 2007.
[115] B. Ye *et al.*, "NBTI mitigation by optimized HKMG thermal processing in a FinFET technology," *IEEE Trans. Electron Devices,* vol. 69, no. 3, pp. 905-909, 2022.
[116] C. A. Mack, "Corner rounding and round contacts," *The Lithography Expert, Summer,* 2000.
[117] MicroChemicals GmbH, "Exposure." [Online]. Available: https://www.microchemicals.com/technical_information/exposure_photoresist.pdf.
[118] H. J. Levinson, "Photoresists," in *Principles of Lithography*, vol. PM304, 4th ed.: SPIE, 2019.
[119] C. A. Mack and J. E. Connors, "Fundamental differences between positive-and negative-tone imaging," in *Optical/Laser Microlithography V*, 1992, vol. 1674: SPIE, pp. 328-338.
[120] MicroChemicals GmbH, "Composition and properties of AZ® and TI photoresists." [Online]. Available: https://www.microchemicals.com/technical_information/photoresists_composition_properties.pdf.
[121] A. D. Bailey III and R. A. G. R. A. Gottscho, "Aspect ratio independent etching: Fact or fantasy?," *Japanese journal of applied physics,* vol. 34, no. 4S, p. 2083, 1995.
[122] R. A. Gottscho, C. W. Jurgensen, and D. Vitkavage, "Microscopic uniformity in plasma etching," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena,* vol. 10, no. 5, pp. 2133-2147, 1992.
[123] C. Hedlund, H. O. Blom, and S. Berg, "Microloading effect in reactive ion etching," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 12, no. 4, pp. 1962-1965, 1994.
[124] H. Jones, R. Bennett, and J. Singh, "Size dependent etching of small shapes," in *Proc. 8th Symp. Plasma Processing*, 1990, p. 45.
[125] A. Misaka and K. Harafuji, "Simulation study of micro-loading phenomena in silicon dioxide hole etching," *IEEE Trans. Electron Devices,* vol. 44, no. 5, pp. 751-760, 1997.
[126] Samco Inc. "Loading effect and microloading effect in Si deep RIE." https://www.samcointl.com/opto/loading-effect-and-microloading-effect-in-si-deep-rie/ (accessed March 12, 2024).
[127] M. Sato, D. Takehara, K. Uda, and K. S. Hara, "Suppression of microloading effect by low-temperature SiO_2 etching," *Japanese journal of applied physics,* vol. 31, no. 12S, p. 4370, 1992.
[128] S. i. Kato, M. Sato, and Y. Arita, "Microloading effect prevention in SiO2 contact-hole etching," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 12, no. 4, pp. 1204-1208, 1994.

[129] L. Feng, Z. Amini, G. Tipton, S. Everist, and R. Jarecki, "Comparison of etch characteristics with oxide masks and resist masks in polysilicon etching," *Electrochemical Society Proceedings,* vol. 96-12, pp. 398-409, 1996.
[130] P. R. Shrestha *et al.*, "Analysis and control of RRAM overshoot current," *IEEE Trans. Electron Devices,* vol. 65, no. 1, pp. 108-114, 2017.
[131] H. Wan *et al.*, "In situ observation of compliance-current overshoot and its effect on resistive switching," *IEEE Electron Device Letters,* vol. 31, no. 3, pp. 246-248, 2010.
[132] D. Gilmer *et al.*, "Effects of RRAM stack configuration on forming voltage and current overshoot," in *2011 3rd IEEE International Memory Workshop (IMW)*, 2011.
[133] A. Kalantarian *et al.*, "Controlling uniformity of RRAM characteristics through the forming process," in *2012 IEEE International Reliability Physics Symposium (IRPS)*, 2012: IEEE, pp. 6C. 4.1-6C. 4.5.
[134] S. Yu, X. Guan, and H.-S. P. Wong, "Understanding metal oxide RRAM current overshoot and reliability using Kinetic Monte Carlo simulation," in *2012 International Electron Devices Meeting*, 2012: IEEE, pp. 26.1. 1-26.1. 4.
[135] H.-S. P. Wong *et al.*, "Metal–oxide RRAM," *Proceedings of the IEEE,* vol. 100, no. 6, pp. 1951-1970, 2012.
[136] K. Kinoshita *et al.*, "Reduction of reset current in NiO-ReRAM brought about by ideal current limiter," in *2007 22nd IEEE Non-Volatile Semiconductor Memory Workshop*, 2007: IEEE, pp. 66-67.
[137] M. Galicia *et al.*, "NEUROTEC I: Neuro-inspired Artificial Intelligence Technologies for the Electronics of the Future," in *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2022: IEEE, pp. 957-962.
[138] J. Borghetti, G. S. Snider, P. J. Kuekes, J. J. Yang, D. R. Stewart, and R. S. Williams, "'Memristive' switches enable 'stateful' logic operations via material implication," *Nature,* vol. 464, no. 7290, pp. 873-876, 2010.
[139] K. M. Kim and R. S. Williams, "A family of stateful memristor gates for complete cascading logic," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 66, no. 11, pp. 4348-4355, 2019.
[140] Y. S. Kim, M. W. Son, and K. M. Kim, "Memristive stateful logic for edge Boolean computers," *Advanced Intelligent Systems,* vol. 3, no. 7, p. 2000278, 2021.
[141] Y. S. Kim *et al.*, "Stateful in-memory logic system and its practical implementation in a TaO_x-based bipolar-type memristive crossbar array," *Advanced Intelligent Systems,* vol. 2, no. 3, p. 1900156, 2020.
[142] C. Bengel *et al.*, "Experimental and theoretical analysis of stateful logic in passive and active crossbar arrays for computation-in-memory," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2022, pp. 2792-2796.
[143] S. Kvatinsky, G. Satat, N. Wald, E. G. Friedman, A. Kolodny, and U. C. Weiser, "Memristor-based material implication (IMPLY) logic: Design principles and methodologies," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* vol. 22, no. 10, pp. 2054-2066, 2013.
[144] Z. Sun, E. Ambrosi, A. Bricalli, and D. Ielmini, "Logic computing with stateful neural networks of resistive switches," *Advanced Materials,* vol. 30, no. 38, p. 1802554, 2018.
[145] C. Bengel *et al.*, "Variability-aware modeling of filamentary oxide-based bipolar resistive switching cells using SPICE level compact models," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 67, no. 12, pp. 4618-4630, 2020.
[146] M. Hu *et al.*, "Dot-product engine for neuromorphic computing: Programming 1T1M crossbar to accelerate matrix-vector multiplication," in *Proceedings of the 53rd annual design automation conference*, 2016, pp. 1-6.

[147] C. Li *et al.*, "CMOS-integrated nanoscale memristive crossbars for CNN and optimization acceleration," in *2020 IEEE International Memory Workshop (IMW)*, 2020: IEEE, pp. 1-4.
[148] F. Lentz, "Integration of redox based resistive switching memory devices," PhD, RWTH Aachen University, 2015.
[149] M. Meier, "Entwicklung einer Nanotechnologie-Plattform für die Herstellung Crossbar-basierter Speicherarchitekturen," PhD, RWTH Aachen University, 2009.
[150] K. Milkove, J. Coffin, and C. Dziobkowski, "Effects of argon addition to a platinum dry etch process," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 16, no. 3, pp. 1483-1488, 1998.
[151] K. Milkove and C. Wang, "Insight into the dry etching of fence-free patterned platinum structures," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 15, no. 3, pp. 596-603, 1997.
[152] T. Shibano, K. Nakamura, T. Takenaga, and K. Ono, "Platinum etching in Ar/Cl_2 plasmas with a photoresist mask," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 17, no. 3, pp. 799-804, 1999.
[153] T. Shibano and T. Oomori, "Sidewall deposition film in platinum etching with Ar/halogen mixed gas plasmas," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena,* vol. 15, no. 5, pp. 1747-1751, 1997.
[154] W. J. Yoo, J. H. Hahm, H. W. Kim, C. O. Jung, Y. B. Koh, and M. Y. Lee, "Control of etch slope during etching of Pt in $Ar/Cl_2/O_2$ plasmas," *Japanese journal of applied physics,* vol. 35, no. 4S, p. 2501, 1996.
[155] C. E. Farrell, K. Milkove, C. Wang, and D. E. Kotecki, "A reactive ion etch study for producing patterned platinum structures," *Integrated Ferroelectrics,* vol. 16, no. 1-4, pp. 109-138, 1997.
[156] W. Kim, D. J. Wouters, R. Waser, and V. Rana, "Tuning the memory window of TaO_x reram using the rf sputtering power," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021: IEEE, pp. 1-5.
[157] C. W. Chung and H. G. Song, "Study on fence-free platinum etching using chlorine-based gases in inductively coupled plasma," *Journal of the Electrochemical Society,* vol. 144, no. 11, p. L294, 1997.
[158] H. Ranpura, D. Butler, L. H. Chang, C. Tracy, and S. Beaudoin, "Aqueous cleaning of sidewall redepositions formed by reactive ion etching of platinum," *Journal of the Electrochemical Society,* vol. 146, no. 8, p. 3114, 1999.
[159] M.-C. Chiang, F.-M. Pan, H.-C. Cheng, J.-S. Liu, S.-H. Chan, and T.-C. Wei, "Dry etching of platinum films with TiN masks in an Ar/O_2 helicon wave plasma," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 18, no. 1, pp. 181-187, 2000.
[160] C. W. Chung and I. Chung, "Platinum etching using a TiO_2 hard mask in an $O_2/Cl_2/Ar$ plasma," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films,* vol. 18, no. 3, pp. 835-839, 2000.
[161] M. U. Gutsche, S. D. Athavale, K. Williams, and D. Hines, "Patterning of 0.175 μm platinum features using Ar/O_2 chemically assisted ion-beam etching," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena,* vol. 18, no. 2, pp. 765-773, 2000.
[162] S. P. Gurrum, S. K. Suman, Y. K. Joshi, and A. G. Fedorov, "Thermal issues in next-generation integrated circuits," *IEEE Transactions on device and materials reliability,* vol. 4, no. 4, pp. 709-714, 2004.

[163] E. Pop, S. Sinha, and K. E. Goodson, "Heat generation and transport in nanometer-scale transistors," *Proceedings of the IEEE,* vol. 94, no. 8, pp. 1587-1601, 2006.
[164] S. S. Salvi and A. Jain, "A review of recent research on heat transfer in three-dimensional integrated circuits (3-D ICs)," *IEEE Transactions on Components, Packaging and Manufacturing Technology,* vol. 11, no. 5, pp. 802-821, 2021.
[165] T. Chang, "Proximity effect in electron-beam lithography," *Journal of vacuum science and technology,* vol. 12, no. 6, pp. 1271-1275, 1975.
[166] C. S. Ea, "Electron beam proximity effect correction," PhD, University of Southampton, 2000.
[167] Raith GmbH. "Proximity effect correction – Getting your design precisely implemented on every sample." https://raith.com/technology/nanofabrication-software/proximity-effect-correction/ (accessed March 18, 2024).
[168] M. McCord and M. Rooks, "Electron Beam Lithography," in *Handbook of microlithography, micromachining, and microfabrication. Volume 1: Microlithography*, P. Rai-Choudhury Ed.: SPIE, 1997.
[169] M. A. Mohammad, M. Muhammad, S. K. Dew, and M. Stepanova, "Fundamentals of Electron Beam Exposure and Development," in *Nanofabrication: Techniques and Principles*, M. Stepanova and S. Dew Eds. Vienna: Springer, 2012, pp. 11-41.
[170] L. Ren and B. Chen, "Proximity effect in electron beam lithography," in *Proceedings. 7th International Conference on Solid-State and Integrated Circuits Technology, 2004.*, 2004, vol. 1: IEEE, pp. 579-582.
[171] R. Wüest, "Proximity-effect induced density limitations for electron-beam patterned planar photonic nanomaterials," *Photonics and Nanostructures-Fundamentals and Applications,* vol. 7, no. 4, pp. 212-219, 2009.
[172] S. R. Ovshinsky, "Reversible Electrical Switching Phenomena in Disordered Structures," *Phys Rev Lett,* vol. 21, no. 20, p. 1450, 1968.
[173] H. S. Chen and T. T. Wang, "On the theory of switching phenomena in semiconducting glasses," *physica status solidi (a),* vol. 2, no. 1, pp. 79-84, 1970.
[174] A. Owen, J. Robertson, and C. Main, "The threshold characteristics of chalcogenide-glass memory switches," *Journal of Non-Crystalline Solids,* vol. 32, no. 1-3, pp. 29-52, 1979.
[175] C. Popescu, "The effect of local non-uniformities on thermal switching and high field behaviour of structures with chalcogenide glasses," *Solid-state electronics,* vol. 18, no. 7-8, pp. 671-681, 1975.
[176] D. Adler, H. K. Henisch, and N. Mott, "The mechanism of threshold switching in amorphous alloys," *Reviews of Modern Physics,* vol. 50, no. 2, p. 209, 1978.
[177] D. Adler, M. Shur, M. Silver, and S. Ovshinsky, "Threshold switching in chalcogenide-glass thin films," *Journal of Applied Physics,* vol. 51, no. 6, pp. 3289-3309, 1980.
[178] A. Pirovano, A. L. Lacaita, F. Pellizzer, S. A. Kostylev, A. Benvenuti, and R. Bez, "Low-field amorphous state resistance and threshold voltage drift in chalcogenide materials," *IEEE Trans. Electron Devices,* vol. 51, no. 5, pp. 714-719, 2004.
[179] A. Redaelli, A. Pirovano, A. Benvenuti, and A. L. Lacaita, "Threshold switching and phase transition numerical models for phase change memory simulations," *Journal of Applied Physics,* vol. 103, no. 11, 2008.
[180] D. Ielmini, "Threshold switching mechanism by high-field energy gain in the hopping transport of chalcogenide glasses," *Physical Review B,* vol. 78, no. 3, p. 035308, 2008.
[181] D. Ielmini and Y. Zhang, "Analytical model for subthreshold conduction and threshold switching in chalcogenide-based memory devices," *Journal of Applied Physics,* vol. 102, no. 5, 2007.

[182] I. Karpov, M. Mitra, D. Kau, G. Spadini, Y. Kryukov, and V. Karpov, "Evidence of field induced nucleation in phase change memory," *Applied Physics Letters,* vol. 92, no. 17, 2008.
[183] V. Karpov, Y. Kryukov, S. Savransky, and I. Karpov, "Nucleation switching in phase change memory," *Applied physics letters,* vol. 90, no. 12, 2007.
[184] M. Nardone, V. Karpov, D. Jackson, and I. Karpov, "A unified model of nucleation switching," *Applied Physics Letters,* vol. 94, no. 10, 2009.
[185] M. Nardone, M. Simon, I. Karpov, and V. Karpov, "Electrical conduction in chalcogenide glasses of phase change memory," *Journal of Applied Physics,* vol. 112, no. 7, 2012.
[186] S. Kim, "Electrical characterization of Si_xTe_{1-x} chalcogenide glass for selector device application," Master, Seoul National University, 2017.
[187] Z. Zhao *et al.*, "Chalcogenide Ovonic Threshold Switching Selector," *Nano-Micro Letters,* vol. 16, no. 1, pp. 1-40, 2024.
[188] H. Li and J. Robertson, "Materials selection and mechanism of non-linear conduction in chalcogenide selector devices," *Scientific Reports,* vol. 9, no. 1, p. 1867, 2019.
[189] M. Anbarasu, M. Wimmer, G. Bruns, M. Salinga, and M. Wuttig, "Nanosecond threshold switching of GeTe6 cells and their potential as selector devices," *Applied Physics Letters,* vol. 100, no. 14, 2012.
[190] S. A. Chekol, J. Yoo, J. Park, J. Song, C. Sung, and H. Hwang, "A C–Te-based binary OTS device exhibiting excellent performance and high thermal stability for selector application," *Nanotechnology,* vol. 29, no. 34, p. 345202, 2018.
[191] T. Gao, J. Feng, and Z. Ma, "Threshold switching electrical responses in AlTe-based selector device," *Applied Physics A,* vol. 126, pp. 1-7, 2020.
[192] J. Ho Lee *et al.*, "Threshold switching in Si-As-Te thin film for the selector device of crossbar resistive memory," *Applied Physics Letters,* vol. 100, no. 12, 2012.
[193] S. Kim *et al.*, "Performance of threshold switching in chalcogenide glass for 3D stackable selector," in *2013 Symposium on VLSI Technology*, 2013: IEEE, pp. T240-T241.
[194] Y. Koo, S. Lee, S. Park, M. Yang, and H. Hwang, "Simple binary ovonic threshold switching material SiTe and its excellent selector performance for high-density memory array application," *IEEE Electron Device Letters,* vol. 38, no. 5, pp. 568-571, 2017.
[195] Y.-J. Lee, M. Han, S.-H. Yoo, and A. Soon, "Tunable threshold voltage of ZnTe-based ovonic switching devices via isovalent cation exchange," *ACS Applied Electronic Materials,* vol. 3, no. 3, pp. 1107-1114, 2021.
[196] A. Velea *et al.*, "Te-based chalcogenide materials for selector applications," *Scientific reports,* vol. 7, no. 1, p. 8103, 2017.
[197] J. Yoo *et al.*, "3D stackable and scalable binary ovonic threshold switch devices with excellent thermal stability and low leakage current for high-density cross-point memory applications," *Advanced Electronic Materials,* vol. 5, no. 7, p. 1900196, 2019.
[198] Y. Koo, K. Baek, and H. Hwang, "Te-based amorphous binary OTS device with excellent selector characteristics for x-point memory applications," in *2016 IEEE Symposium on VLSI Technology*, 2016: IEEE, pp. 1-2.
[199] R. Gu *et al.*, "Structural features of chalcogenide glass SiTe: An ovonic threshold switching material," *APL Materials,* vol. 9, no. 8, 2021.
[200] Y. Feutelais, A. Schlieper, and S. G. Fries, "Thermodynamic evaluation of the system Silicon-Tellurium," *Calphad,* vol. 23, no. 3-4, pp. 365-378, 1999.
[201] A. Schlieper, Y. Feutelais, S. G. Fries, B. Legendre, and R. Blachnik, "Thermodynamic evaluation of the Germanium-Tellurium system," *Calphad,* vol. 23, no. 1, pp. 1-18, 1999.

Band / Volume 95
A brain inspired sequence learning algorithm and foundations of a memristive hardware implementation
Y. Bouhadjar (2023), xii, 149 pp
ISBN: 978-3-95806-693-9

Band / Volume 96
Characterization and modeling of primate cortical anatomy and activity
A. Morales-Gregorio (2023), ca. 260 pp.
ISBN: 978-3-95806-698-4

Band / Volume 97
Hafnium oxide based memristive devices as functional elements of neuromorphic circuits
F. J. Cüppers (2023), vi, ii, 214 pp
ISBN: 978-3-95806-702-8

Band / Volume 98
Simulation and theory of large - scale cortical networks
A. van Meegen (2023), ca. 250 pp
ISBN: 978-3-95806-708-0

Band / Volume 99
Structure of two-dimensional multilayers and topological superconductors: surfactant mediated growth, intercalation, and dopingY.-R. Lin (2023), x, 111 pp
ISBN: 978-3-95806-716-5

Band / Volume 100
Frequency mixing magnetic detection for characterization and multiplex detection of superparamagnetic nanoparticles
A. M. Pourshahidi (2023), X, 149 pp
ISBN: 978-3-95806-727-1

Band / Volume 101
Unveiling the relaxation dynamics of Ag/HfO2 based diffusive memristors for use in neuromorphic computing
S. A. Chekol (2023), x, 185 pp
ISBN: 978-3-95806-729-5

Band / Volume 102
Analysis and quantitative comparison of neural network dynamics on a neuron-wise and population level
R. Gutzen (2024), xii, 252 pp
ISBN: 978-3-95806-738-7

Band / Volume 103
3D Scaffolds with Integrated Electrodes for Neuronal Cell Culture
J. Abu Shihada (2024), vii, 163 pp
ISBN: 978-3-95806-756-1

Band / Volume 104
Advances in Photoemission Orbital Tomography
A. Haags (2024), ix, 254 pp
ISBN: 978-3-95806-766-0

Band / Volume 105
Quantitative investigation of point defects and their dynamics in focused ion beam-prepared group III-nitride lamellas by off-axis electron holography
K. Ji (2024), 164 pp
ISBN: 978-3-95806-782-0

Band / Volume 106
NeuCoNS and Stacked-Net: Facilitating the Communication for Accelerated Neuroscientific Simulations
R. Kleijnen (2024), xx, 110, xxi-xxxiv pp
ISBN: 978-3-95806-788-2

Band / Volume 107
Construction of a Spiking Network Model of Macaque Primary Visual Cortex: Towards Digital Twins
A. Kurth (2024), xvi, 207 pp
ISBN: 978-3-95806-800-1

Band / Volume 108
Spin selectivity of chiral molecules on surfaces
M.R. Safari (2025), xiv, 165 pp
ISBN: 978-3-95806-810-0

Band / Volume 109
Redox-based Random Access Memory Arrays for Computing-In-Memory and Neuromorphic Computing
H. Chen (2025), x, 154 pp
ISBN: 978-3-95806-814-8

Information
Band / Volume 109
ISBN 978-3-95806-814-8

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum