# Wafer-level IV Characterizations of CMOS-transistors and ReRAM Devices

Electrical Characterization of Conventional and Emerging Memory Devices

School of Physical and Mathematical Sciences

Division of Physics and Applied Physics

Prof. Lew Wen Siang

Email: WenSiang@ntu.edu.sg

Version 3

| Version | Updated on | Updated by |
|---|---|---|
| 1 | Nov 2022 | Funan and Andhita |
| 2 | Feb 2023 | Funan and Andhita |
| 3 | Nov 2024 | Andhita |

A. Introduction to Memory Storage Technologies ..... 3

B. Resistive Random Access Memory (ReRAM) ..... 5

1. Recommended Readings ..... 5

2. Overview of ReRAM and CMOS transistor technology ..... 6

3. Objectives ..... 9

4. Experiment ..... 9

4.1. Experiment Setup ..... 9

4.1.1 Semi-Auto Probe Station ..... 9

4.1.2 Keysight Semiconductor Parameter Analyser ..... 14

4.2. Experimental Methodology ..... 18

4.3. ReRAM Characterisation ..... 19

4.4. Transistor Characterisation ..... 20

5. Questions ..... 20

C. Schedule ..... 21

D. Assessment ..... 22

# A. Introduction to Memory Storage Technologies

The advancement in data storage technology has allowed applications like cloud storage, internet of things (IoT) applications and artificial intelligence (AI) to be possible for consumer uses. These systems run on a variety of complementary metal-oxide semiconductor (CMOS) based random access memories (RAM) which can be classified as volatile memories and non-volatile memories. The difference between a volatile memory and a non-volatile memory (NVM) lies in the memory's ability to store it's data bits with or without a constant power supply respectively. A volatile memory, like the static random-access memory (SRAM) and the dynamic random access memory (DRAM), requires a constant voltage to retained its memory. On the other hand, NVM like the flash memory can continue to retain either bit '1' or '0' after the memory cell have been programmed. Apart from conventional NVM like flash memory, there are a number of emerging non-volatile memories which are being heavily researched on (f*igure 1*). These emerging memories, namely Magnetoresistive RAM (MRAM), Phase Change RAM (PCRAM) and Resistive RAM (ReRAM), have been reported to exhibit faster operating speed, higher switching endurance, higher memory retention and better scalability as compared to conventional memories.

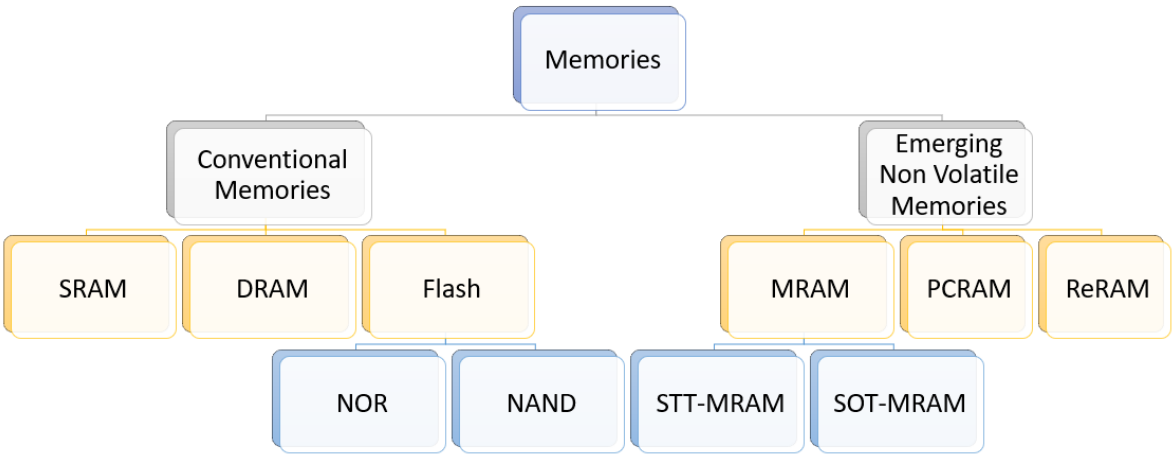



*Figure 1: Flow Chart Classification of Conventional and Emerging Non-Volatile Memories*

Volatile memories and charged based memory are widely used in commercial products, however due to the integration limitation (due to the structure of SRAM and DRAM) and memory degradation issues (flash memory), alternative memories with better scalability and better memory retention are being explored. Emerging memories such as MRAM, PCRAM

and ReRAM are potential candidates to replace NVM such as flash memory. These memories are reported to have high operation speed, good scalability for high density storage, high endurance and multilevel operation (*Table 1*).

| | Conventional Memories | | | | Emerging Memories | | |
|---|---|---|---|---|---|---|---|
| | | | FLASH | | | | |
| | SRAM | DRAM | NOR | NAND | MRAM | PCRAM | ReRAM |
| Cell Area | >100F$^2$ | 6F$^2$ | 10F$^2$ | <4F$^2$ (3D) | 6-50F$^2$ | 4-30F$^2$ | 4-12F$^2$ |
| Multibit | 1 | 1 | 2 | 3 | 1 | >2 | >2 |
| Voltage | <1V | <1V | >10V | >10V | <1.5V | <3V | <3V |
| Read Time | ~1ns | ~10ns | ~50ns | $\sim 10\mu s$ | <10ns | <10ns | <10ns |
| Write Time | ~1ns | ~10ns | $10\mu s - 1ms$ | $100\mu s - 1ms$ | <10ns | ~50ns | <10ns |
| Retention | NA | ~64ms | >10 years | >10 years | >10 years | >10 years | >10 years |
| Endurance | >$10^{16}$ | >$10^{16}$ | >$10^5$ | >$10^4$ | >$10^{15}$ | >$10^9$ | >$10^6$-$10^{12}$ |
| Write Energy (J/bit) | ~fJ | ~10fJ | ~100pJ | ~10fJ | ~0.1pJ | ~10pJ | ~0.1pJ |

*Table 1: Memories Performance Table*

These emerging memories can be considered as resistive switching memories due to their operating principles, utilising the change of device resistance values by the modulation of electronic spins and structural properties, i.e., phase, atomic defects, under external magnetic and/or electric field to represent memory bit “1” and “0” in digital sense. Typical MRAM devices consist of magnetic/insulator/magnetic structure, which work based on the different relative spin configurations between the fixed and free magnetic layer. When the magnetisation of the free layer is the same as the fixed layer, a low resistance state (LRS) is achieved, while high resistance state (HRS) can be produced when both magnetisations are in the opposite direction. PCRAM devices are made of a tri-layer metal/active material/metal structure where the active material can be tuned between amorphous phase (HRS) during RESET operation and crystalline phase (LRS) during SET operation. It utilizes the Joule heating generated under external electric field applied through the device. The RESET operation requires a shorter pulse with higher amplitude as compared to SET operation to enable rapid cooling down known as “quenched” to 'freeze' the atomic structure into a disordered (amorphous) state. Similarly, ReRAM devices comprise of a tri-layer structure with an insulator layer sandwiched in between

two metal electrodes. These types of devices work based on atomic defects modulation within the structure by the means of Joule heating and/or ionic migration induced under external electric field. Depending on the choice of electrodes and insulating layers, the type of defects involved in the switching operation of such devices might vary, i.e., oxygen vacancy and metal atoms. In general, more defects present in the oxide layer represent LRS, while lower defects represent HRS.

This experiment is split into two sections: Section B – Resistive Random Access Memory (ReRAM), and section C – nMOS transistors.

1. The working principles of ReRAM devices and their IV characterization
2. Device fundamentals and characterization of nMOS transistor

You are expected to give analysis of the electrical data and record your findings in a final report and presentation.

# B. Resistive Random Access Memory (ReRAM)

## 1. Recommended Readings


- Huang, P., et al. (2021). RRAM Device Characterizations and Modelling in Emerging Non-volatile Memory Technologies: Physics, Engineering, and Applications. (pp. 345-381). Springer, Singapore.
- P. R. Shrestha et al., "Analysis and Control of RRAM Overshoot Current," in IEEE Transactions on Electron Devices, vol. 65, no. 1, pp. 108-114, Jan. 2018, doi: 10.1109/TED.2017.2776860.
- F.-C. Chiu, “A review on conduction mechanisms in dielectric films,” Advances in Materials Science and Engineering, vol. 2014, pp. 1–18, 2014.

## 2. Overview of ReRAM and CMOS transistor technology

ReRAM have been extensively studied due to its promising characteristics in terms of scalability (two-terminal nature), energy consumption, cost per bit, and memory reliability. Most of these reported ReRAM structures are also known to be compatible with existing CMOS fabrication processes implemented in the semiconductor industry, thus ReRAM have been seen as one of the most promising candidates to replace the conventional memories. As mentioned, the HRS and LRS in a ReRAM is achieved by modulating the amount of defects presence within the oxide layer. In this experiment, we will particularly investigate a specific type of ReRAM device, which works based on localized defects induced under external voltage. The localized defects are usually referred as conductive filaments, which are formed during the SET process, connecting top and bottom electrode, resulting in LRS. The conductive filaments can be ruptured under RESET voltage operation to obtain HRS, as depicted in Figure 2.

The pristine devices are required to undergo an initial FORMING process, whereby its operation mimics the SET operation under higher input voltage. The conductive filament is formed through a soft dielectric breakdown in the metal oxide layer which produces the oxygen vacancies. This is known as the LRS or bit '1'. To create the HRS or bit '0', a voltage of opposite polarity will be applied through the top electrode. This will force the oxygen anions to combine with the oxygen vacancies and break the conductive filament. The switching between the HRS and the LRS can be represented in a Current-Voltage curve (IV curve), where the set voltage and the reset voltage can be determined directly.

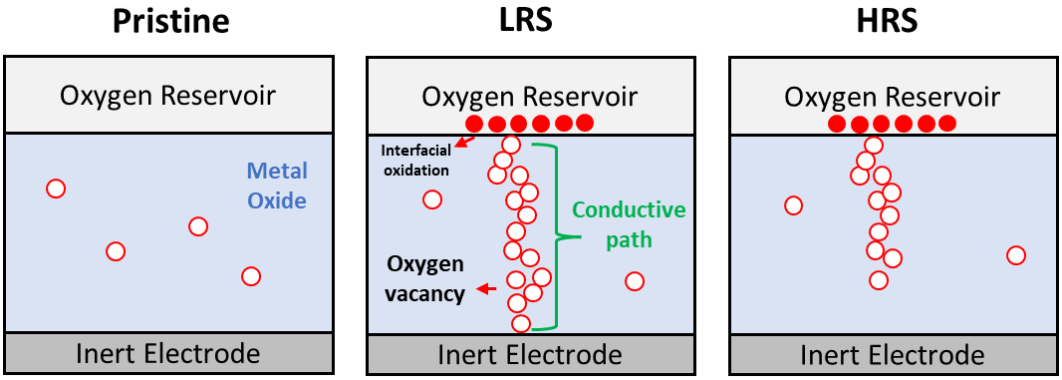



*Figure 2: Physical Representaton of the HRS (left) and the LRS (right) Resistive RAM*

The difference in resistance between the binary states of a ReRAM is known as the On-off Ratio. Ideally, the on-off ratio of a ReRAM is targeted to have be at least 10× to reduce the

possibility of having an overlap of resistance states. A lower LRS and a higher HRS can be created by controlling the thickness of the conductive filament and controlling the gap between the broken filament and the bottom electrode respectively. With different compliance current during the set process (switching from HRS to LRS), the thickness and the number of conductive filaments in the metal oxide layer will increase with higher compliance current. This will create a lower LRS. To control the HRS in the reset process (switching from LRS to HRS), the reset voltage can be varied accordingly. A higher reset voltage will break the conductive filament to a higher extent and a higher HRS can be achieved. From experiments, distinct binary states are hard to achieve. This is due to the variability in the set and reset voltage during its respective operation where the thickness of the conductive filament and the extend of rupture of the filament cannot be control precisely. Hence, this resulted in the perturbation of the states when they are being analysed statistically. Studies in the metal oxide material optimisation, film deposition techniques and AC pulse characteristics are ongoing to have a better control in this variability challenge hindering the application of ReRAM.

Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) is a type of field-effect transistor that makes use of the voltage applied at the gate terminal to control the conductivity of the channel between the source and drain terminals (*figure 3*). In the circuit level, a ReRAM is usually integrated on top of the drain terminal of a MOSFET for storage applications. The channel between the source and drain terminal determines the current flowing through the transistor. Transistor acts as an amplifier or a switch to the output signals from the connected memory cells and there are 2 classes of MOSFET known as the depletion mode and the enhancement mode MOSFET. These classes can be further divided into nMOS or pMOS which are n-channel (electron inversion layer) and p-channel (hole inversion layer) MOS transistor respectively. Both nMOS and pMOS are required to form a complementary metal-oxide-semiconductor (CMOS) for logic gate usage. A simple CMOS inverter, also known as the NOT gate, only makes use of 1 nMOS and 1 pMOS transistor while more complicated logic gate like the NAND gate and NOR gate make use of 2 nMOS and 2 pMOS each for their application.

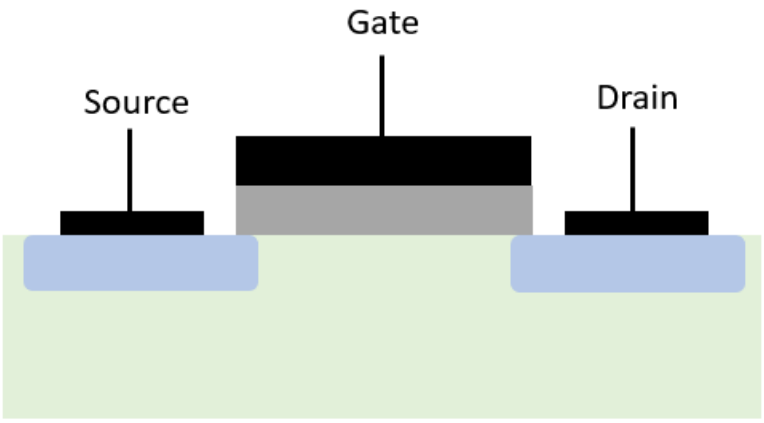



*Figure 3: Physical Representation of MOSFET*

A nMOS transistor is a more commonly used transistor as its carriers have a better mobility, allowing a faster operating speed as compared to pMOS. This type of transistor has a n-type source and drain terminals fabricated into a p-type substrate with a gate terminal made up of polysilicon. As mentioned, there are 2 types of nMOS where its channel is either the depletion mode or the enhancement mode type (*figure 4*). The enhancement type nMOS will be in the 'off' state unless being turned on using the gate voltage while the depletion type nMOS will be in a constant 'on' state unless being turned off. In general, an enhancement transistor will create an inversion layer near the surface, which will connect the source and drain terminals when the gate voltage ($V_{GS}$) exceeds the threshold voltage ($V_T$). When a small voltage is being applied to the source ($V_{DS}$), a current ($I_D$) will flow through the channel. As $V_{DS}$ increases, the $V_{GD}$ decreases and causes less inversion. The n-channel will be pinched off when the gate to n-channel voltage decreases to just below $V_T$. This happens when $V_{DS} = V_{DS(SAT)}$. On the other hand, the depletion transistor operates in the opposite way whereby the channel connects the source and drain terminals due to the diffusion of dopants during the fabrication process. For small $V_{DS}$, $I_D$ increases with $V_{DS}$. When $V_{GD} = - V_{DS}$, depletion layer will be formed and it will continues to be extended into the channel until $V_{DS} = V_{DS(SAT)}$, where the channel will be pinched off. Both nMOS transistors have a slight difference in their operation and it is common to have both transistors in a single circuit.

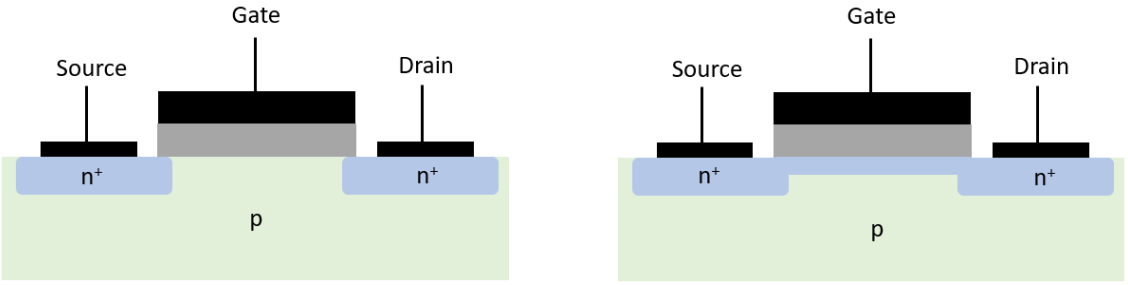



*Figure 4: nMOS Enhancement Mode (left) nMOS Depletion Mode (right) transistors*

## 3. Objectives

Students will be performing electrical testing of a ReRAM memory cell and a nMOS transistor by making use of the semi-auto probing machine and the Keysight semiconductor perimeter analyser. The electrical characteristics of each ReRAM device will be represented statistically to justify the performance between each cycle and between devices. Through the analysis, students are required to explain the application of ReRAM cells with a nMOS transistor in a 1-transistor-1-resistor (1T1R) architecture.

## 4. Experiment

### 4.1. Experiment Setup

#### 4.1.1 Semi-Auto Probe Station




*Figure 5: Semi-auto probe station 8 inch wafer setup*




*Figure 6: (Left) Power on button of the system. (Right) Emergency off (EMO) button in case of emergency.*

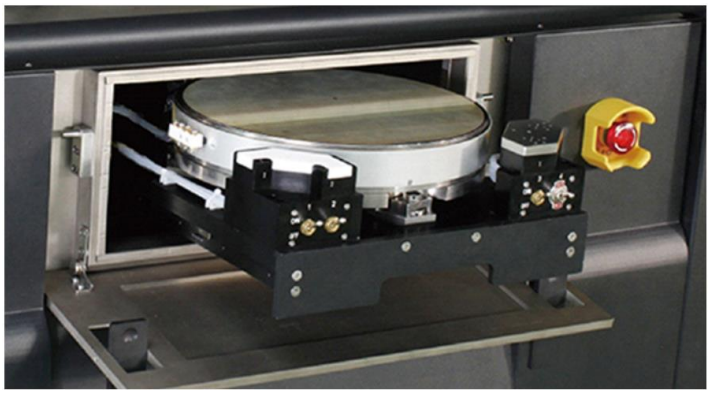


*Figure 7: Wafer stage with vacuum suction to hold the samples onto the stage*

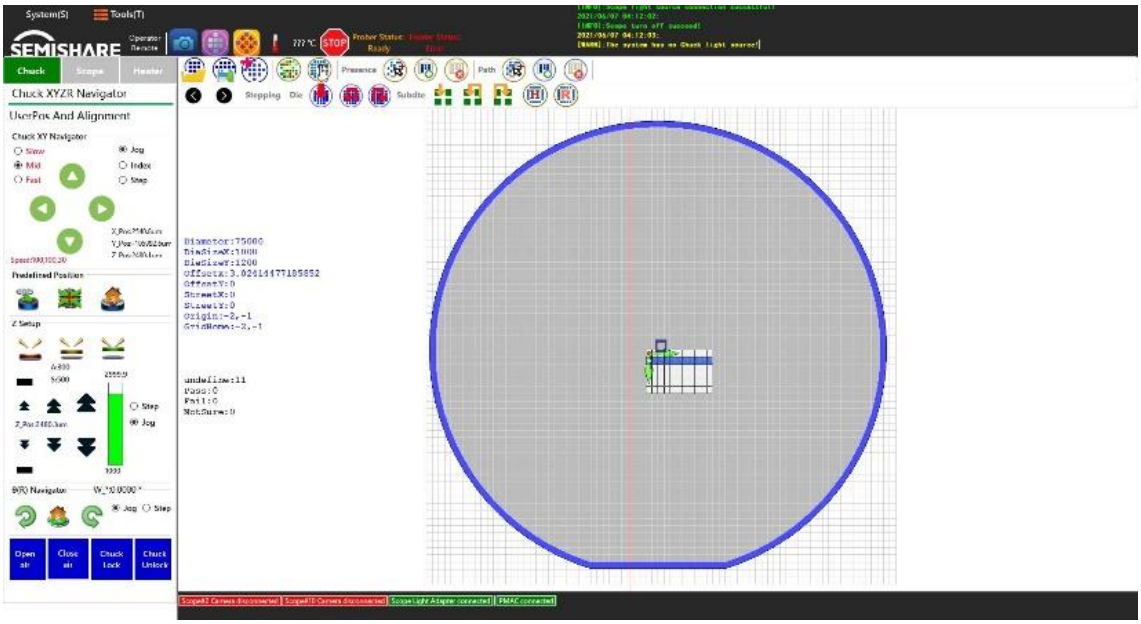



*Figure 8: User interface for the semi-auto probe machine*

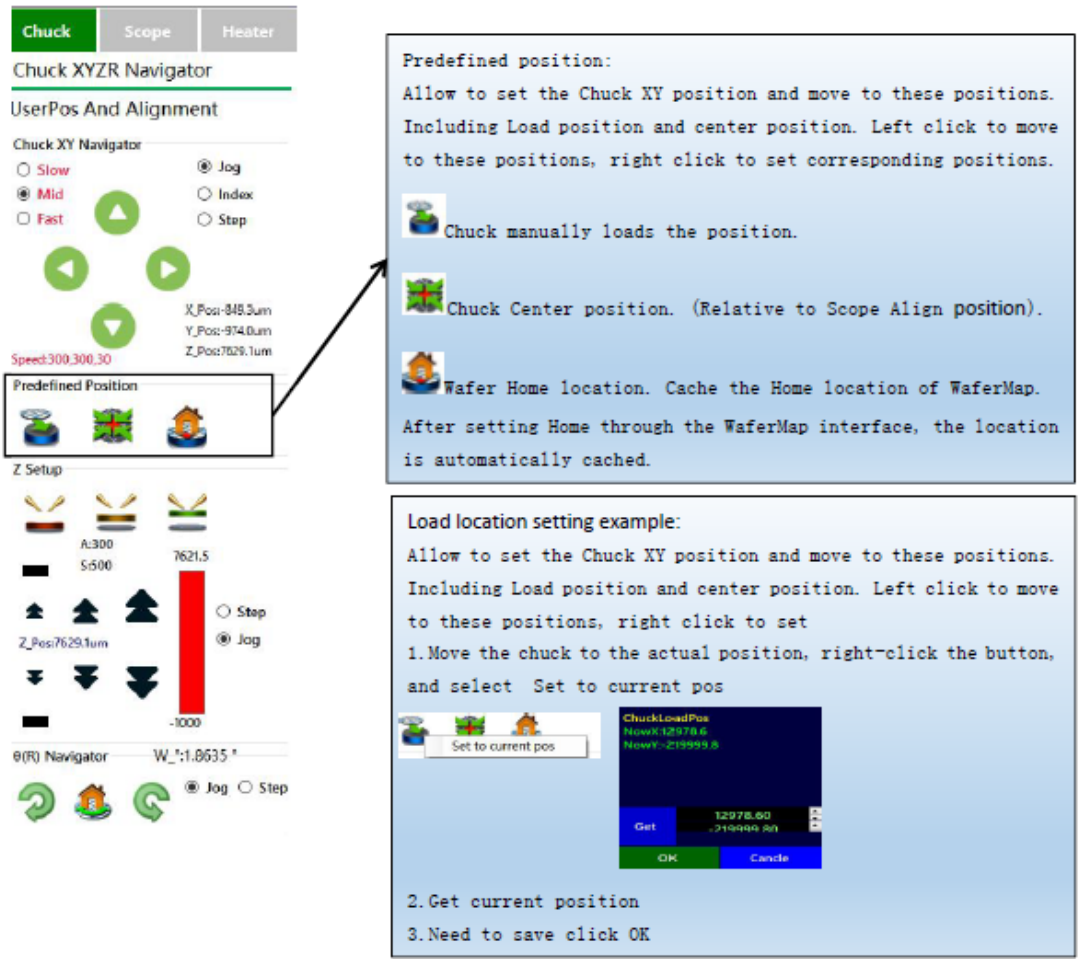



*Figure 9: Chumk movement control buttons for the semi-auto probe machine*

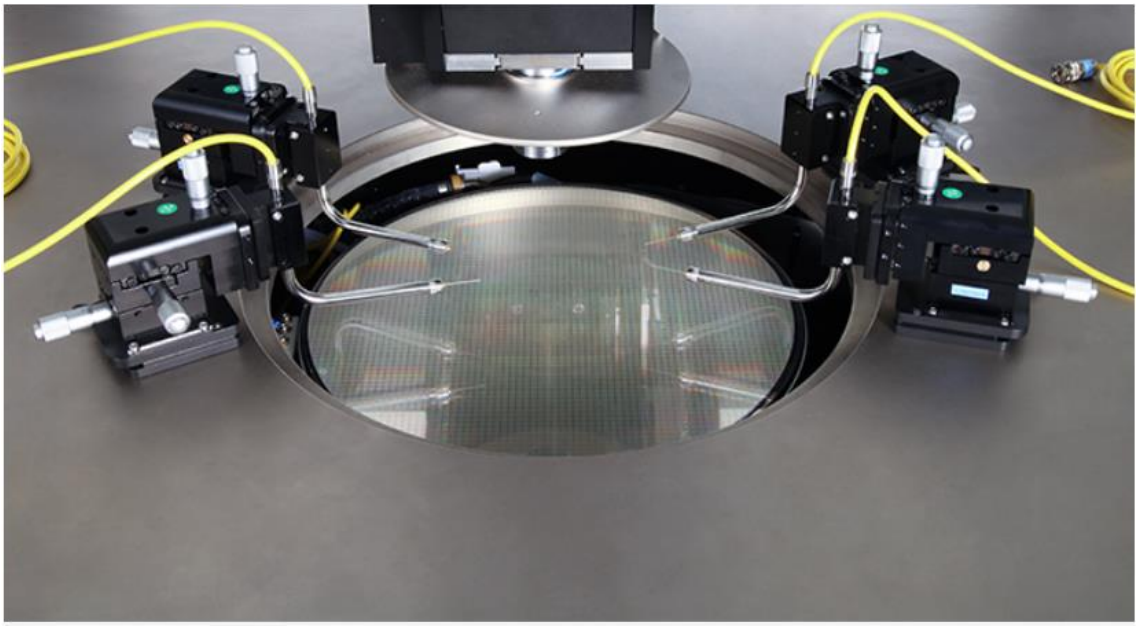


*Figure 10: Wafer stage at the 'Chunk Center' position*

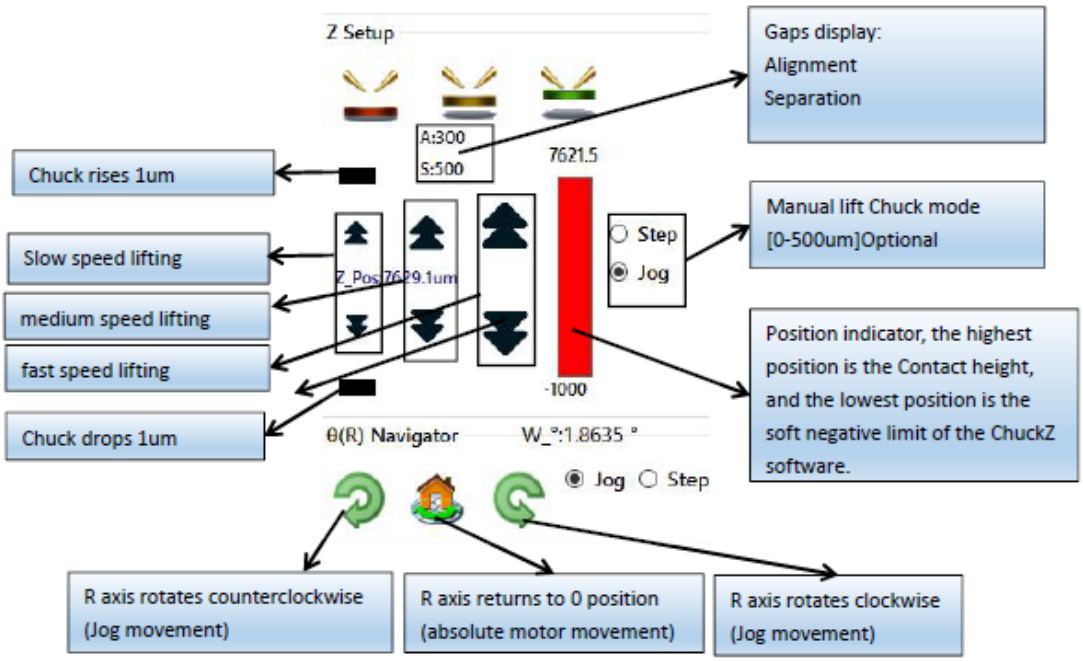



*Figure 11: Stage movement to control the probe tips distance from the wafer*

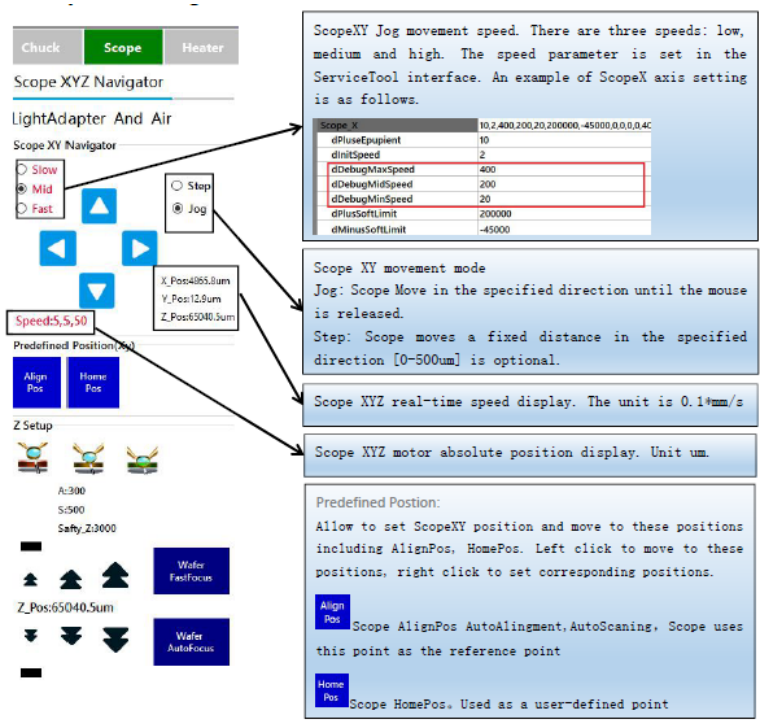



*Figure 12: Microscope movement to control the position of the microscope on the wafer*

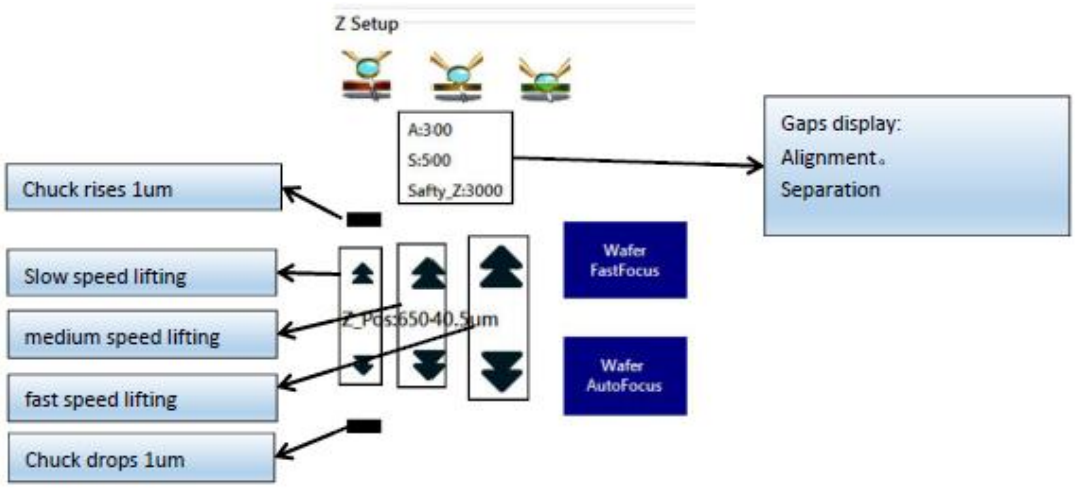



*Figure 13: Microscope movement to control the lens distance from the wafer*

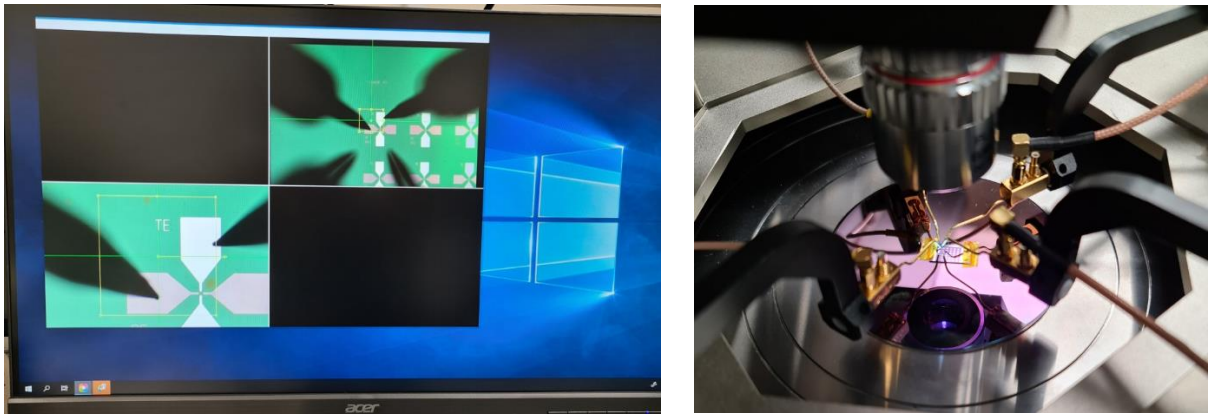


*Figure 14: Screen view and stage view of the wafer after successful probing on the contact pads*

### 4.1.2 Keysight Semiconductor Parameter Analyser

The semiconductor parameter analyser is a separate unit that works parallel with the semi-auto probe station to ensure that the electrical characterisation data is being obtained.



*Figure 15: Keysight semiconductor parameter analyser*




*Figure 16: Voltage input and information output on the source measurement unit (SMU)*

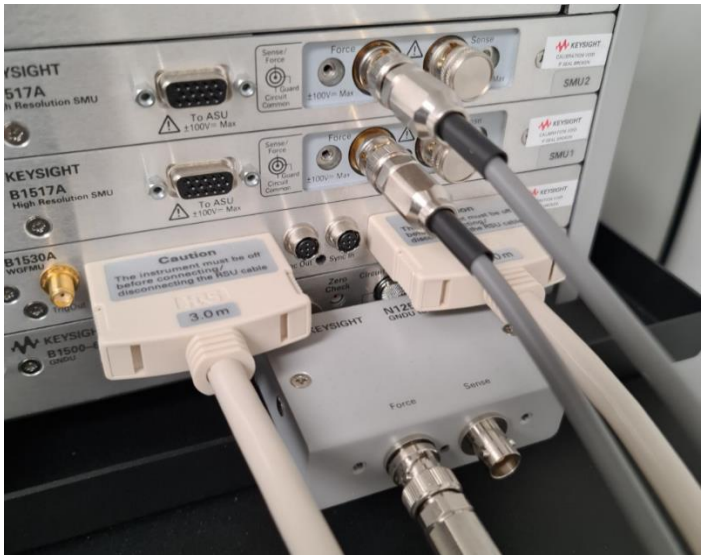



*Figure 17: Voltage and information output from the source measurement unit (SMU) to Keysight semiconductor parameter analyser*

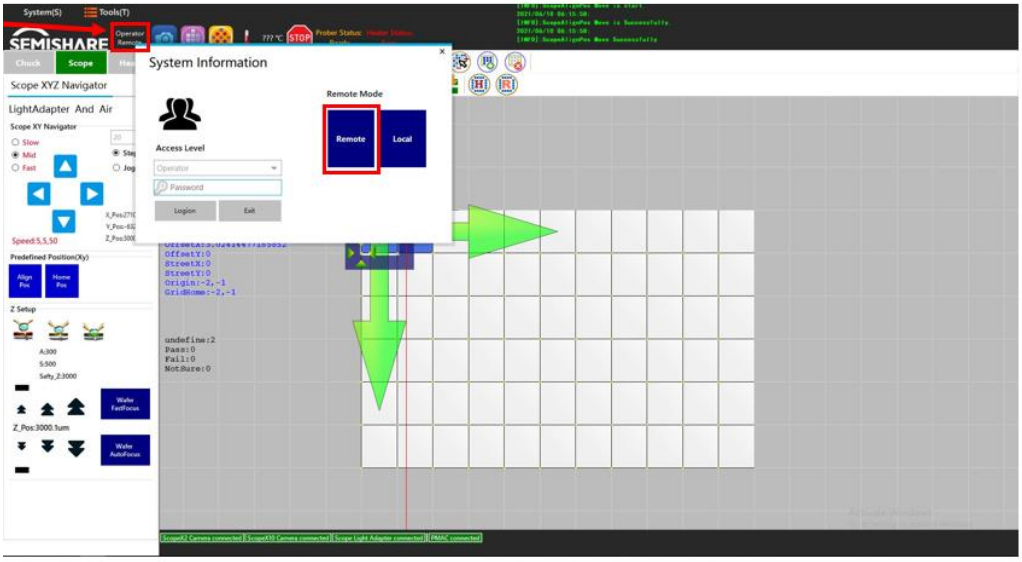



*Figure 18: Changing the control of the semi-auto probe station from 'Local' control to 'Remote' control for usage of Keysight analyser*

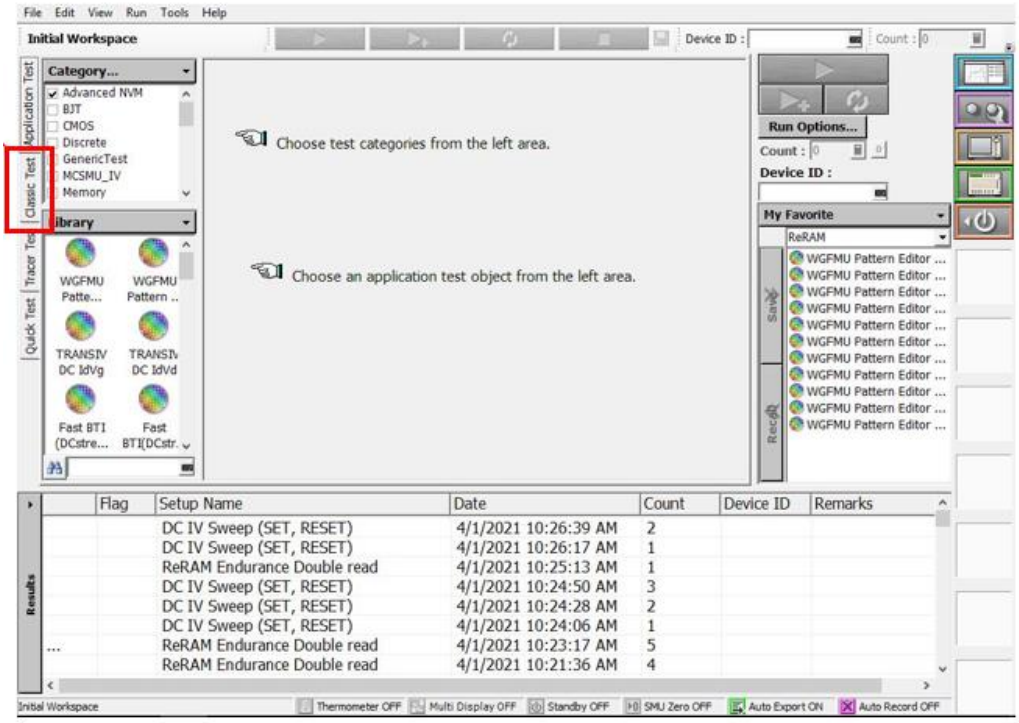



*Figure 19: Selection of 'Custom Test' to perform the IV sweep in the Keysight analyser*

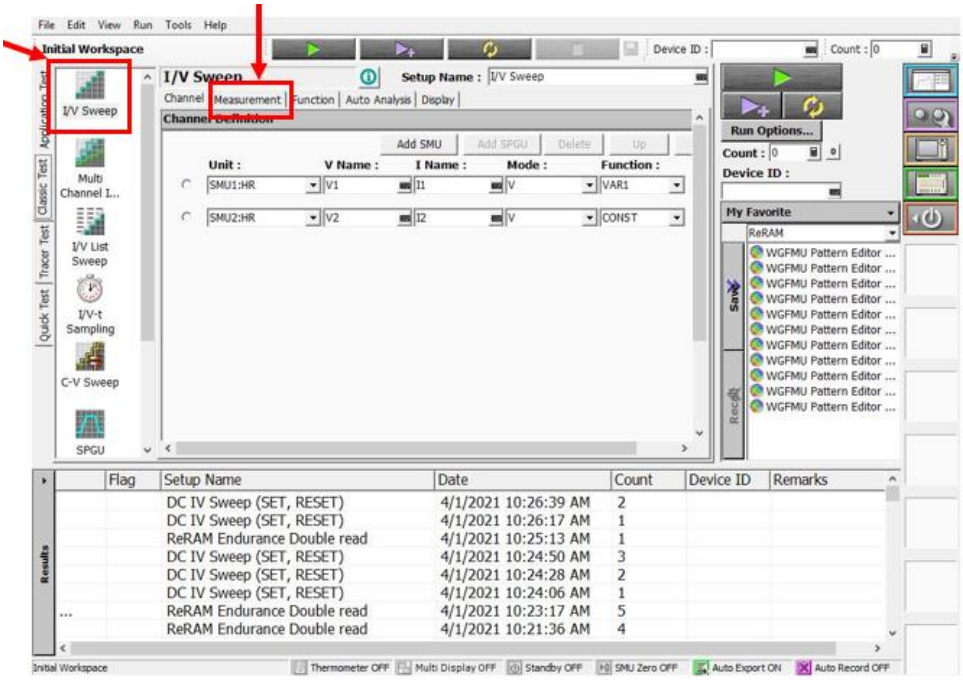



*Figure 20: Selection of 'IV Sweep' and 'Measurement' tab to perform the IV sweep in the Keysight analyser*

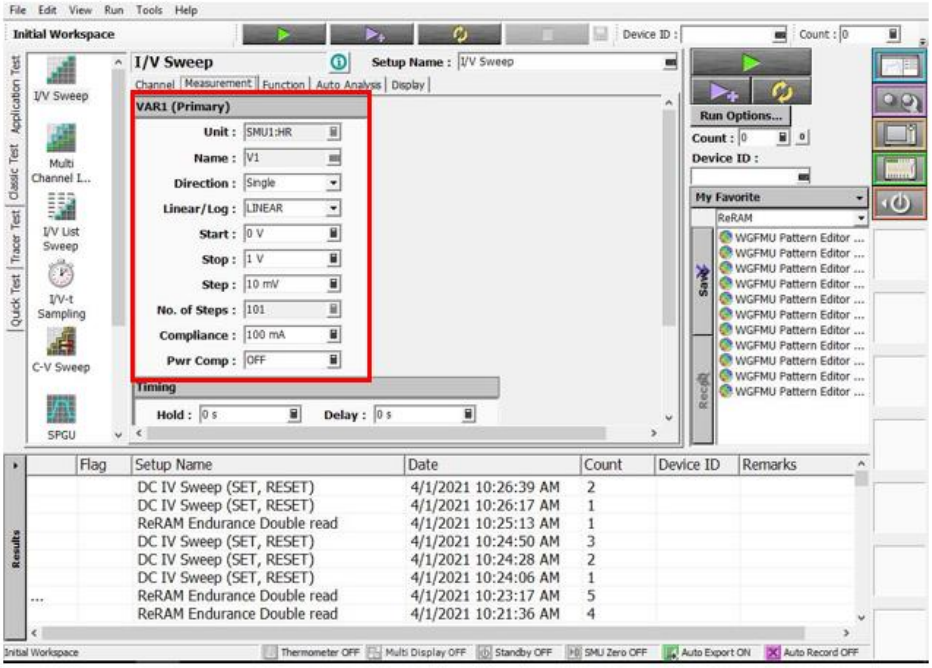



*Figure 21: Interface for inputting the starting voltage, ending voltage and the compliance current for the measurement*

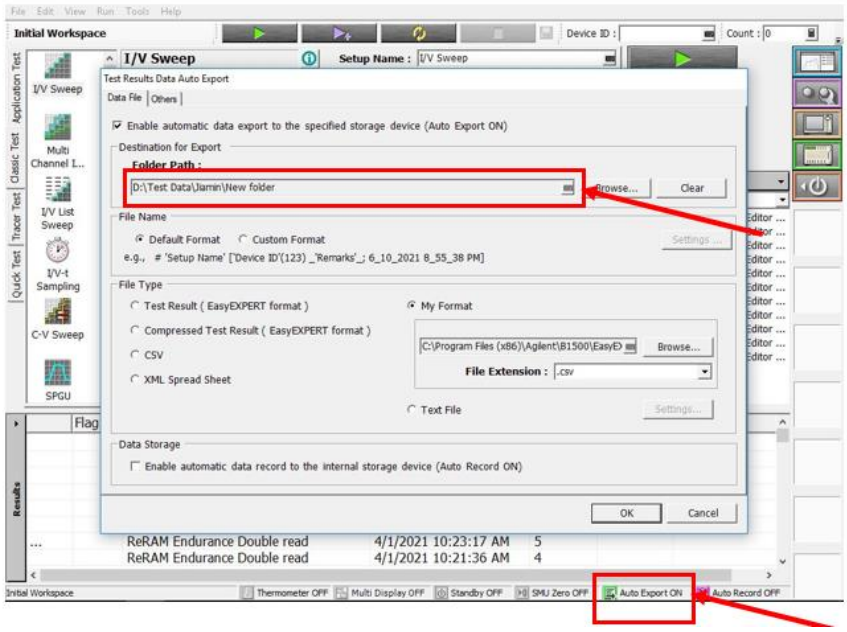



*Figure 22: Interface for the file directory that the data will be saved to*

## 4.2. Experimental Methodology

In this experiment, the students will investigate the electrical properties of there-terminal nMOS transistors (with different dimensions) and two-terminal ReRAM devices. The step-by-step required before conducting the electrical test as follows:

**Semishare Semi-automated probe station**

1. Please ensure you are wearing gloves when handling the samples.
2. Before loading the sample, the probe needles must be placed within the distance similar to the sample size and sufficiently far from the surface of the sample to avoid scratching the surface or damaging the needles.
3. Place the coupon-sized sample on the stage (chuck) and turn ON the vacuum suction to ensure the sample does not move throughout the measurement.
4. Manually load the chuck into the station and press the PRESET X-Y position using the SEMISHARE software.
5. Tune the microscope height by manual Z control or autofocusing feature
6. X-Y sample alignment step can be done by an automated two-point alignment feature or manual rotation.

**Keysight B1500 Semiconductor Analyser**

1. Check the necessary connections from the SEMISHARE probe station to the Keysight B1500 semiconductor analyser (SA).
2. The B1500 SA is operated by using EasyEXPERT software.
3. Depending on the number of input terminals required on the device, the program will be chosen accordingly.
4. For nMOS transistor measurement, the Id-Vd and Id-Vg IV characterization will be performed using B1530A waveform generator/fast measurement unit (WGFMU) with two timely synchronized channels.
5. For ReRAM devices, the IV characteristics will be measured by two B1517A High resolution source/monitor units (HRSMUs)

## 4.3. ReRAM Characterisation

The ReRAM device measurement will be divided into 3 parts:

1. Basic IV characterization, which involves the initial measurement of the device pristine resistance states, FORMING step, and subsequent SET and RESET. The students will optimize the input parameters, i.e., voltage and current limit (known as current compliance, CC), to achieve sufficiently high ON/OFF memory window (>10).
2. Fully automated measurement in which the students will use the optimized input parameters to run multiple cycles of SET and RESET across multiple devices. This is to investigate the temporal (cycle-to-cycle) and spatial (device-to-device) variation of the ReRAM device.
3. Multibit per cell characterization that consists of tuning the CC level during SET (to achieve multiple LRS) and the stopping voltage during RESET (to achieve multiple HRS).

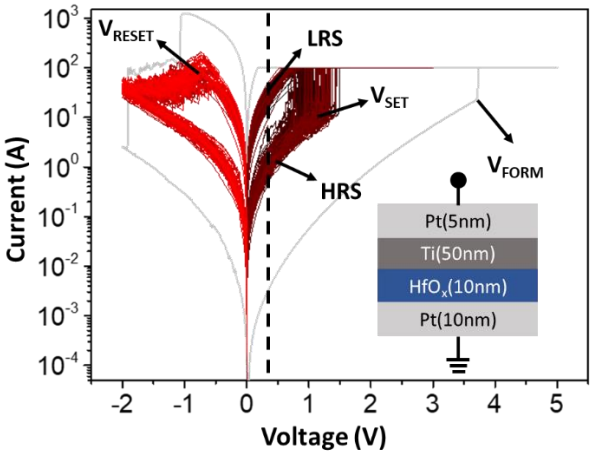



*Figure 23. Single DC cycle of a ReRAM IV performance curve*

Your report should include the following figures:

1. DC IV curves of ReRAM to explain the device performance through:
    a. On-off ratio
    b. Set/Reset Voltage
2. Cumulative probability distribution plots for the different resistance states of ReRAM
3. Box plots for ReRAM set and reset voltages in terms of:
    a. Cycle to cycle variability
    b. Device to device variability
4. Multibit per cell demonstration

## 4.4. Transistor Characterisation

The nMOS IV characterization is divided into two parts:

1. Output characteristics: Id-Vd plot under different Vg
2. Transfer characteristics: Id-Vg plot under different Vd

The same set of IV characterizations will be performed on two different transistor sizes with multiple devices each.

Your report should include the following figures:

1. Id-Vd and Id-Vg plot of the transistors
2. Statistical comparison of the current level under specific input voltages across devices

# 5. Questions

The report should also contain answers to the questions in the following section.

1) Are the ReRAM devices showing consistent performance in terms of initial resistance value and forming voltage? What is the yield of the sample and possible source of variation?
2) After the FORMING process, how is the cycle-to-cycle and device-to-device variation? What are the possible sources of the variability?
3) Did any of the devices failed to switch? If so, why?
4) Are the DC power supply measurements destructive to the device?
5) What are the possible ways to reduce the performance variability?
6) The ReRAM device is driven by nonlinear conduction mechanism, what is the ON/OFF ratio (memory window) achieved under different reading voltages?
7) What is the mode of nMOS transistor? What is the voltage range when the current is linear? (nMOS linear region)
8) What is the optimal $V_{GS}$ and $V_{DS}$ to ensure an optimal ReRAM device performance in 1T1R?

# C. Schedule

| Wafer-level IV Characterizations of CMOS-transistors and ReRAM Devices | |
|---|---|
| **Course Schedule (Five Weeks)** | |
| **Session 1** | 1. The students will learn how to operate the semi-auto prober and the Keysight semiconductor parameter analyser<br>2. IV characterization of the nMOS transistors |
| **Session 2** | 1. The student will perform DC IV measurement on the ReRAM devices to achieve optimized input parameters<br>2. The student will learn how to set up a wafer map for the fully automated measurements<br>3. The students will conduct the fully automated measurements to obtain statistical performance data of the ReRAM devices. |
| **Session 3** | 1. The students will tune the operating parameters of the ReRAM to achieve multibit per cell capability<br>2. The students will work on achieving at least four level of resistance states |
| **Session 4** | 1. The students will continue the work from session 3<br>2. Any additional measurements proposed by students or discussion on the analysis is welcome |
| **After completion of the experiment** | 1. Discussion on the report writing<br>2. Preparation of the slides for final presentation |

# D. Assessment

## Report Writing

At the end of this experiment, you are expected to prepare a report and a slideshow presentation to **elaborate on your understanding of the equipment, methods, and experimental findings**. The report should consist of the following sections:

- **Introduction**
  - Introduce different type of devices being measured and their corresponding working principles.
  - Describe the motivation and significance for performing the electrical measurements and analysis.
  - You are not allowed to copy the materials from this manual. However, you can add new information from other sources such as textbook, research journals, and review articles.
- **Methods**
  - Discuss briefly the different experimental setup and apparatuses being use for data collection.
  - Describe how the measurements were conducted, and important techniques and methods used in the experiment and results.
- **Results and Discussions** (this should form the main bulk of the report)
  - Present the collected data in the form of graphs and/or table form
  - Discuss the qualitative observations and quantitative analysis of the results based on the data collected. Explain its implications in terms of device performance and its applications.
- **Summary**
  - Conclude the experimental work and the discussion on the results.
  - Describe any challenges encountered and its proposed mitigation approaches.

All figures, diagrams, tables, and equations in the report should be **properly labelled** and placed appropriately alongside the description/discussion of itself. The report should seek to answer the questions. Emphasis is placed on **consistent and proper labelling, units, fitting, and error bars.**

# Presentation

- The presentation time for **single student is 20 min**, while joint presentation for **two students is 30 min with a combined slide.**
- There will be **20% - 40%-mark penalty** for late report submission or delayed presentation, without valid reasons.
- **Do not reuse** diagrams/descriptions in experiment manual / briefing slide significantly for your report. You are required to do reference search and write your analysis.
- Each report must show statement of **iThenticate check, with similarity less than 30%.**