THE BOOLEAN
SNAPSHOTS OF DOCTORAL RESEARCH AT UCC

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ionic Transistor - A New Generation Memory Device

Debarati Podder[1,*] and Paul Hurley[1]

[1] *School of Chemistry, University College Cork, Cork, Ireland*

[*] *Corresponding author: 119222129@umail.ucc.ie*

Abstract

We have come a long way since Alan Tuning first proposed the Artificial Intelligence (AI) in modern computers in 1950s enabling them to response like a human brain under certain conditions. But in order to perform various machine-learning operations such as image or speech recognition, huge datasets need to be processed leading to massive power consumption. Hence for the practical implementation and progress of AI with energy efficiency there is a pressing need of new class of memory devices which can mimic the performance of human brain at equivalent low energy. The focus of my PhD project is to develop such memory element by controlled incorporation of metal ions into the insulating layer in Metal Oxide Semiconductor (MOS) transistor which can be an innovative solution for muti-level (Analog; for reference, Binary system represents two levels), non-volatile (stored data retained even after power is off), Neuromorphic (mimics human brain response) memory device. Here I have reported controlled incorporation of lithium ions in an additional deposited insulating polymer layer in a metal-oxide-semiconductor capacitor and have shown that lithium ions motion in this layer can be controlled externally which enables it to modify the conductivity of the device, overall making it a promising candidate for the new generation memory element. Successfully integrating this with present silicon-based integrated circuits can lead to a breakthrough in AI in the future.

Keywords: artificial intelligence, neuromorphic computing, MOS transistor or capacitor, multi-level, non-volatile.

Nobody phrases it this way, but I think that artificial intelligence is almost a humanities discipline. It's really an attempt to understand human intelligence and human cognition.

— Sebastian Thrun

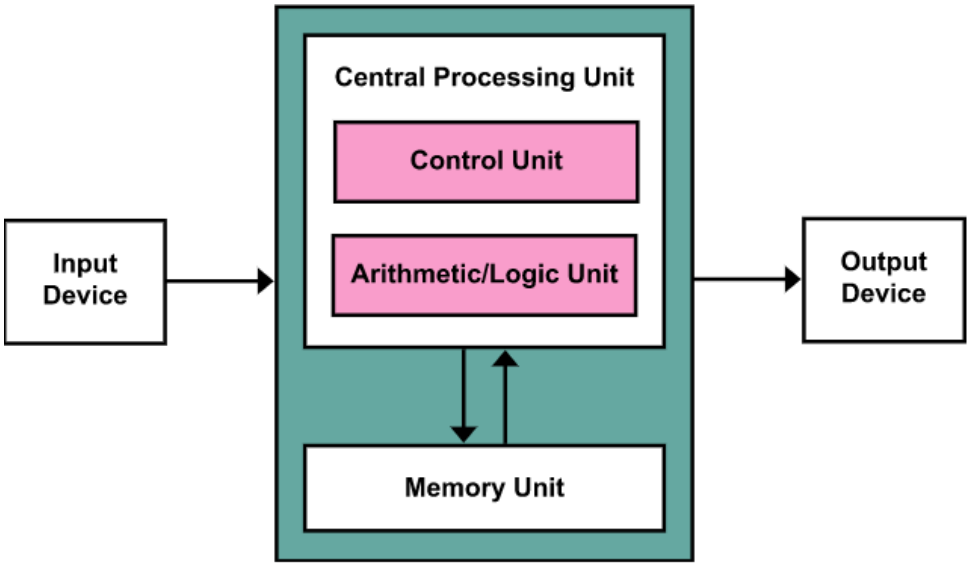

Figure 1: von Neumann architecture[4]

Introduction

Modern computers are based on the architecture that is called von Neumann architecture, proposed by John von Neumann in 1945. The basic principle of this architecture is that data and instructions are stored in memory units, while the computation is carried out in processors. So, the data and instructions need to be moved between memory and processors which are physically separated. This leads to the Von Neumann bottleneck: most of the energy and time are consumed in data movement, rather than computation.

The biological brain has a radically different architecture and function compared to conventional von Neumann computers. It has the following qualities:

- massively parallel, three-dimensionally organized, extremely compact
- amazingly energy efficient
- combines storage and computation
- fault and variation tolerant and robust
- self-learning and adaptive to changing environments

Hence, in data-centric computations, (e.g. machine-learning operations such as object, image, and speech recognition) von Neumann machines are inefficient in terms of computation time and energy consumption, whereas, Brain-inspired or in other words Neuromorphic computation system incorporating the above features can significantly improve data-centric computation tasks.

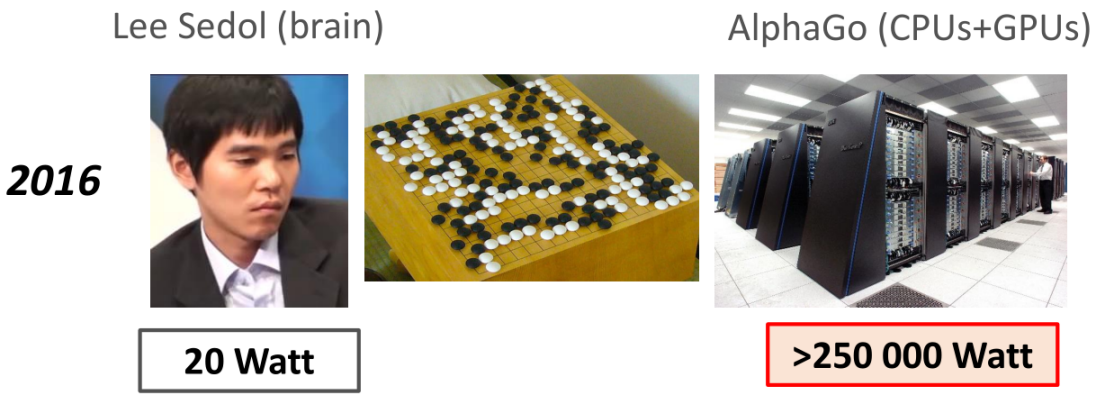

Figure 2: Real vs Conventional Computation.[1]

Why do we need New Generation Computation?

We are living now in the era of data, incorporating over 50 billion devices networked together; access to big computers, graphics cards and to gigantic datasets (billions of images for image recognition task) to train neural networks leads to massively huge data sets that need to be processed. Therefore, we need new models and device architectures for computation which can process information without any significant physical separation between computation and memory. This is where research into dedicated hardware for Neuromorphic Computing enters the picture.

Neuromorphic Computing: Hardware Realisation

Because of the requirement for processing vast sets of training data, AI computation software is only accessible to limited groups with extensive computation resources. The approach for personal smartphone apps is to upload to the cloud (i.e., a large server) with the results being sent back to the mobile phone. Hence, for environmental and practical reasons there is a pressing need to realise AI computation in hardware, at low power, on a local device.

Dedicated Integrated Circuits for Neuromorphic Computation: What is needed?

As mentioned above, the bottleneck in using conventional von Neumann architectures for performing AI tasks is the large amount of data exchange between the memory and computation sections of the computer. This suggests that the memory needs to be co-integrated with the microprocessor. One approach is to integrate the memory directly above the core logic of the processor using a monolithic 3D integration technology where multiple transistor layers are fabricated above a single substrate and thus being 3-dimensional integration reducing the feature size of the device. Also, if the memory could be comprised of multiple non-volatile memory levels as opposed to two levels – it could save on chip area as well as power! This requires innovative solutions for multi-level non-volatile memory, which can be processed at temperature which allows it to be integrated directly above an existing silicon microprocessor.

Neuromorphic Computation Element: Basic Examples

In the search for new type non-volatile memory solutions, different technologies have emerged in research over the past years which includes different materials and their properties to modify the conductivity of the device; despite being promising candidates for Neuromorphic hardware elements they all have some limitations, such as their conductivity switching characteristics between high conductive and low conductive states is non-ideal, also the number of times that the memory device can perform write and erase cycle is limited. More importantly, they have difficulty in achieving multilevel memory in true sense as they can mostly deliver two or three level memory states (for reference present memory devices works at binary system, so they have only two states 0 and 1).

My project: Overview of one of the most promising Neuromorphic Memory Hardware Elements

The Mainstream non-volatile memory technology of today is based on the complementary metal-oxide semiconductor (CMOS) technology, where complementary and symmetrical pairs of p-type and n-type MOS transistors incorporated circuits are used to perform various logic operations. So, the ideal scenario for hardware-based computation would be on a single integrated circuit, with the new neuromorphic functions being assimilated above an existing silicon-based CMOS integrated circuit.

The essential feature for on-chip, brain inspired computing comprises of:[2,3]

- low power, multilevel and non-volatile memory which can be programmed, read, and reset electrically.
- the device should ideally exhibit linear and symmetric programming and erase response.

Successfully achieving this device will open the door to energy-efficient brain-inspired computation on-chip, without the necessity to go to the cloud, with the associated latency and energy consumption.

Device Outline and Basic understanding of Switching

Human brain memory cells include neurons and synapses, where neurons transmit electrical and chemical signals which travel via synapses reaching other neurons and thus communication between different parts of brain and nervous system takes place. To mimic this action non-volatile electrochemical switches have been proposed as artificial synapses for neuromorphic computing. Such devices are referred to as ECRAM (Electro Chemical Random Access Memory).

Fig. 3 illustrates the device structure of ECRAM, which forms the focus of my PhD, where lithium (Li) ions in the electrolyte layer are controlled by the applied voltage and their movement results in change of its conductance for synaptic weight update. This device develops on

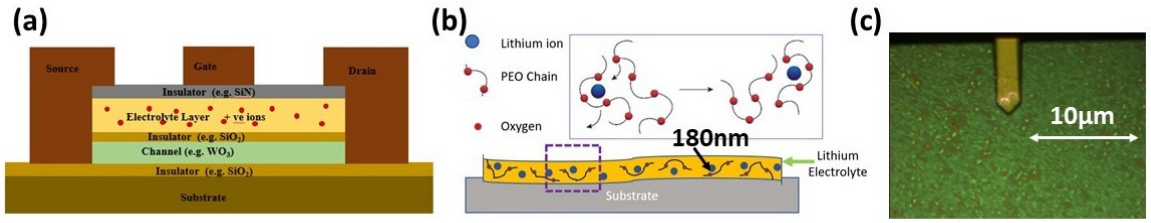

Figure 3: Device: Schematic diagram (a), deposited electrolyte polymer layer including lithium ions and polymer chain, lithium ions movement confined to this layer - Schematic Diagram (b), Optical Image of the polymer layer by AFM (c).

previously published work as the Li ions can modify the channel conductance without insertion and removal of the Li ions in the semiconducting channel. The effect is purely through electrostatics, and this has the potential to enhance the stability of the device. The amount of Li ions drifting in the electrolyte layer is precisely controlled by the gate voltage and this process is reversible, enabling symmetric update.

Electrical Measurement Results

To test the concept, a simplified device called a metal-oxide-capacitor (MOS) structures has been fabricated. This cuts down on the process steps needed to fabricate the full transistor, while alowing detailed examination of how the Li ion drifts in the polymer layer. To fabricate the device at this first stage, the MOS capacitor has been prepared with a silicon substrate incorporating a thin (around 85nm) silicon dioxide layer on top of it, cleaned and spun on with an electrolyte layer combining the polymer film doped with different concentration of a lithium compound in an appropriate solvent. On applying gate voltage the response of the lithium ions has been recorded, this has been acheieved using advanced probe station, voltage signal has been applied on the sample and corresponding impedance and phase angle are meaured, the applied voltage is a DC bias, super-imposed by a small AC voltage on top of that. The impedance and phase angle are then converted into a corresponding capacitance and conductance, which varies with the DC gate voltage and the AC signal frequency, hence we get capacitance vs gate voltage graph for a range of AC frequency for this device.

Figure 4 plots the capacitance versus gate voltage varying from -20V to + 20V for AC signal frequencies ranging between 1 kHz to 1 MHz. On applying negative gate voltage positive lithium ions are attracted towards gate making the polymer layer more conductive while positive gate voltage does the opposite, it repulses positive ions from the gate towards the other interface resulting decrease in conductivity of the polymer layer. For lower AC signal frequencies (1 kHz), the Li ions in the polymer can fully follow the AC signal, and as a result the electrolyte layer performs as an extension of the conductive gate, and the capacitance of the device is the same as maximum capacitance of the oxide layer in absence of the polymer layer. At higher frequencies, the Li ions can no longer follow the AC signal, and the electrolyte layer

acts as another dielectric layer, reducing the overall capacitance of the structure. These results demonstrate that the Li ion movement through the electrolyte layer can be controlled by the gate voltage and the AC signal frequency.

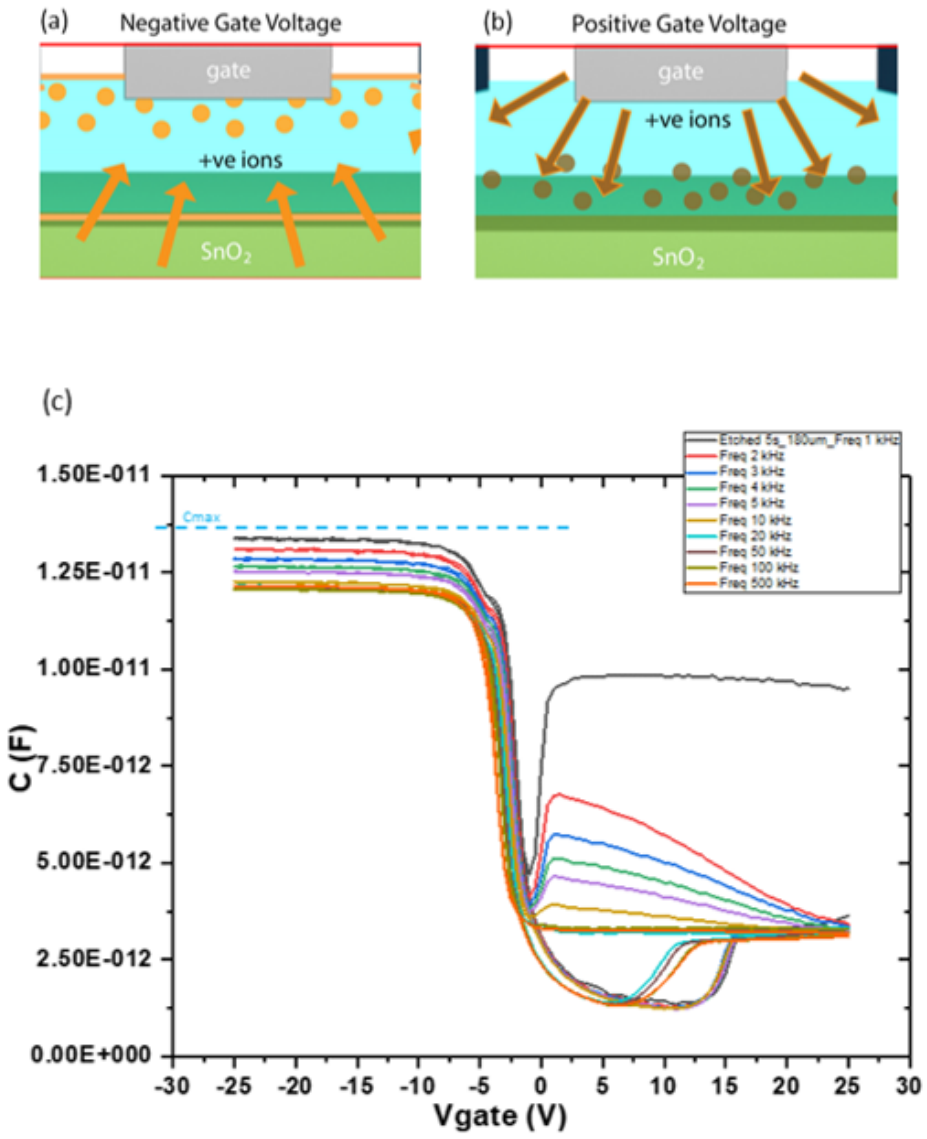

Figure 4: Schematic diagram of lithium ions response to applied gate voltage – negative gate voltage (a), positive gate voltage (b), Capacitance vs Voltage graph of the device (c).

Conclusion

The successful controlled incorporation of group 1 or group 2 metals like Lithium, Sodium, and Potassium (Li, Na, K) into the dielectric layer during deposition and to further explore the diffusion process and its dependence on voltage, time, and temperature gives us the opportunity to explore this process as a new class of multi-level non-volatile memory and to design a full device system to merge the fabric of memory and logic in future computation and this will enable to achieve a local device performing complex tasks at low power mimicking the biological

brain.

Acknowledgements

My sincere thanks to my supervisor Prof. Paul Hurley, co-supervisor Prof. Justin D. Holmes and my advisor Karim Cherkaoui for their guidance and motivation.

I would like to acknowledge the AMBER research centre (SFI-12/RC/2278) as the funding body for my research. Also, many thanks to all the staff and students of the Department of Chemistry (UCC) and the MNS Materials and Devices Group (Tyndall National Institute).

References

[1] Daniele Garbin. *A variability study of PCM and OxRAM technologies for use as synapses in neuromorphic systems*. PhD thesis, Université Grenoble Alpes, 2015.

[2] Daniele Ielmini and Stefano Ambrogio. Emerging neuromorphic devices. *Nanotechnology*, 31(9):092001, 2019.

[3] Jianshi Tang, Douglas Bishop, Seyoung Kim, Matt Copel, Tayfun Gokmen, Teodor Todorov, SangHoon Shin, Ko-Tao Lee, Paul Solomon, Kevin Chan, et al. ECRAM as scalable synaptic cell for high-speed, low-power neuromorphic computing. In *2018 IEEE International Electron Devices Meeting (IEDM)*, pages 13–1. IEEE, 2018.

[4] Kapooht Own Work. A von Neumann architecture scheme. CC BY-SA 3.0 https://commons.wikimedia.org/w/index.php?curid=25789639.