Mario Lanza[1*]
Yuanyuan Shi[1,2], Xianhu Liang[1], Bin Yuan[1], Victoria Chen[2], Haitong Li[2], Fei Hui[1], Zhouchangwan Yu[2], Fang Yuan[2,3], Eric Pop[2], H.-S. Philip Wong[2]
[1] Institute of Functional Nano & Soft Materials, Collaborative Innovation Center of Suzhou Nanoscience and Technology, Soochow University, Suzhou, 215123, China
[2] Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
[3] Department of Applied Physics, The Hong Kong Polytechnic University, Hong Kong

Mlanza@suda.edu.cn

Building the hardware of future artificial intelligence systems: two-dimensional materials based electronic synapses

Artificial intelligence (AI) systems are essential for modern societies because they can produce important progress in fields like economy (e.g. new jobs, powerful analysis tools), health (e.g. take care of elderly and children) and national security (e.g. autonomous machinery). Current AI systems rely on advanced computers to process a massive amount of data and carry out complex operations very fast (<1 ns/operation), and by using sophisticated algorithms they have been able to emulate some functionalities of animal brains (e.g. spiders, mice, cats). However, their computing capability and energy efficiency is still very far from that of human brains. The main reason is that traditional computers have a von Neuman structure, in which the data is computed in the central processing unit and stored in the memory unit. This produces a bottleneck that strongly limits the performance and enhances the power consumption of the entire system. On the contrary, the human brain uses an ultra-dense neural network to process and store the information in parallel. A human brain contains ~10^{12} neurons, each of them electrically connected to other ~1000 neurons through synapses; therefore, the total number of synapses in a human brain is ~10^{15}. The synapses are biological membranes that can change their conductivity (when they receive an electrical impulse from a neuron) by segregating Ca^{2+} ions. The learning is achieved by: i) generating new synapses, and ii) modifying the conductivity of the synapses (a.k.a. synaptic weight). Depending on the learning process, the changes on the synaptic weight (produced by neuronal spikes) can last in a timescale from milliseconds to minutes (i.e. short term plasticity, STP) or for minutes or longer (i.e. long term plasticity, LTP).

In order to create more powerful and efficient AI systems, electronic engineers have started to consider the possibility of designing new electronic circuits that act as artificial neural networks. Several different electronic components, including three-terminal devices or filed effect transistor based devices, ferroelectric switches and memory devices, have been suggested as the hardware implementation of electronic synapses for artificial neural networks. However, the designs have always resulted very complex and the performance too low (i.e. unable to emulate several simple learning rules that are readily carried out by biological synapses). Recently, two-terminal resistive switching (RS) devices based on a metal/insulator/metal (MIM) nanocells have been suggested as the ideal candidate device for the implementation of electronic synapses because their working principle resembles very well that of a biological synapse [1], i.e. changes in resistivity can be induced by applying electrical impulses to one of the electrodes, which produces the segregation of metallic and/or oxygen ions in the MIM cell. By using a wide range of metallic and insulating materials combinations, the emulation of LTP and its characteristic figures of merit has been readily achieved, including long term potentiation, long term depression, paired-pulse facilitation (PPF), paired-pulse depression (PPD), spike-timing dependent plasticity (STDP) and spike-rate dependent plasticity (SRDP) [2]. However, the traditional materials used in MIM-like electronic synapses have strong difficulties to generate progressive and linear resistivity changes, implement both STP and LTP in the same device, and generate stable and reproducible resistivity changes.

In our group we have solved some of these technological challenges by introducing two-dimensional (2D) materials in the structure of the RS devices. Some attempts of building electronic synapses using 2D materials have been reported [3], but those prototypes used planar configurations, which occupy much more space than vertical MIM cells, cannot be stacked three dimensionally and suffer from a higher device-to-device variability [4] (in fact, the RS industry only uses vertical configurations). Moreover, in most cases the 2D material was synthesized via mechanical exfoliation [5], which is not scalable and therefore unsuitable for building large-area synaptic networks. Here we report the first realization of vertical MIM-like electronic synapses using 2D materials produced by chemical vapor deposition (CVD), which is a scalable method [6]. By using multilayer hexagonal boron nitride (h-BN) sheets as RS medium, we have been able to implement both volatile and non-volatile RS simultaneously, which allows emulating several STP and LTP synaptic behaviors, including PPF, PPD, relaxation and STDP. The working regime can be selected by tuning the amplitude, duration and interval of the electrical stimuli. While until now all previous synaptic studies reported slow (0.1-100 s) erratic relaxation process during 0.1-100 s, here we show a fast (~200 µs) and stable relaxation during more than 500 cycles, with a very low variability. The power consumption of the synapses in volatile regime is 0.1 fW in standby and 600 pW per transition, and we find that the pulse voltage plays a more important role in the potentiation of the synapses than the pulse time/interval. These performances are enabled by a novel switching mechanism that combines characteristics from CBRAM and ReRAM. The volatile and non-volatile nature of the RS has been confirmed at the nanoscale via CAFM, demonstrating excellent potential for scalability. This work represents an important advancement for the development of electronic synapses in terms of performance and, as only scalable processes have been used, these methods may be employed to build 2D materials synaptic networks.

References

[1] P. A. Merolla, **Science** 345, 668-673 (2014).
[2] M. A. Zidan et al. **Nat. Electron**. 1, 22-29 (2018).
[3] H. Tian et al. **Nano Lett**. 15, 8013-8019 (2015).
[4] F. Hui et al. **Adv. Electron. Mater**. 3, 1600195 (2017).
[5] M. Wang et al. **Nat. Electron**. 1, 130-136 (2018).
[6] Y. Shi et al. **Nat. Electron.** 1, 458–465 (2018)

Figures

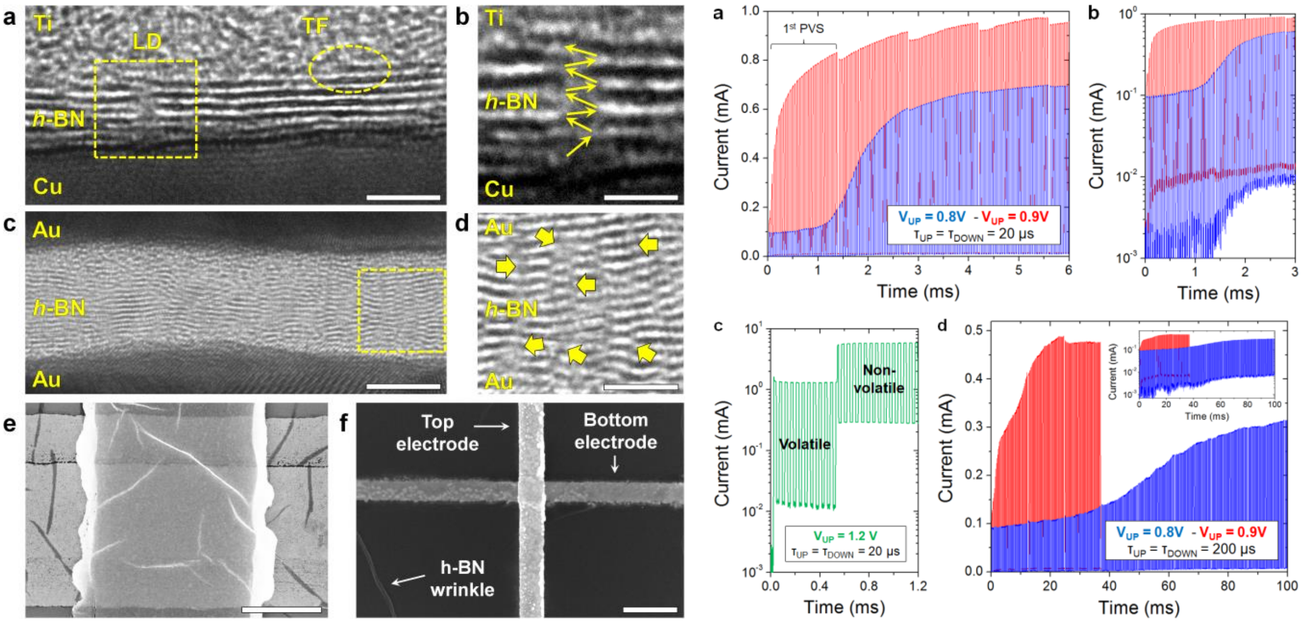

Figure 1: (Left) (a-d) Crsos sectional TEM images of the metal/h-BN/metal electronic synapses. (e-d) Top-view SEM images of the metal/h-BN/metal synapses with cross-point structure. (Right) (a-d) Potentiation of a h-BN based synapse by applying pulsed voltage stresses with different height, duration and interval. Reprinted with permission from Ref. [6]