Session 20: Memory Technology - RRAM for Neuromorphic Applications
Tuesday, December 4, 2:15 PM
Continental Ballroom 4
Co- Chairs: B. Magyari-Kope, Stanford University
R. Dittmann, Forschungszentrum Juelich

2:20 PM - 2:45 PM
20.1 20x Retention Improvement by Eliminating Resistance Relaxation with High Temperature Forming in 28 nm RRAM Chip, *X. Xu, L. Tai, T. Gong, J. Yin, P. Huang***, J. Yu, D. Nian Dong, Q. Luo, J. Liu, Z. Yu, X. Zhu, X. Long Wu**, Q. Liu, H. Lv, M. Liu, Chinese Academy of Sciences, Beijing, *University of the Chinese Academy of Sciences, **Anhui University, ***Peking University*

In this work, we proposed a high temperature forming scheme for 28 nm 1Mb RRAM test chip. Compared with room temperature forming scheme, the average forming voltage performed at 125 ºC could be greatly reduced from 2.5 V to 1.7 V. Resistance relaxation resulted from the recombination of Vo and O2- that generally occurred after programming was effectively eliminated as the residual O2-in the filament was highly decreased. Benefit from this,retention improvement of more than 20× times was successfully achieved, especially for LRS.

2:45 PM - 3:10 PM
20.2 Characterizing Endurance Degradation of Incremental Switching in Analog RRAM for Neuromorphic Systems, *M. Zhao, H. Wu, B. Gao, X. Sun*, Y. Liu, P. Yao, Y. Xi, X. Li, Q. Zhang, K. Wang**, S. Yu*, and H. Qian, Tsinghua University, *Georgia Institute of Technology, **Huawei Technologies Co., LTD.*

Resistive random access memory (RRAM) is attractive for neuromorphic computing systems as synaptic weights. In the neural network training, incremental switching occurs between the analog conductance states, thus the analog RRAM devices have unique endurance degradation behaviors compared to the convention digital memory application. In this work, a fast measurement platform is developed to characterize the endurance of incremental switching in analog RRAM. It is found that under weak weight update pulses, the incremental switching cycles of RRAM can be increased for more than 5 orders of magnitude compared with full window switching under strong programming pulses. The 1e11-cycle endurance of analog RRAM is proved to be sufficient for training neural networks online for various datasets (from MNIST to ImageNet). However, the nonlinearity and dynamic range of analog RRAM degrade during cycling, which may influence the learning accuracy of the neural network when it re-trains with new datasets.

3:10 PM - 3:35 PM
20.3 In-depth Characterization of Resistive Memory-based Ternary Content Addressable Memories, *D. R. B. Ly, B. Giraud, J-P Noel, A. Grossi, N. Castellani, G. Sassine, J-F Nodin, G. Molas, C. Fenouillet-Beranger, G. Indiveri*, E. Nowak and E. Vianello, CEA Leti, *University of Zurich and ETH Zurich*

Resistive Memory-based Ternary Content Addressable Memories (TCAMs) were developed to reduce cell area, search energy and standby power consumption beyond what can be achieved by SRAM-based TCAMs. In previous works, RRAM-based TCAMs have already been fabricated, but the impact of RRAM reliability on TCAM performance has never been proven until now. In this work, we fabricated and extensively tested a RRAM-based TCAM circuit. We show that a trade-off exists between search latency and reliability in terms of match/mismatch detection and search/read endurance. We show that a RRAM-based TCAM is an ideal building block in multi-core neuromorphic architectures. These ones would not be

affected by long latency time and limited endurance, and could greatly benefit from their high-density and zero standby power consumption.

3:35 PM - 4:00 PM
20.4 Mixed-Signal Neuromorphic Inference Accelerators: Recent Results and Future Prospects (Invited), *M. Bavandpour, M.R. Mahmoodi, H. Nili, F. Merrikh, Bayat, M. Prezioso, A. Vincent, and D.B. Strukov, K.K. Likharev*, University of California, Santa Barbara, *Stony Brook University*

Recent advances in analog-grade dense nonvolatile memories now enable extremely fast, compact, and energy efficient analog circuits. Such circuits are perfectly suited for implementations of the inference operation in neuromorphic networks. Here, we review implementations mixed-signal circuits, describe recent experimental demos of mixed-signal neuromorphic networks and outline urgently needed work.

4:00 PM Coffee Break

4:25 PM - 4:50 PM
20.5 Temporal sequence learning with a history-sensitive probabilistic learning rule intrinsic to oxygen vacancy-based RRAM, *J. Doevenspeck, R. Degraeve, A. Fantini, P. Debacker, D. Verkest, R. Lauwereins, and W. Dehaene, imec, KU Leuven ESAT*

Widely spread and low value resistance distributions inhibit the use of filamentary resistive RAM (RRAM) at low currents for deep learning training and inference. An entirely different approach which employs RRAM as active computational elements is proposed. For this means, the history-sensitive probabilistic reset in Tantalum-Oxide (TaOx)-based RRAM is characterized and explained. This intrinsic RRAM effect is used as a local learning rule in a novel temporal sequence learning algorithm.

4:50 PM - 5:15 PM
20.6 In-Memory and Error-Immune Differential RRAM Implementation of Binarized Deep Neural Networks, *M. Bocquet, T. Hirztlin*, J.-O. Klein*, E. Nowak**, E. Vianello**, J.-M. Portal and D. Querlioz*, Université de Toulon, CNRS, IM2NP, *Univ Paris-Sud, **CEA, LETI*

We fabricated differential HfO2-based memory arrays and their CMOS circuitry for in-Memory Computing. Our approach reproduces the reliability benefits of error correction, but without the associated CMOS overhead. It can implement ultralow-energy Binarized Deep Neural Networks, and allows using RRAMs at low voltage, where they feature outstanding endurance.

5:15 PM - 5:40 PM
20.7 A new hardware implementation approach of BNNs based on nonlinear 2T2R synaptic cell, *Z. Zhou, P. Huang, Y. C. Xiang, W. S. Shen, Y. D. Zhao, Y. L. Feng, B. Gao*, H. Q. Wu*, H. Qian*, L. F. Liu, X. Zhang, X. Y. Liu, and J. F. Kang, Peking University, *Tsinghua University*

For the first time, we propose a new hardware implementation approach which can utilize the non-linear synaptic cells to build a Binarized-Neural-Networks (BNNs)for online training. A 2T2R-based synaptic cell is designed and demonstrated by the fabricated RRAM array to achieve the basic functions of synapse in BNNs: binary weight (sign(W)) reading and analog weight updating (W+ΔW). The performance of BNNs based on 2T2R synaptic cells is evaluated by MNIST, and the recognition accuracy of 97.4% can be achieved. A novel refresh operation is proposed to enhance the network performance.