Confined PCM-based Analog Synaptic Devices offering Low Resistance-drift and 1000 Programmable States for Deep Learning

W. Kim[1], R.L. Bruce[1], T. Masuda[3], G.W. Fraczak[1], N. Gong[1], P. Adusumilli[1], S. Ambrogio[2], H. Tsai[2], J. Bruley[1], J.-P. Han[1], M. Longstreet[1], F. Carta[1], K. Suu[3] and M. BrightSky[1]

[1]IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, email: wkim@us.ibm.com
[2]IBM Almaden Research Center, San Jose, CA, USA, [3]ULVAC, Inc., Shizuoka, Japan

Abstract

We have demonstrated, for the first time, a combination of outstanding linearity of analog programming with matched PCM pairs, small analog programming noise, an extremely low resistance drift (R-drift) coefficient (0.005, median) and high endurance for a CVD-based confined phase change memory (PCM) with a thin metallic liner. In-depth analysis of linear analog programming is also presented. MNIST simulations using a pair of these confined PCM devices as a synaptic element yield a high test accuracy of 95%.

Keywords: PCM, confined PCM, metallic liner, analog programming, linearity, R-drift, noise, synaptic array

Introduction

PCM is a good candidate for use as synaptic elements to build prototype neural networks (NN). Because of the asymmetry between SET and RESET programming, synapses implemented with two PCM cells (2T2R) allow us to use only the SET programming for good accuracy of the NN (Fig. 12a). However, most reported PCM cells show considerable R-drift, non-linear analog programming behavior, and small number of programming states [1, 2]. In order to overcome these issues, we fabricated and characterized optimized confined PCM devices with a thin metallic liner (Fig. 1) based on [3, 4].

Results and Discussions

A. Improved characteristics with confined PCM A, B and C

After full integration, median R-V curve and SET speed evaluation of the A-type confined PCM array with a metallic liner are shown in Fig. 2. In order to improve endurance, we fabricated B-type (Fig. 3) and C-type confined PCM that also show good median R-V curve, fast SET speed (100ns), a low R-drift coefficient (0.005, median) and improved endurance (> $2x10^{11}$ cycles) as illustrated in Fig. 4. The metallic liner of confined PCM B is found to be continuous even after 10^{10} cycles (Fig. 3). A large grain size is observed in the void-free GST (Fig. 6a) approximately the same size as the pore dimensions. This fully crystalline GST with stable metallic liner results in enhanced endurance properties (Fig. 4b) and excellent linearity of analog programming curve (Fig. 5b,c), while maintaining low R-drift (Fig. 4e,f). Linear regression correlation coefficient (R^2) is utilized to measure linearity of analog programming. Characterization of data retention with these confined PCM devices is in progress. In addition, the improved CVD process is extremely scalable and can fill high aspect ratio pores with small critical dimensions (Fig. 6b).

Fig. 7 shows three different types of PCM cells that were programmed to a near minimum conductance level and monitored as a function of time. These measurements were analyzed to extract read noise and R-drift coefficient, with A-type PCM exhibiting both the lowest R-drift coefficient (0.006) and the lowest read noise. Using confined PCM A, we also performed analog programming followed by read-only measurements (Fig. 8). Programming noise is evaluated by the methodology based on Gaussian Process Regression (GPR) [5], and is found to be small and comparable to the read noise at a state close to the maximum conductance of the cell. In addition, this high-conductance state shows extremely small R-drift coefficient (0.0018). In summary, A-type, B-type and C-type PCM exhibit outstanding linearity of analog programming, small noise and mitigated R-drift (Fig. 5a).

B. Analog programming of A-type confined PCM devices

Different voltage pulse amplitudes and durations lead to different slopes of the analog programming curves in Fig. 9. By optimizing these critical parameters, we can achieve good linearity while also modulating the total number of states (200 to 1000 states, Fig. 10). With optimized programming conditions we found in Fig. 9, A-type confined PCM devices illustrate excellent linearity of analog programming (R^2: 0.993 with 1000 states) and small programming noise (small normalized difference in Fig. 11a). The coefficient (0.993) close to 1 indicates excellent linearity of analog programming. Tight device-to-device variation and cycle-to-cycle variation are also observed in Fig. 11b,c. By plotting the conductance difference of two neighboring PCM devices ($G = G_1 - G_2$), we observed symmetric behavior of the pair with confined PCM A (Fig. 12). R-drift measurements after analog programming of these devices to states between 40kΩ and 500kΩ all show extremely low R-drift (0.005 as median, Fig. 13).

C. MNIST simulation results with A-type confined PCM

By utilizing a 3-layer fully-connected neural network [6] with 528-200-10 neurons, MNIST simulations yield high accuracy (95% in Fig. 14) after 20 epochs with 2T2R scheme (2 PCM devices per synapse [7], Fig. 12b). These results (Fig. 14) show network behavior for simulated device arrays, where every device exhibits the same G-vs-pulse trajectory (blue), or array devices are randomly drawn from a small pool of measured trajectories, either individually (red) or in coupled pairs (orange). These accuracies are higher than the previously reported result (93.77% in [8]) due to better linearity of the confined PCM. Similar simulations with a 4T4R scheme encoding weights into two pairs of PCMs (one Least Significant Pair and one Most Significant Pair, Fig 12d) showed promise for even higher accuracy (97.4%).

Conclusions

We have demonstrated, for the first time, excellent linearity of analog programming with matched PCM pairs, mitigated resistance drift, great endurance and small analog programming noise for a confined PCM with a metallic liner. These outstanding characteristics lead to high test accuracy (95%) during a simulation using the MNIST database. Thus, the confined PCM shows great potential as elements in a future analog synaptic array for deep learning.

References

[1] M. Suri, *IEDM*, 2011. [2] I. Boybat, *PRIME*, 2017.
[3] M. BrightSky, *IEDM*, 2015. [4] W. Kim, *IEDM*, 2016.
[5] N. Gong, *Nature Communications*, 2018. [6] W. Haensch, *Proc. of IEEE*, 2019. [7] G.W. Burr, *IEEE TED*, vol. 62, no. 11, 2015.
[8] S. Ambrogio, *Nature*, vol. 558, pp. 60–67, 2018.

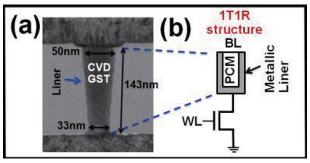

Fig. 1 (a) Confined PCM A with metallic liner after full integration. (b) Schematic view of the confined PCM cell (1T1R).

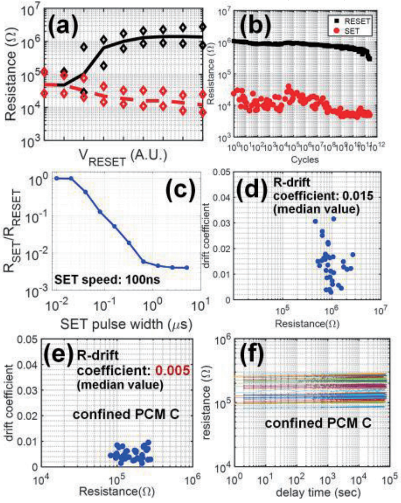

Fig. 4 (a), (b), (c), (d) Improved electrical characteristics with confined PCM B. (e), (f) Further reduced R-drift with confined PCM C (R-drift coefficient: 0.005).

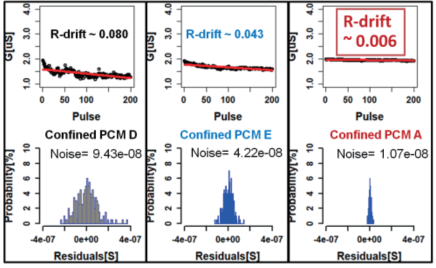

Fig. 7 Analysis of read noise and R-drift with different types of confined PCM. Confined PCM A showed lowest R-drift coefficient as well as smallest read noise.

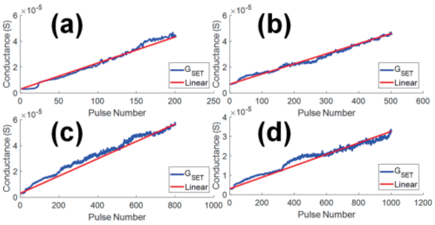

Fig. 10 Linear analog programming curves with different number of states from 200 states to 1000 states (confined PCM A).

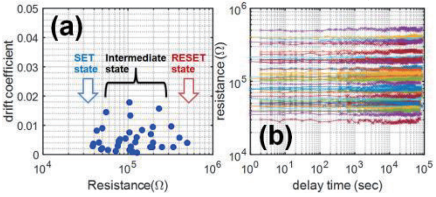

Fig. 13 R-drift of confined PCM A after analog (gradual SET) programming. Different analog programming results in different states between SET state and RESET state. This characterization after analog programming is performed, for the first time, with confined PCM array.

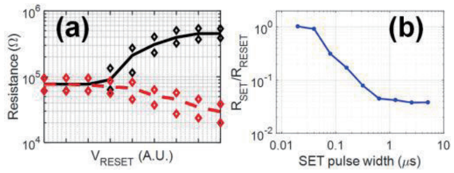

Fig. 2 (a) Median R-V curve and (b) SET speed measurement of the confined PCM A

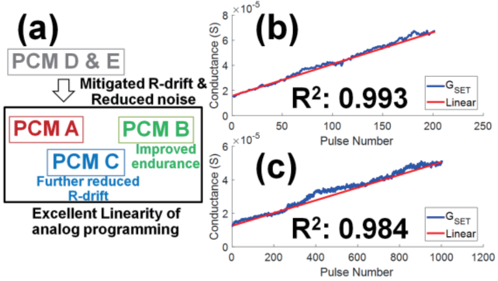

Fig. 5 (a) Schematic figure for different types of PCM. (b), (c) Linear analog programming curves with confined PCM C.

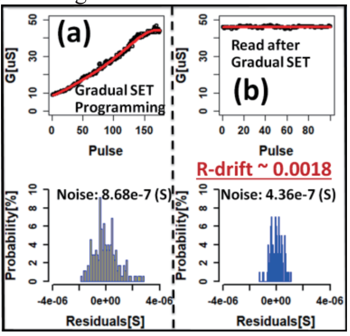

Fig. 8 Analysis of programming noise and read noise with confined PCM A. The high-conductance state shows extremely small R-drift coefficient (0.0018).

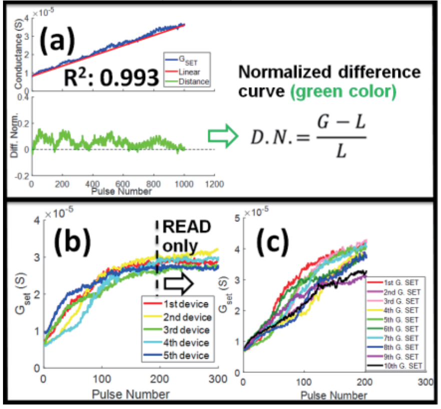

Fig. 11 (a) Outstanding linearity, (b) device-to-device variation and (c) cycle-to-cycle variation of analog programming with confined PCM A. Identical pulses are used for (b) and (c).

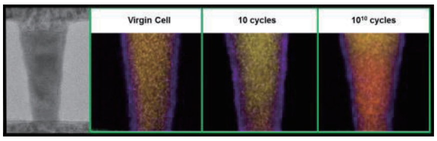

Fig. 3 TEM/EDX map overlay of confined PCM B showing composition of Sb (red) and Te (yellow) over number of cycles. The thin metallic liner is found to be continuous after 10^{10} cycles.

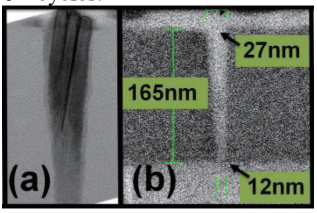

Fig. 6 (a) Void-free fully crystalline GST with a large grain size is observed (confined PCM B). (b) Great GST fill has been demonstrated on small dimensions with high aspect ratios.

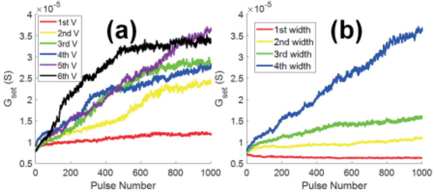

Fig. 9 Two critical parameters of gradual SET programming (pulse amplitude and pulse duration).

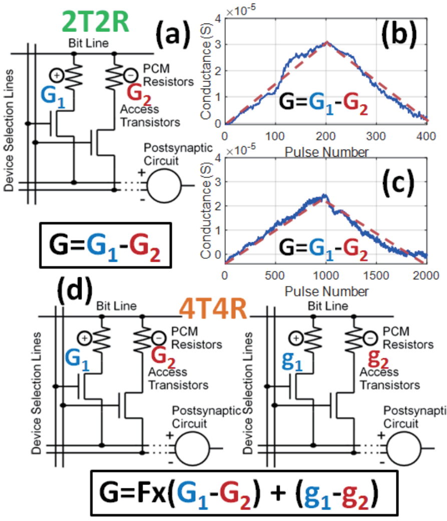

Fig. 12 (a), (b), (c) 2T2R scheme and (d) 4T4R scheme (4 PCM devices per synapse, [8]) for gradual weight increase and decrease with confined PCM A. Two conductance curves with great symmetry in Fig. 12 (b),(c) are measured, for the first time, from an identical set of PCM devices.

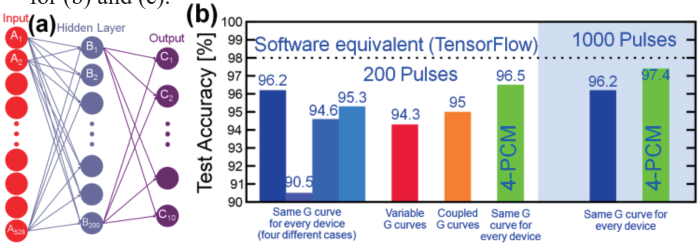

Fig. 14 (a), (b) By utilizing a 2T2R scheme, MNIST simulation yields a high training accuracy of 95% (orange color) with a small neural network (528-200-10 neurons) after 20 epochs. Similar simulations with a 4T4R scheme (Fig. 12d) encoding weights into two pairs of PCMs show even higher accuracy (97.4%, green) based on the data in Fig. 12c.