Supplementary Information: Shape-Dependent Multi-Weight Magnetic Artificial Synapses for Neuromorphic Computing

Thomas Leonard[1], Samuel Liu[1], Mahshid Alamdar[1], Can Cui[1], Otitoaleke G. Akinola[1], Lin Xue[2], T. Patrick Xiao[3], Joseph S. Friedman[4], Matthew J. Marinella[3], Christopher H. Bennett*[3] and Jean Anne C. Incorvia*[1]

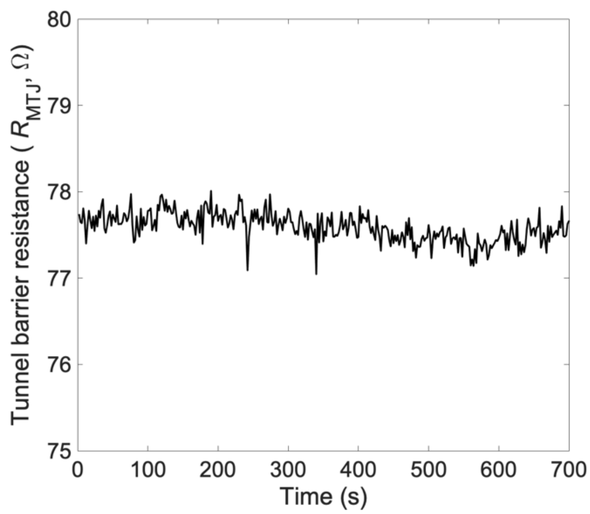
[1]*Electrical and Computer Engineering Dept., University of Texas at Austin, Austin, TX, USA.*

[2]*Applied Materials, Santa Clara, CA, USA.*

[3]*Sandia National Laboratories, Albuquerque, NM, USA.*

[4]*Electrical and Computer Engineering Dept., University of Texas at Dallas, Richardson, TX, USA.*

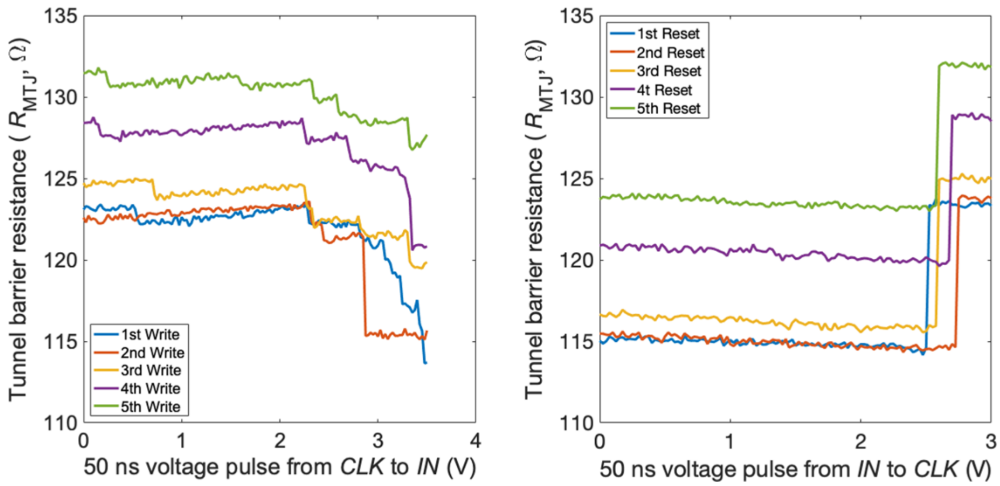
S.1 MULTI-WEIGHT STABILITY



Supplementary Figure 1. Trapezoidal DW-MTJ resistance over time when in middle MW state at room temperature.

In Supplementary Fig. 1, the trapezoidal synapse was set to the middle state and measured repeatedly for over 10 minutes, showing nonvolatility at room temperature.

S.2 SECOND TRAPEZOIDAL SYNAPSE BEHAVIOR



Supplementary Figure 2. Second trapezoidal synapse behavior. a, Write and **b**, reset cycling showing MW switching in one direction and single-shot reset behavior after each MW write.

In Supplementary Fig. 2, we show results of five write/reset cycles of a different trapezoidal synapse device from the same set of devices as the main text. Here, the same method is used as in the main text: $+V$ is applied to push the DW from *IN* to *CLK* showing multiple weights, and then $-V$ is applied showing a single switch back as the DW moves from *CLK* to *IN*, and this is repeated over 5 cycles without resetting the device in-between. The DC bias field is adjusted to promote efficient switching, alternating between -5 mT for MW writing and -45 mT for single-shot resetting. Five consistent states are observed in the MW direction and a single switch in the reset direction, but this device had lower *TMR* = 8% due to fabrication imperfections, which also led to drift of the resistance states during measurement.

S.3 ALGORITHMIC REPRESENTATION OF METAPLASTIC UPDATE

The metaplastic function modifies the returned optimizer function using the following algorithm:

for W^H in $\mathbf{W^H}$**:**

  if $U_W * W^b > 0$: (if the update decreases the magnitude of the weight)

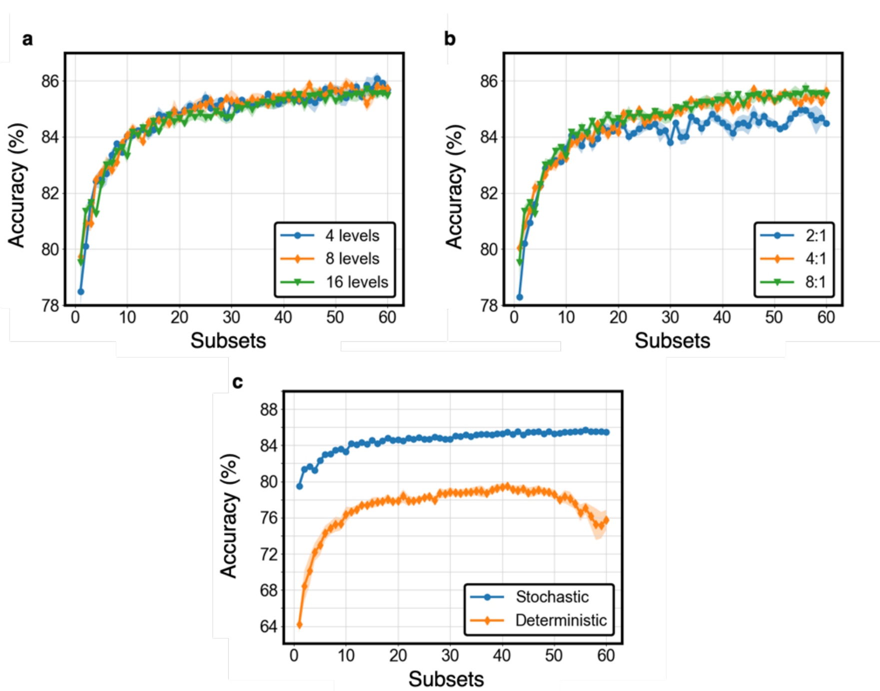
    $W^h \leftarrow W^h - \eta U_W * f_{meta}(W^h)$

  else: (if the update increases the magnitude of the weight)

    $W^h \leftarrow W^h - \eta U_W * f_{meta}(W^h) * \delta$

where U_W is the prescribed update obtained from the optimizer, W^b is the binarized weight, η is the learning rate, and δ denotes an asymmetric modification in the update. Typically, $\delta > 1$ is set so that significant weights are learned more quickly and $\delta = 1$ indicates a symmetric response.

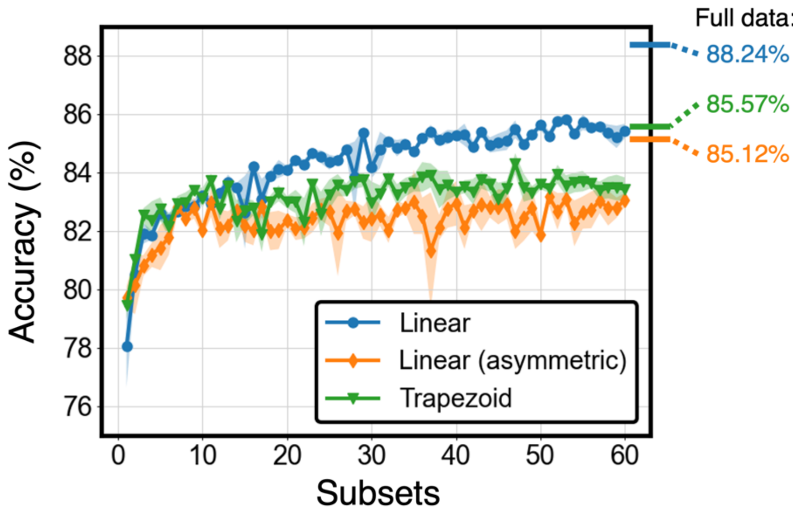
S.4 EXTENDED BINARIZED METAPLASTIC SYNAPSE STREAM LEARNING PERFORMANCE



Supplementary Figure 3. Additional data on trapezoidal DW-MTJ stream learning. a, Validation accuracy of binarized networks quantized to 4, 8, and 16 levels per device. **b**, Validation accuracy of a binarized network with 16-level quantized metaplastic weights at various width ratios. **c**, Validation accuracy of binarized network with 16-level quantized metaplastic weights using stochastic and deterministic rounding.

In Supplementary Fig. 3, stream Fashion-MNIST training performance of a binarized neural network at varying metaplastic weight quantization is shown. The quantization does not impact the performance since quantization is masked by binarization and small updates are not lost due to the stochastic rounding scheme. Supplementary Fig. 3b shows a sweep of varying trapezoidal slopes which modify the metaplastic function. The best performance comes from the largest ratio, although there are diminishing returns. However, the binarized network with the trapezoidal network performed similarly to that of the linear synapse in the non-binarized case. This demonstrates that a binarized network is still preferable to continuous weights in online learning situations due to less information overhead and the ability to reduce the size of the sensing MTJ. In Supplementary Fig. 3c, the effect of stochastic rounding is demonstrated. With weight quantization, there is a significant accuracy penalty for deterministic rounding. This is because small updates that fail to move the DW to the next notch are lost. In the case of stochastic rounding, when this is done over thousands of updates, the result is an effectively higher resolution than the physical level of quantization.

S.5 NON-BINARIZED SYNAPSE STREAM LEARNING PERFORMANCE



Supplementary Figure 4. Non-binarized DW-MTJ synapse stream learning. Validation accuracy of simulated network with continuous weights for stream Fashion-MNIST learning. Ticks at the top-right denote training using the full dataset for 30 epochs.

Supplementary Fig. 4 shows training performance of the stream learning task on a NN comprised of MW DW-MTJ synapses. The network architecture is the same as that of the binarized neural network: 784 input, 2 hidden layers of 512 units each, and 10 output units, all followed by batch normalization. The sign activations are replaced with ReLU activations. The weights are quantized stochastically to 32 levels. The linear synapse with symmetric updates performs the best of all three in both the stream learning and full data training case. Since the weight is continuous, linear and symmetric updates are essential for high accuracy training using backpropagation. The linear synapse with symmetric updates performs significantly better than the other two cases when shown the full data, but this advantage decreases greatly in the stream-learning case, most likely due to overfitting problems. The trapezoidal synapse can still overcome some overfitting problems as evidenced by the slightly better performance compared to the linear synapse with asymmetric updates. However, since the performance ceiling is relatively low, it still performs slightly worse than the linear synapse with symmetric updates.