# Current Mode Memristor Crossbars for Neuromorphic Computing

Neuro-Inspired Computational Elements Workshop, 2019

Cory Merkel, Ph.D.

Assistant Professor

Department of Computer Engineering

Rochester Institute of Technology

brain·lab

# Agenda

brain·lab

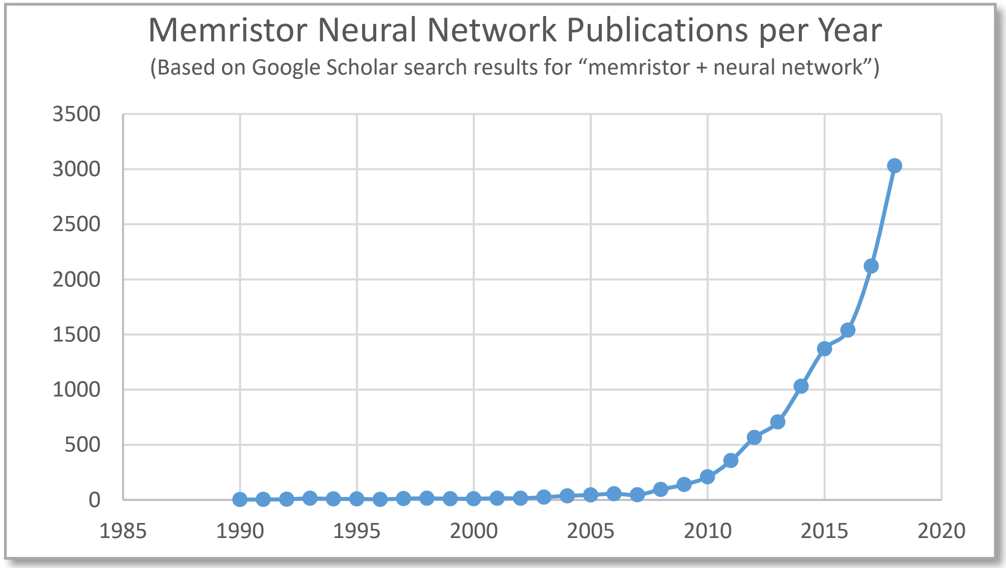
- Background and Motivation
- Current-Mode Crossbars
- MNIST Experiments
- Conclusions and Future Work

# Agenda

- Background and Motivation
- Current-Mode Crossbars
- MNIST Experiments
- Conclusions and Future Work

# Motivation

- Memristors → Avenue for low-power neuromorphic hardware




- Previous work proposes voltage-driven crossbars for neural network weight matrices[2,3]

[1]C. Merkel et al., “Neuromemristive systems: Boosing efficiency through brain-inspired computing,”IEEE Computer, October 2016.
[2]G. Indiveri et al., “Integration of nanoscale memristor synapses in neuromorphic computingarchitectures,”Nanotechnology, 2013.
[3]G. W. Burr et al., “Neuromorphic computing using non-volatilememory,”Advances in Physics: X, vol. 2, no. 1, pp. 89–124, 2017.

# Motivation

- Voltage-mode neurons driving memristor crossbars suffer from a few drawbacks:
  - Large memristor conductances → Neuron output degradation
  - IR drop over long-distance communication pathways
- Some solutions: 1T1R, spiking networks
- Current-mode designs offer
  - Better input/output swing, reduced loading issues
  - Low voltage operation and higher bandwidth
  - Nice design techniques (e.g. translinear principle)

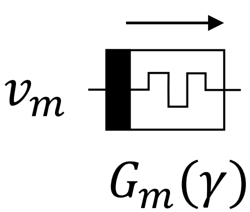
**Can we operate a memristor crossbar in current mode (current in, current out) for neuromorphic computing?**

# Overview of Memristor Device

brain·lab

- Memristor = "Memory + Resistor"
  - 2-Terminal thin-film device with state-dependent Ohm's Law (Chua, 2011)

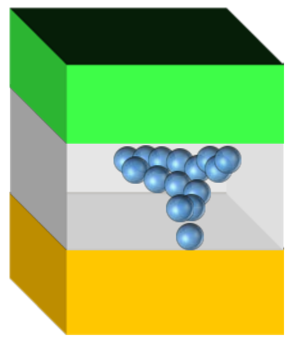
$$i_m = G_m(\gamma)v_m$$

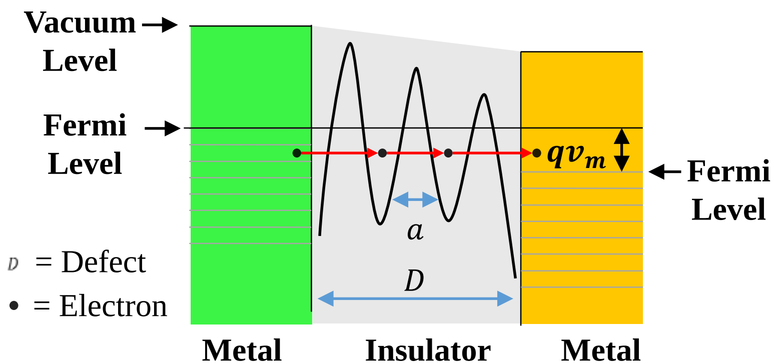



Top Electrode (Metal)

Switching Layer (Insulator)

Bottom Electrode (Metal)

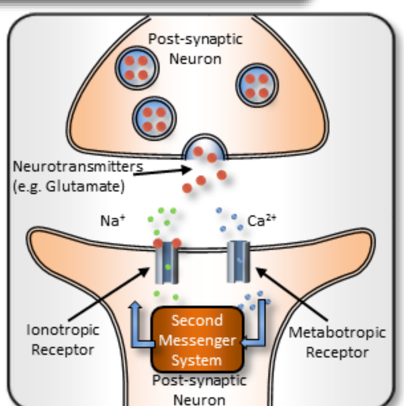





**Similar to Biological Synapses**

Facilitate:

1.) Computation
2.) Memory
3.) Learning




| | | Memristors | | | |
|---|---|---|---|---|---|
| Metric | Flash | PCM | STT-RAM | RRAM | Targets |
| Dynamic Range | - | >1000 | 2 | 1000 | >4 |
| Number of States | 8-16 | 100 | 4 | 100 | 20-100 |
| Retention | Several years at room temp. | | | | Years |
| Energy (pJ/bit) | >100 | 2-25 | 0.1-2.5 | 0.1-3 | 0.01 |
| Endurance (cycles) | $10^4$ | $10^9$ | $10^{15}$ | $10^{12}$ | $10^9$ |

Table compiled from (Yang et al., 2013; Kuzum et al., 2013; Devised et al., 2013; Ishigaki et al., 2

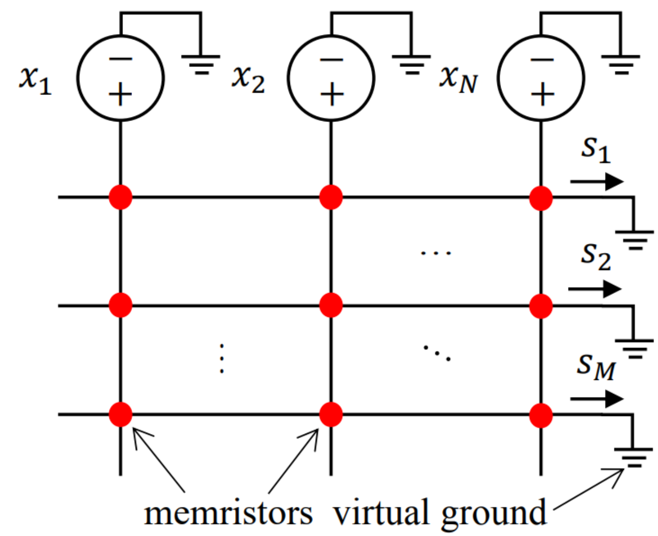
# Agenda

- Background and Motivation
- Current-Mode Crossbars
- MNIST Experiments
- Conclusions and Future Work

# Voltage-Mode Memristor Xbar

brain·lab

- Matrix-vector multiplication implemented by xbar as
$$\mathbf{s} = \mathbf{W}\mathbf{x}$$




- Each weight is defined by a single memristor conductance
$$w_{ij} = G_{ij}/G_{max}$$

- Weight range
$$g \le w_{ij} \le 1$$
where $g \equiv G_{min}/G_{max}$

# Current-Mode Memristor Xbar

brain·lab

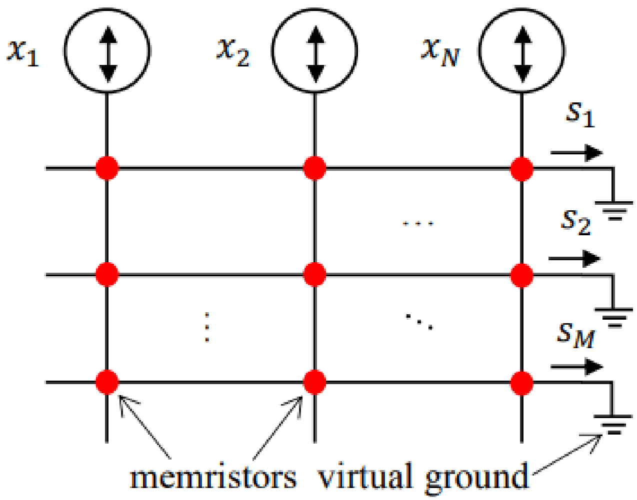
- The weight definition becomes:

$$w_{ij} = \frac{G_{ij}}{\sum_{k=1}^{M} G_{kj}}$$

- The maximum driving current should be

$$I_{max} < V_{th} \sum_{i} G_{ij}$$

to avoid read disturb

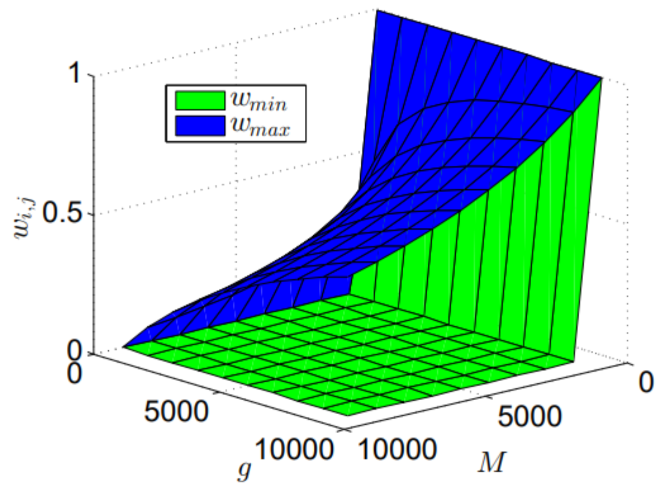


# Weight Range

brain·lab

- Weight range depends on the fanout:

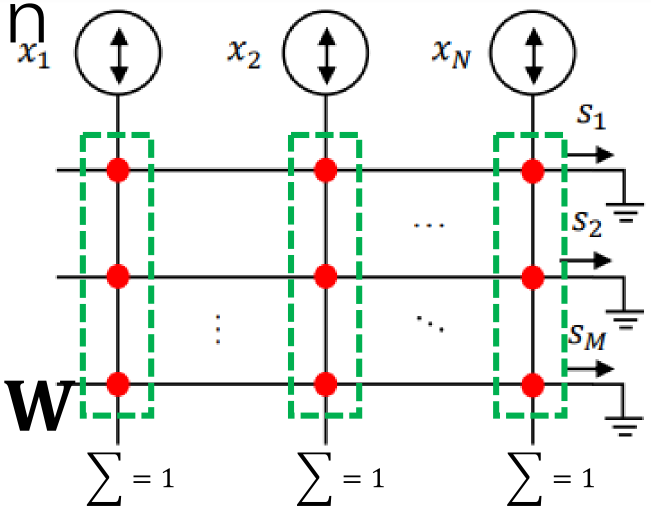
$$\frac{1}{(M-1)g+1} \leq w_{ij} \leq \frac{g}{M-1+g}$$




- Range approaches [0,1] for large $g$
- May need to split into multiple crossbars for very large $M$

# Weight Distribution

- Constraint: Weights in a column must sum to 1
- Given a target weight matrix $\overline{\mathbf{W}}$, find closest constrained matrix $\mathbf{W}$




$$\text{minimize} \quad \sum_{i=1}^{M}\sum_{j=1}^{N}\left(w_{ij} - \overline{w}_{ij}\right)^2$$

$$\text{s.t.} \quad \sum_{i} w_{ij} = 1 \;\forall\, j = 1,2,\dots,N$$

$$w_{min} \le w_{ij} \le w_{max}$$

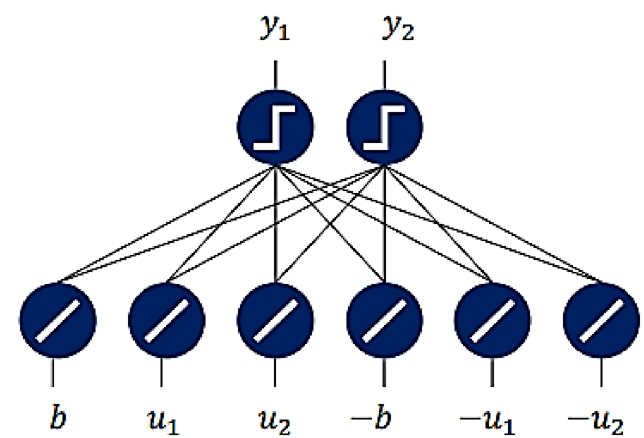
$$w_{ij} = \frac{1}{M}\left(1 - \sum_{k=1}^{M}\overline{w}_{kj}\right) + \overline{w}_{ij}$$

$$G_{ij} = \frac{w_{ij}}{w_{kj}} G_{kj}$$

# AND-OR Perceptron

brain·lab

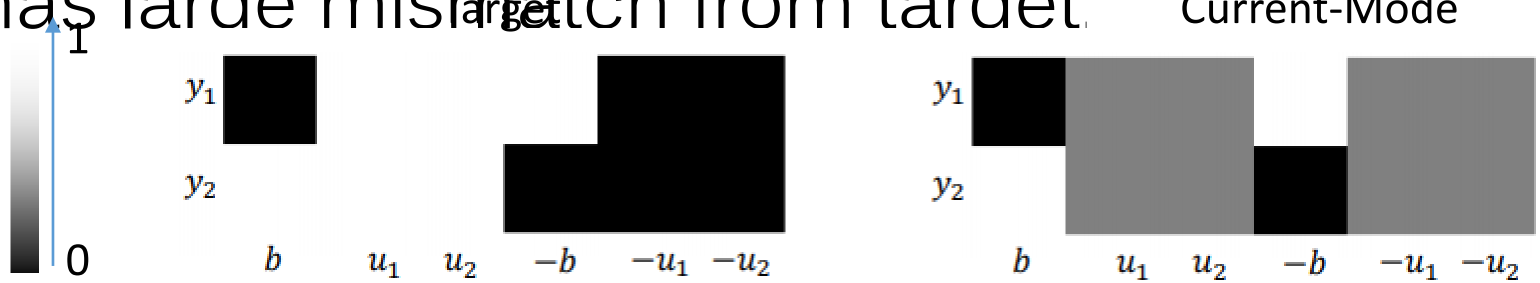
- Simple perceptron trained to perform AND and OR logic functions then mapped to current-mode xbar




- Accuracy:

| Voltage-Mode | Current-Mode |
|---|---|
| 100% ☺ | 50% ☹ |

- Current-mode weight matrix has large mismatch from target:



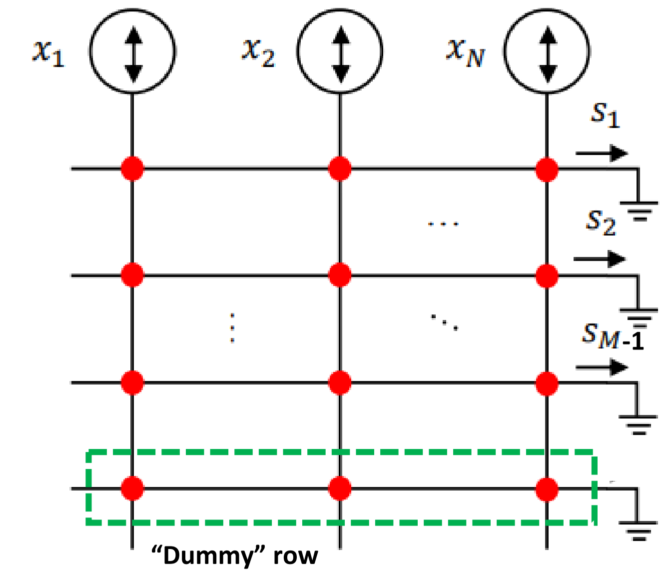

Can't simultaneously control relative and absolute weights in current-mode

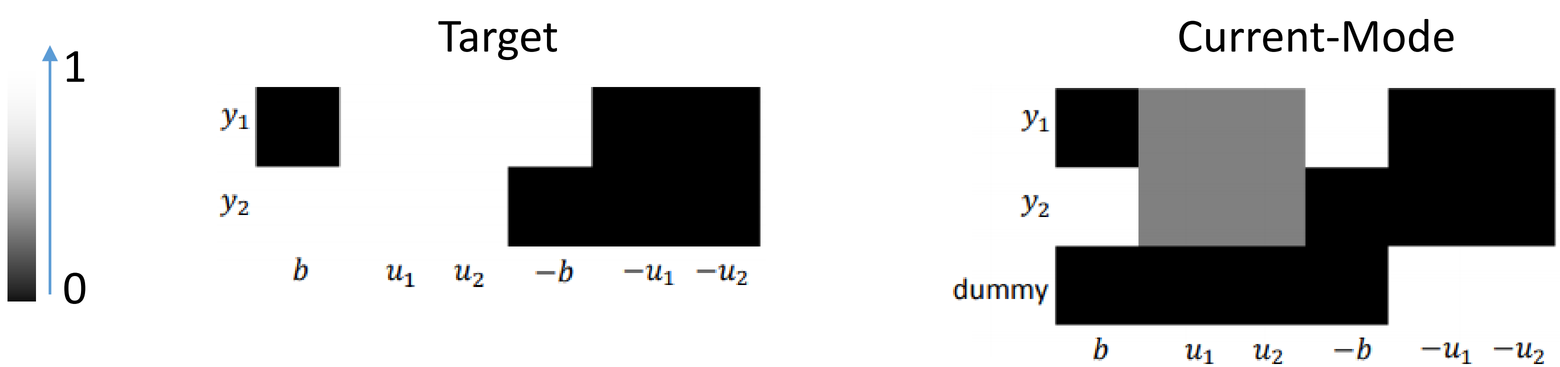
# Addition of Dummy Row

- Dummy row allows magnitude of entire column to be adjusted
- Now, set weights as:

$$w_{ij} = \overline{w}_{ij} \;\forall\, i < M$$

$$w_{Mj} = 1 - \sum_{k=1}^{M} \overline{w}_{kj}$$



# Restrict Target Weights

brain·lab

- Finally, to make sure a target matrix can be mapped to the current-mode crossbar:

$$\max\left(\frac{1-w_{max}}{M-1}, w_{min}\right) \leq \overline{w}_{ij} \leq \min\left(\frac{1-w_{min}}{M-1}, w_{max}\right)$$

- The target matrix constraint can be satisfied using regularization techniques

With these constraints, any voltage-mode crossbar maps uniquely to an equivalent current-mode crossbar.

# Bipolar Weights

brain·lab

- We can create bipolar (positive and negative) weights **without the need for two memristors**
- We know the total current from the pre-synaptic neurons
  - Just subtract a fraction of it to get positive and negative weights

$$s_i = \sum_{j=1}^{N} \frac{G_{ij}}{\sum_{k=1}^{N} G_{kj}} x_j - \theta \sum_{j=1}^{N} x_j = w_{ij}^{*} x_j$$

where $w_{ij}^{*} = w_{ij} - \theta$

- Good choice for $\theta$ is usually $\frac{\overline{w}_{max} + \overline{w}_{min}}{2}$

# Gradient Descent

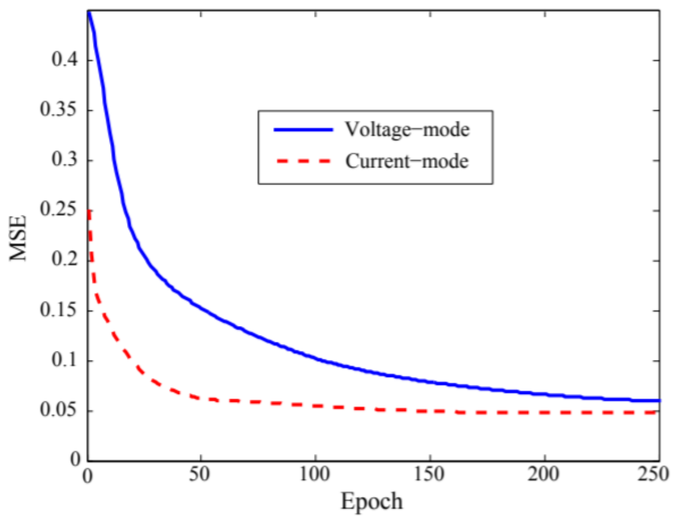
brain·lab

- We can train offline using the proposed method, but online training is more desirable
- Gradient descent can be implemented as

$$\Delta G_{ij}^{l} = -\alpha \frac{\partial J}{\partial G_{ij}^{l}} = \alpha \sum_{k=1}^{M^l} \delta_k^l x_j^{l-1} \frac{\partial w_{kj}^{l}}{\partial G_{ij}^{l}}$$

$$\approx \begin{cases} \alpha \delta_i^l x_j^l & i \neq M \\ -\langle \Delta G_{1:M-1,j}^{l} \rangle & i = M \end{cases}$$

**AND/OR Perceptron**




**Just use delta rule for non-dummy rows and average the updates for dummy row**

# Agenda

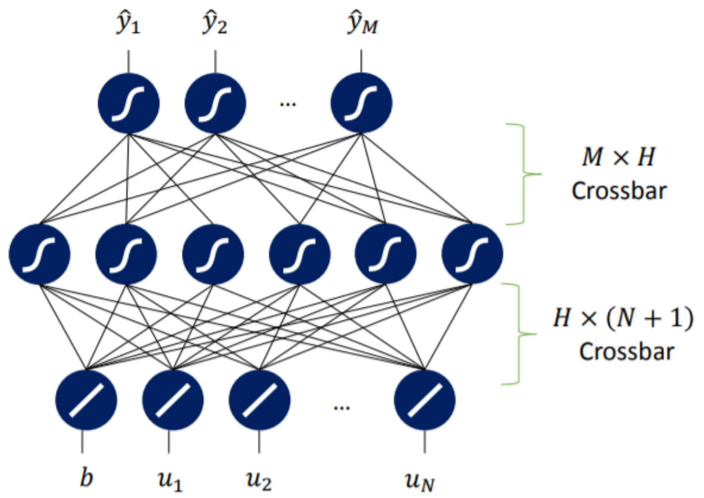
brain·lab

- Background and Motivation
- Current-Mode Crossbars
- MNIST Experiments
- Conclusions and Future Work

# MNIST Experiments

brain·lab

- Circuits modeled in MATLAB based on Ag chalcogenide devices[1]:
  - $G_{min} = 2.1 \times 10^{-5}$ S
  - $G_{max} = 1 \times 10^{-3}$ S
- MLP with 49 inputs, 50 hidden units, and 10 outputs

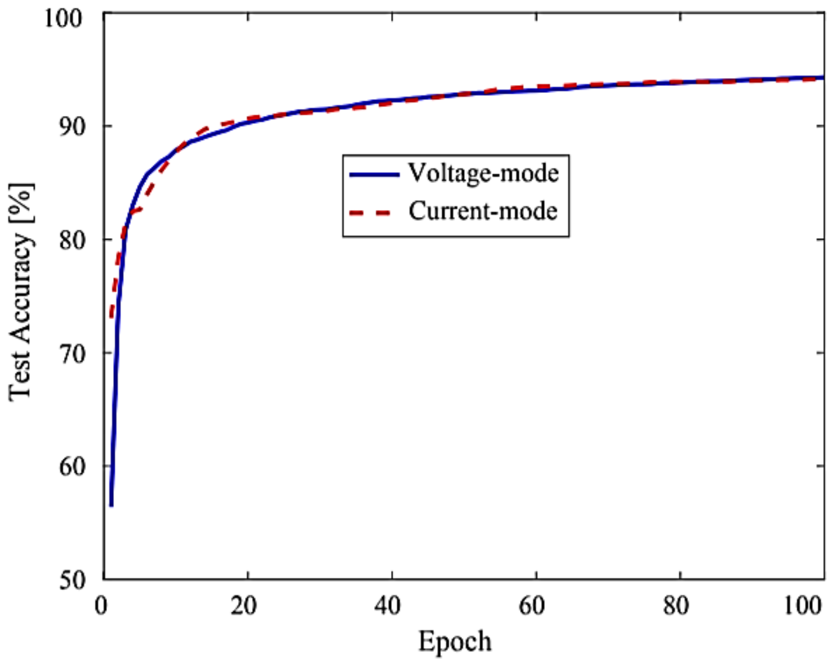



- Classification on MNIST handwritten digits (60000 train, 10000 test)

[1]A. S. Oblea et al., "Silver Chalcogenide Based Memristor Devices," Proceedings of the IEEE , vol. 3, 2010.

# Classification Accuracy

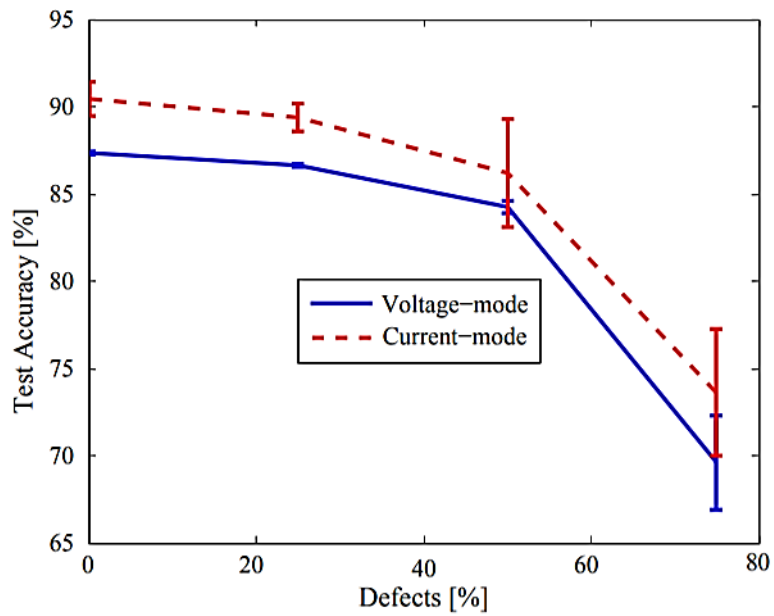
- Trained using online backpropagation
  - Delta rule for voltage-mode
  - Derived rule for current-mode




- Similar convergence and final accuracy values

# Defect Tolerance

- Defects in the current-mode crossbar (e.g. stuck-at faults) will affect the weight range of each row in the colum




- Results indicate that current-mode design may be as robust as voltage-mode
  - More results are needed to verify

# Agenda

brain·lab

- Background and Motivation
- Current-Mode Crossbars
- MNIST Experiments
- Conclusions and Future Work

# Conclusions and Future Work

brain·lab

- Conclusions:
  - Current-mode memristor crossbars may be a viable alternative to voltage-mode designs
    - Similar training convergence and defect tolerance
  - Some considerations
    - Reduced weight range
    - Need for a dummy row
- Future work
  - SPICE simulations for power consumption and performance analysis
  - Deeper analysis of weight distributions and feature extraction

# Thank You!

brain·lab