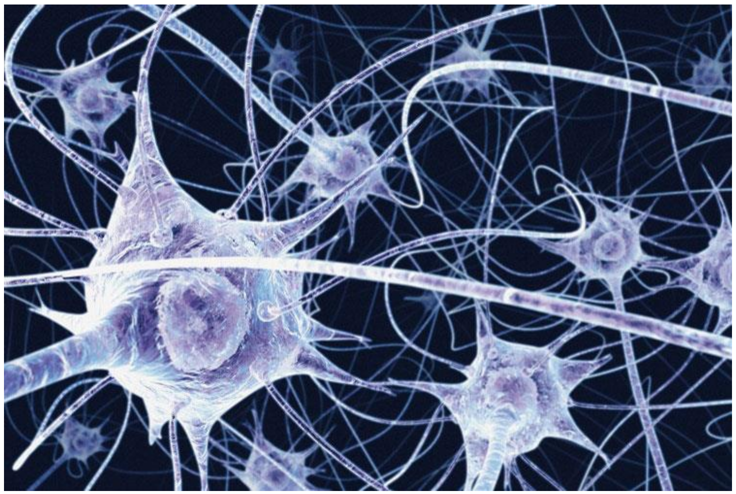
Feature Extraction and Image analysis using memristor networks

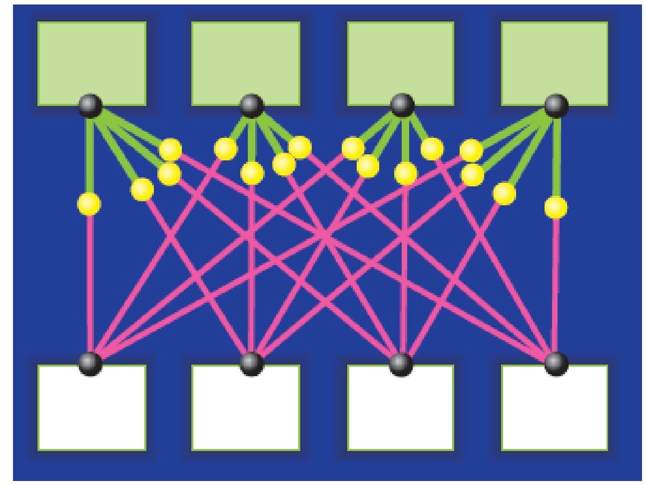
Wei Lu

University of Michigan
Electrical Engineering and Computer Science

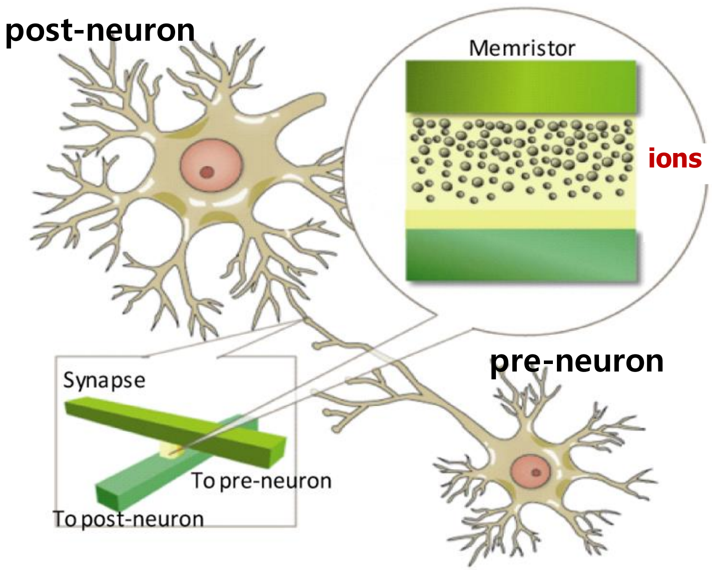
Memristor Based Neural Network Hardware

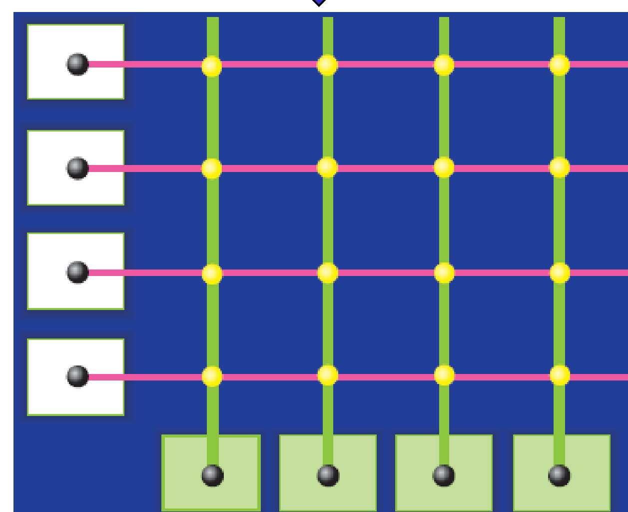
Synapse – reconfigurable two-terminal resistive switches





Goal: building bio-inspired, efficient artificial neural networks





U. Michigan

S. H. Jo, T. Chang, I. Ebong, B. Bhavitavya, P. Mazumder, W. Lu, Nano Lett. **10, 1297 (2010).**

Computing with Memristor Arrays

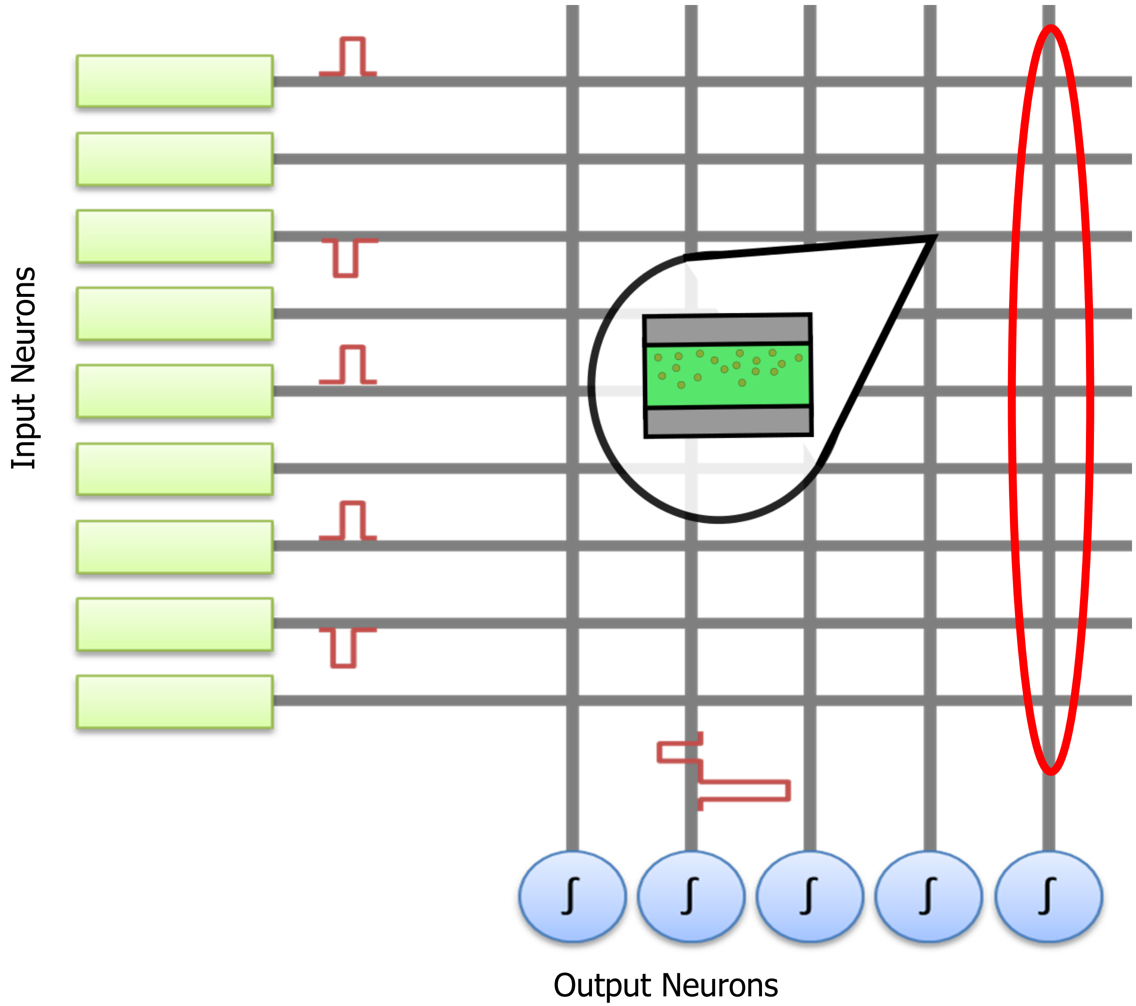
Memristors perform learning and inference functions

DARPA UPSIDE program

- **Memristor weights form dictionary elements (features)**
- **Image input, Pixel intensity represented by widths of pulses**
- **Memristor array natively performs matrix operation**

$$\vec{I} = \vec{v} \cdot \overleftrightarrow{\Phi}$$

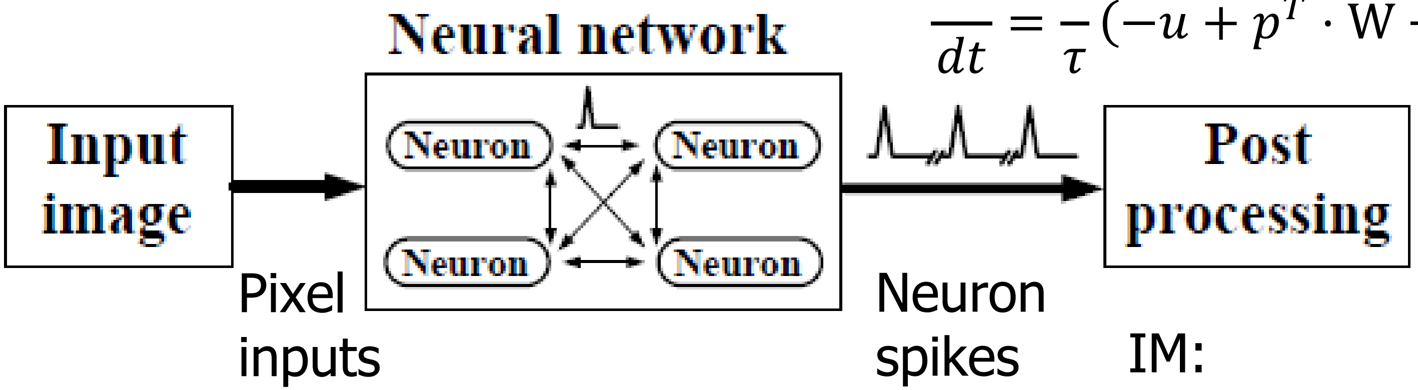
- **Integrate and fire neurons**
- **Learning achieved by backpropagating spikes**



Lu Group
U. Michigan

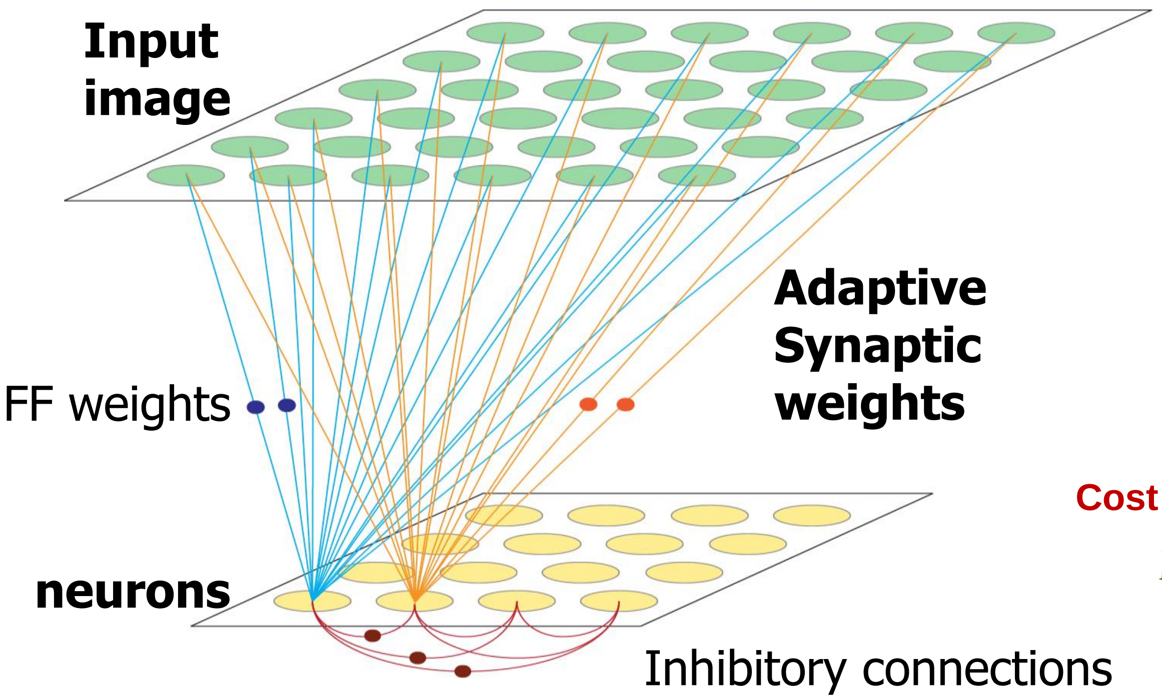
Neural Network for Image Processing based on Sparse Coding

$$\frac{du}{dt} = \frac{1}{\tau}\left(-u + p^T \cdot \mathrm{W} - a \cdot (\mathrm{W}^\top \mathrm{W} - I)\right)$$



IM:

1. Network adapt during training following local plasticity rules
2. FF weights form neuron receptive fields (dictionary elements)
3. Output as neuron firing rates



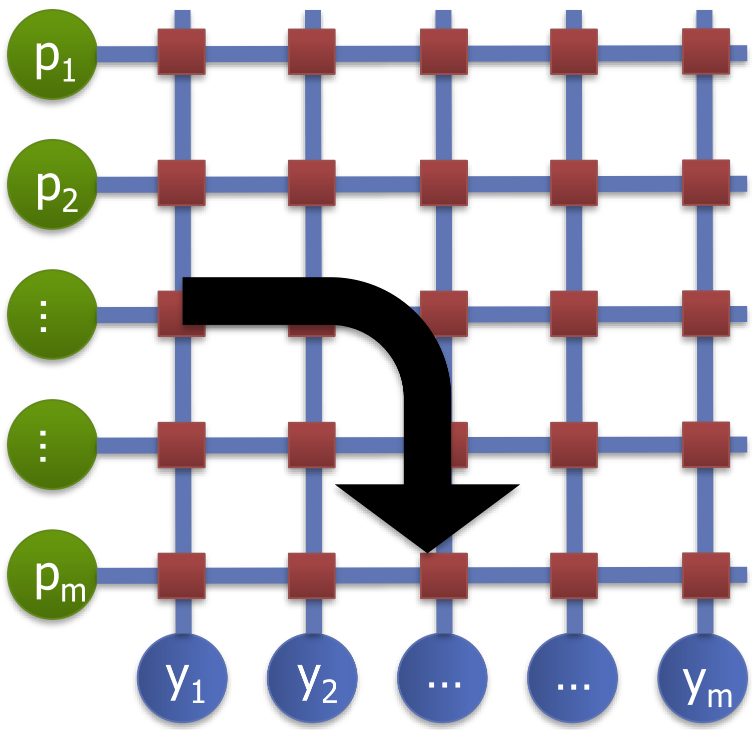
Cost Function:

$$E(t) = \frac{1}{2}\left|\left|s(t) - \hat{s}(t)\right|\right|^2 + \lambda \sum_m C(a_m(t)).$$

Sparse Coding Implementation in Memristor Array

Forward Pass

Update neurons/activities



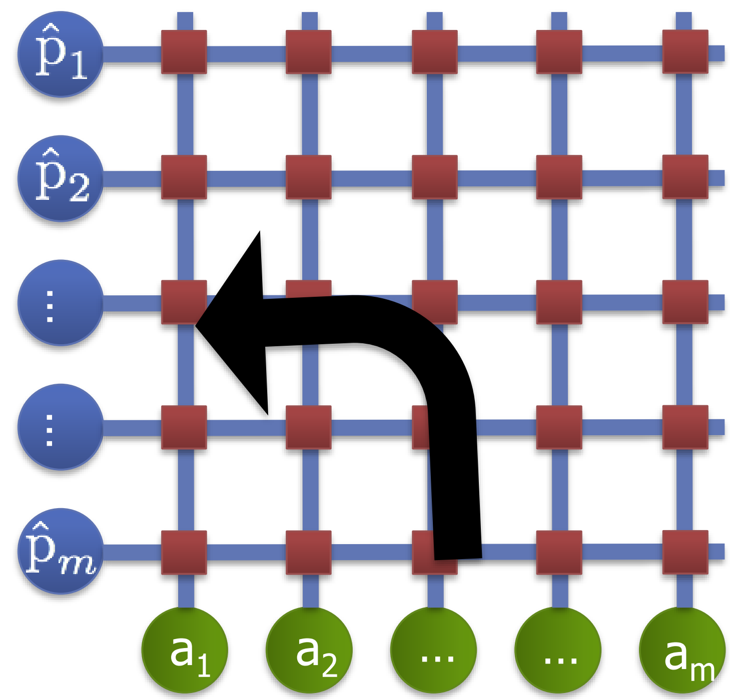
$$y = p^{\intercal} W$$

$$\frac{du}{dt} = \frac{1}{\tau}(-u + p^{T} \cdot \mathrm{W} - a \cdot (\mathrm{W}^{\intercal}\mathrm{W} - I)\,)$$

$$\frac{du}{dt} = \frac{1}{\tau}\,(-u + (p - \hat{p})^{\intercal} W + a)$$

Backward pass

Update residual

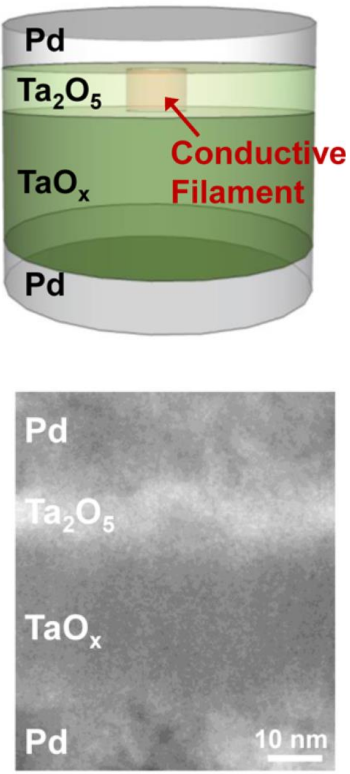


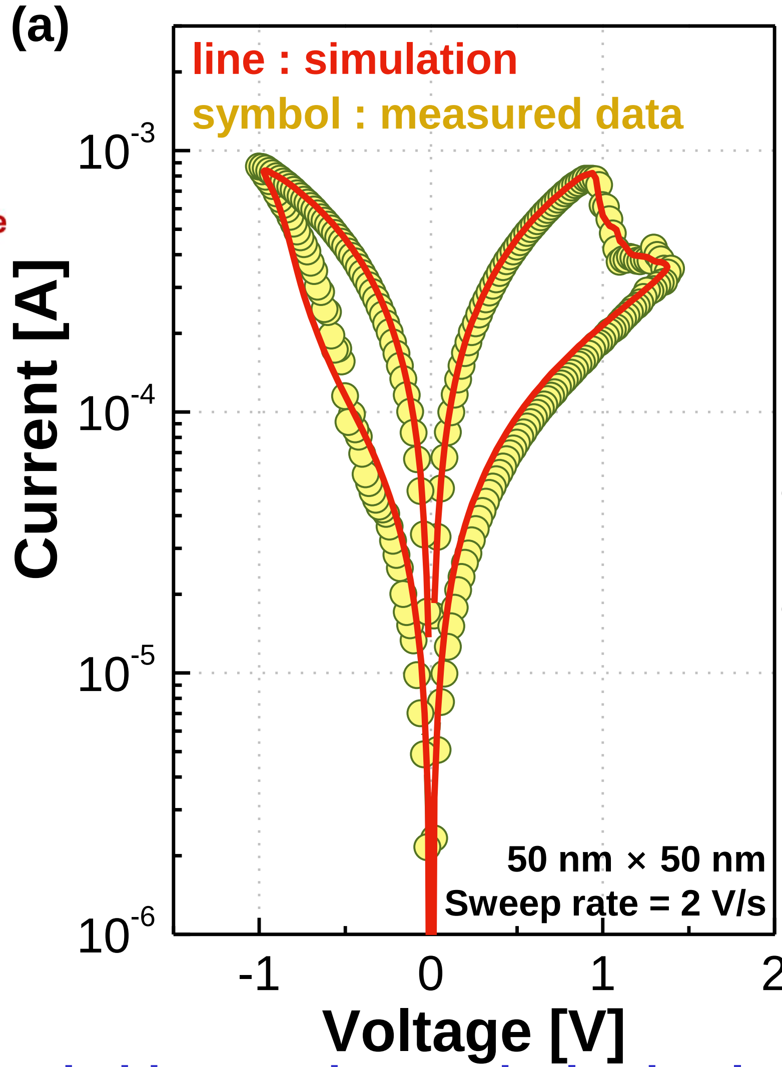
$$\hat{p} = a W^{\intercal}$$

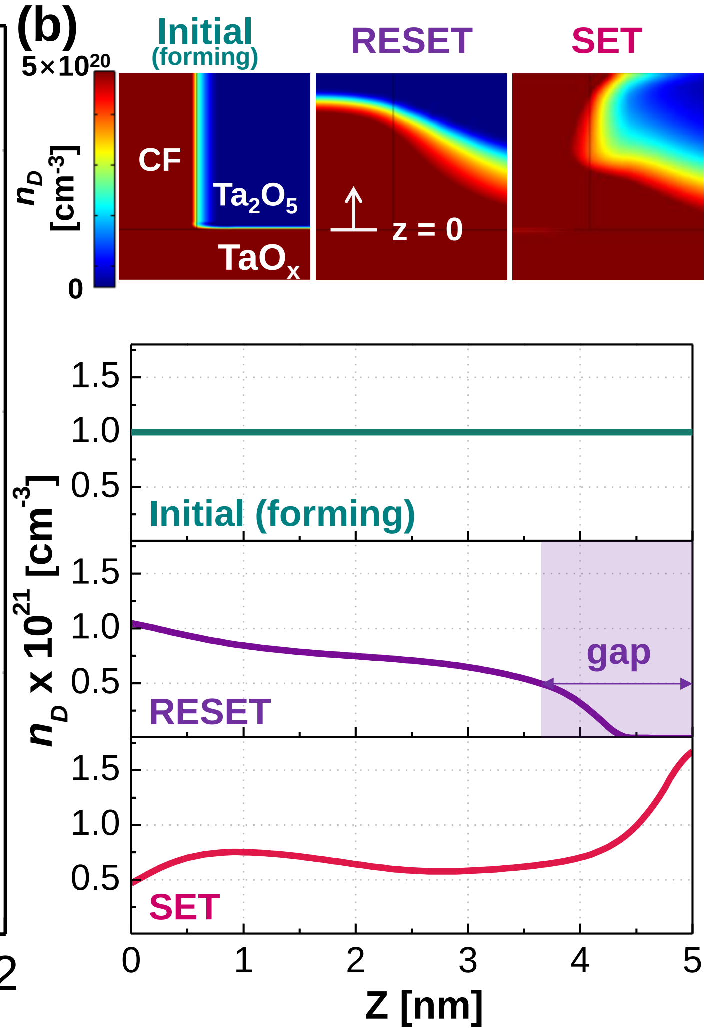
Neuron membrane potential

Sheridan et al., *Nature Nanotechnology*, 12, 784–789 (2017)

Analog Oxide Memristors







- Resistive switching can be precisely simulated after considering V_O diffusion, drift and thermophoresis effects

S. Kim, S. Choi, W. Lu, ACS Nano, 8, 2369–2376 (2014).

Simulation of Switching Process

• **Dependent variables**		• **Constants**	
n_D	Concentration of V_o [cm^{-3}]	a	Hopping distance, 0.1 nm
T	Temperature [K]	f	Escape-attempt frequency, 10^{12} Hz
ψ	Potential [V]	E_a	Diffusion barrier, 0.85 eV

• **Oxygen vacancy transport**

Eq.(1) $\frac{\partial n_D}{\partial t} = \nabla \cdot (D \nabla n_D - v n_D + D S n_D \nabla T)$

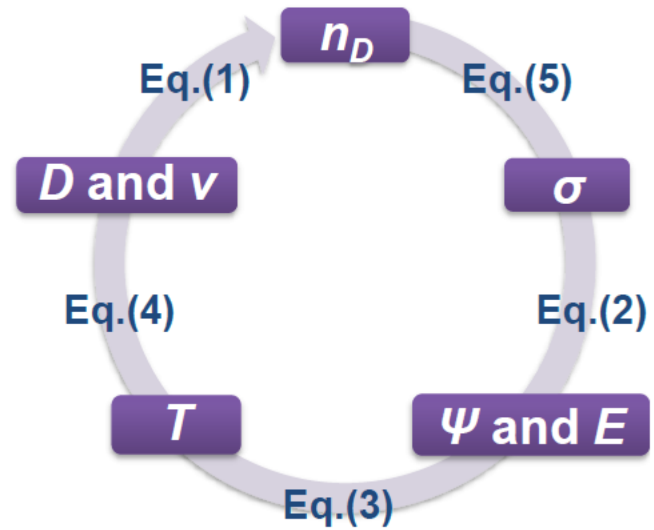
• **Current continuity**

Eq.(2) $\nabla \cdot \sigma \nabla \psi = 0$

• **Heat (Joule heating)**

Eq.(3) $-\nabla \cdot k_{th} \nabla T = J \cdot E = \gamma \cdot \sigma |\nabla \psi|^2$

(γ = 1 for DC, and γ = 2 for AC simulation)



• **Parameters** - Eqs.(4)

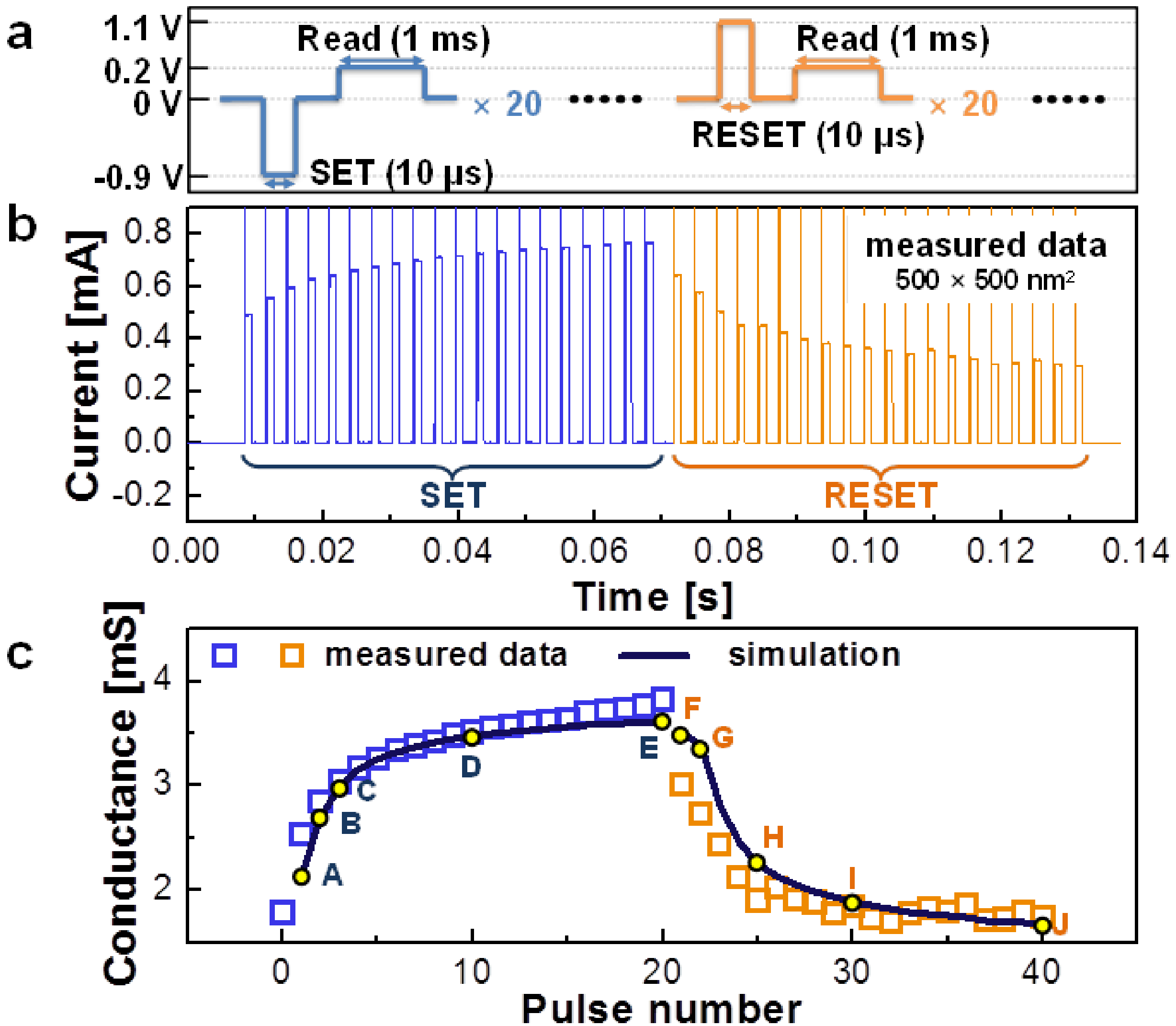
$D = 1/2 \cdot a^2 \cdot f \cdot \exp(-E_a / kT)$ Diffusivity of V_o [cm^2s^{-1}]

$v = a \cdot f \cdot \exp(-E_a / kT) \cdot \sinh(qaE / kT)$ Drift velocity of V_o [cm/s]

$S = -E_a / kT^2$ Soret diffusion coefficient [1/K]

S. Kim, S. Choi, W. Lu, ACS Nano, 8, 2369–2376 (2014).

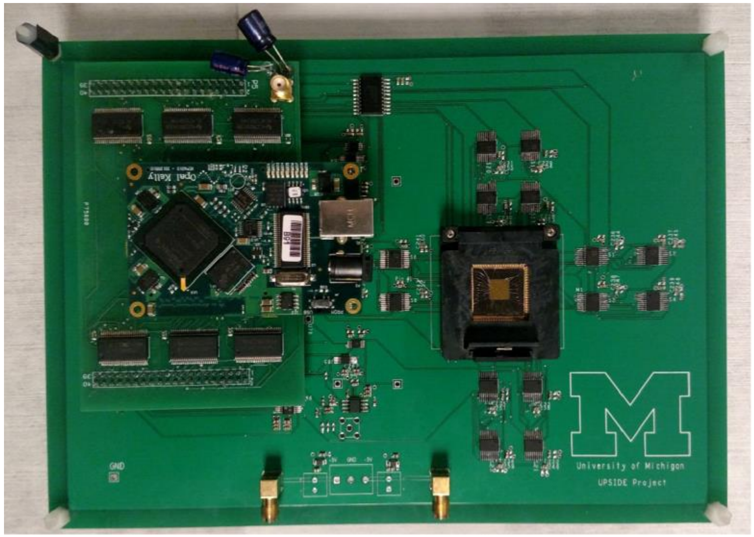
Simulation of Filament Growth

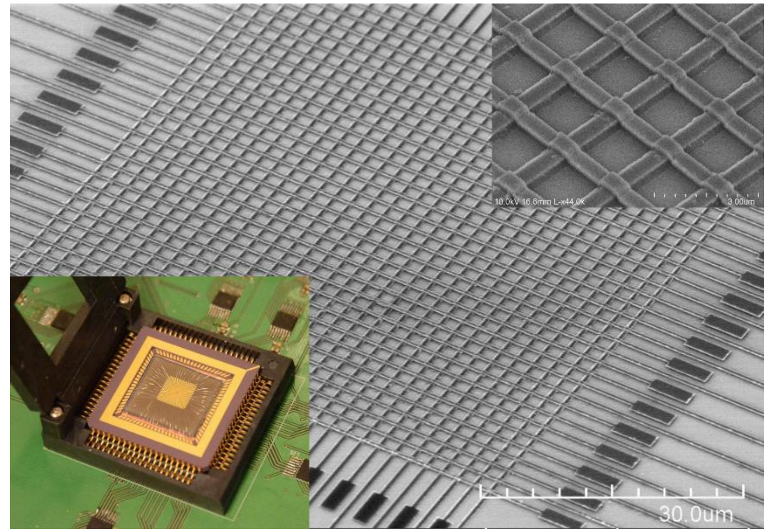


- Same set of parameters can explain both DC and pulse response

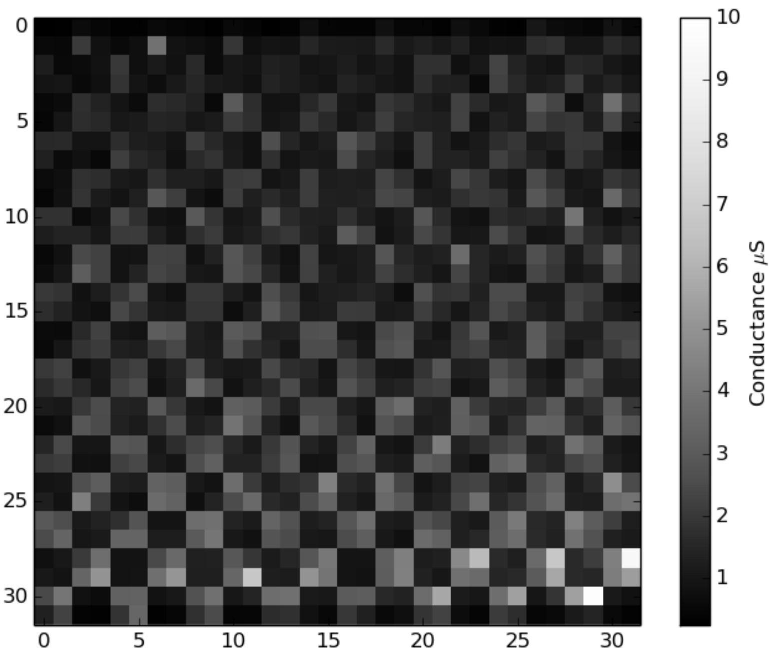
S. Kim, S. Choi, W. Lu, ACS Nano , 8, 2369–2376 (2014)

Neuromorphic Hardware Implementation





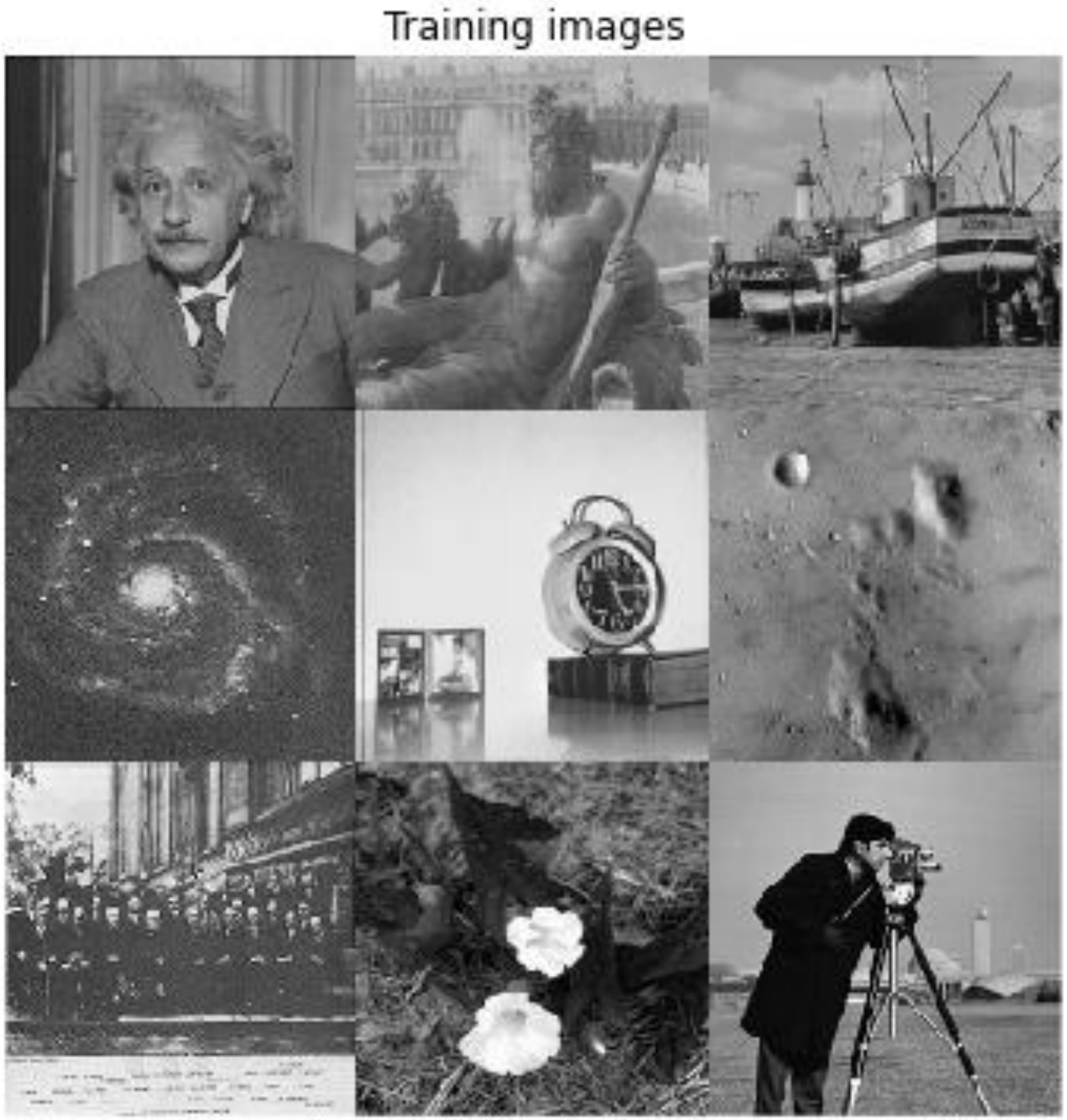
32x32 memristor array



- Checkerboard pattern
- 32 x 32 array
- Direct storage and read out
- No read-verify or re-programming

Sheridan et al., *Nature Nanotechnology*, 12, 784–789 (2017)

Training



9 Training Images

128x128px

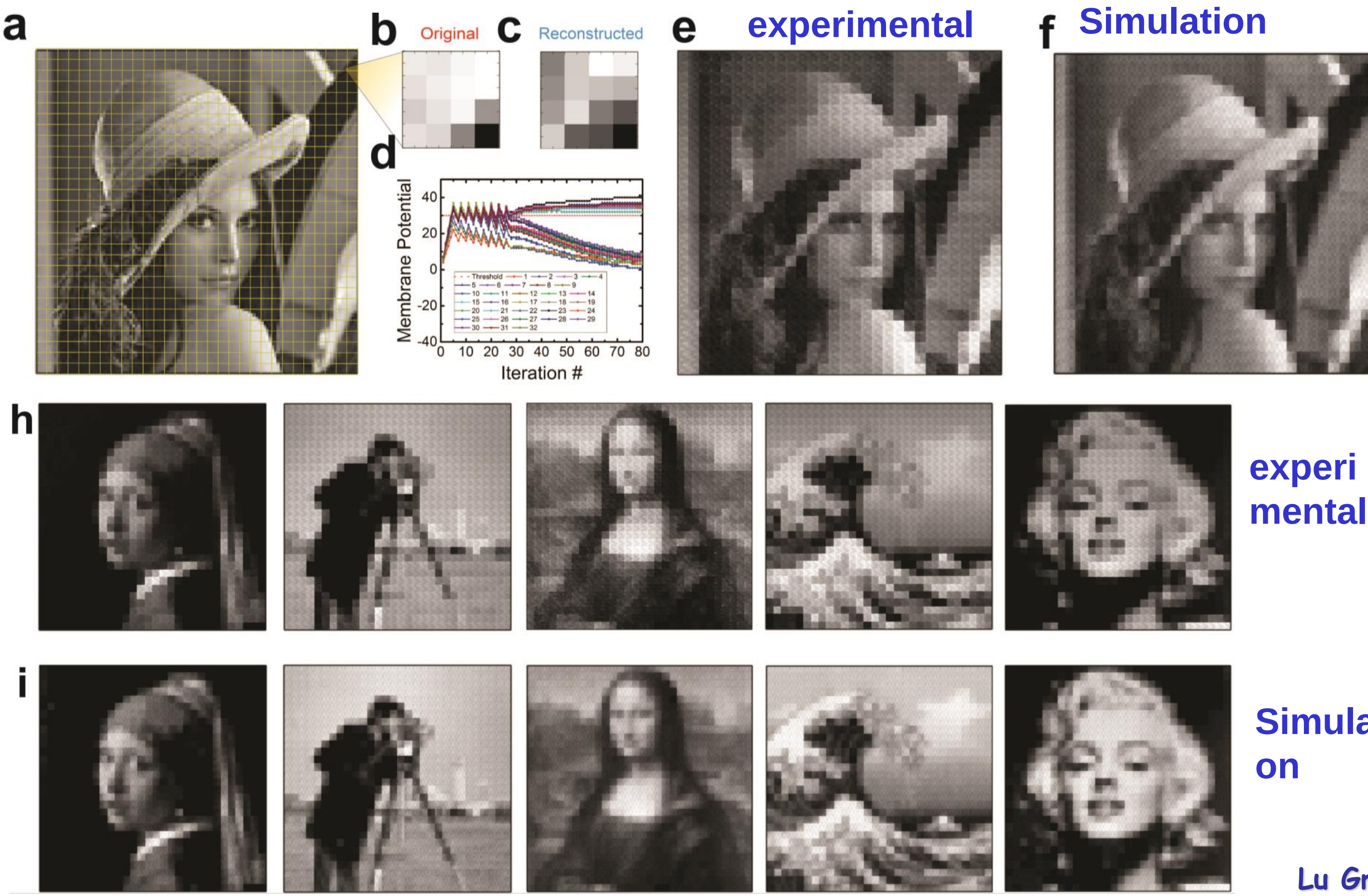
4x4 patches

127449 training patches (overlaps allowed)

Trained in random order

Sheridan et al., *Nature Nanotechnology*, 12, 784–789 (2017)

Image Reconstruction with Memristor Crossbar



Sheridan et al., *Nature Nanotechnology*, 12, 784–789 (2017)

Lu Group
U. Michigan

PCA Analysis Using Memristor Arrays

Wisconsin Breast Cancer Data

Sensory data from malignant or benign cells

$$\begin{pmatrix} clump\ thickness \\ uniformity\ of\ cell\ size \\ uniformity\ of\ cell\ shape \\ marginal\ adhesion \\ single\ epithelial\ cell\ size \\ bare\ nuclei \\ bland\ chromatin \\ normal\ nucleoli \\ mitoses \end{pmatrix} \quad \begin{pmatrix} 9 \\ 5 \\ 8 \\ 1 \\ 2 \\ 10 \\ 8 \\ 9 \\ 8 \end{pmatrix}, \begin{pmatrix} 10 \\ 10 \\ 6 \\ 3 \\ 3 \\ 10 \\ 3 \\ 5 \\ 3 \end{pmatrix}$$

- Principal Component Analysis (PCA) for data clustering
- Unsupervised training using Sanger's rule

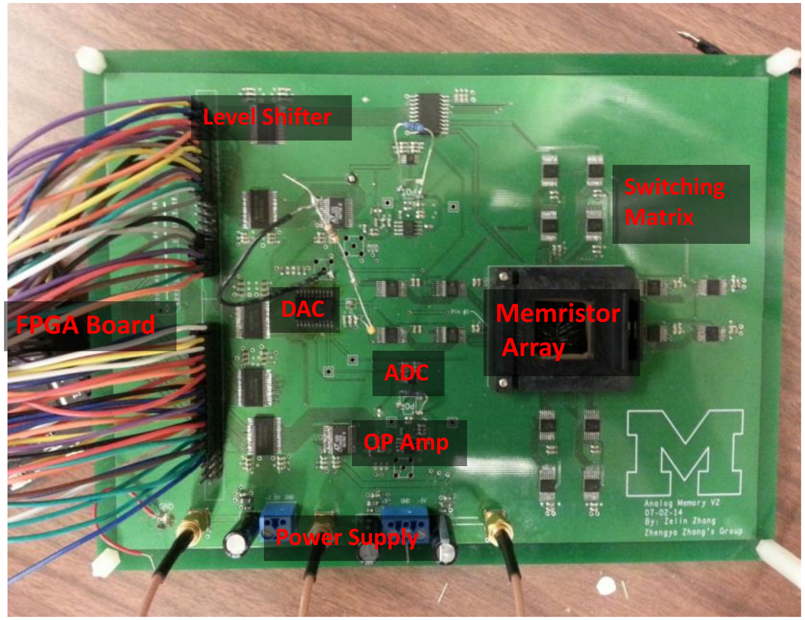
$$\Delta g_{ij} = \eta y_j (x_i - \sum_{k=1}^{j} g_{ik} y_k)$$

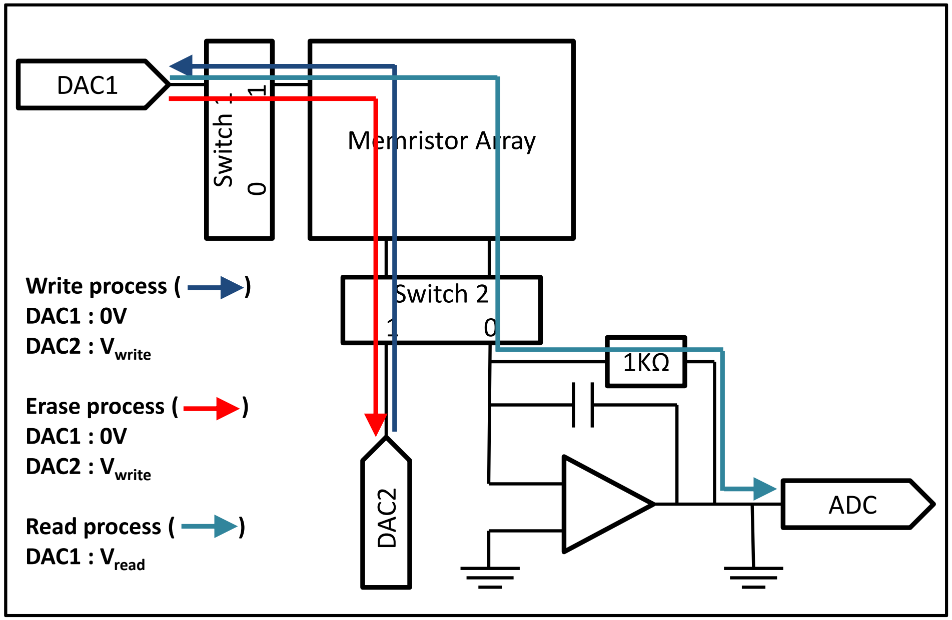
Training set: 100 points
Testing set: 583 points
Input voltage pulse:
Amplitude: fixed
Width: $\propto$ the values from the data

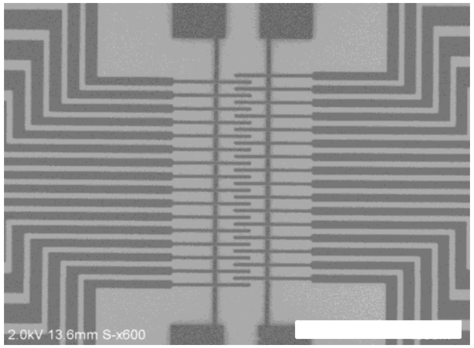
PCA network $\longrightarrow$ $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

S. Choi, P. Sheridan, J. Shin, W. Lu, Nano Lett. 2017, 17, 3113–3118

Experimental Implementation







- **9x2 memristor array**
- **Unsupervised learning using Sanger's rule**

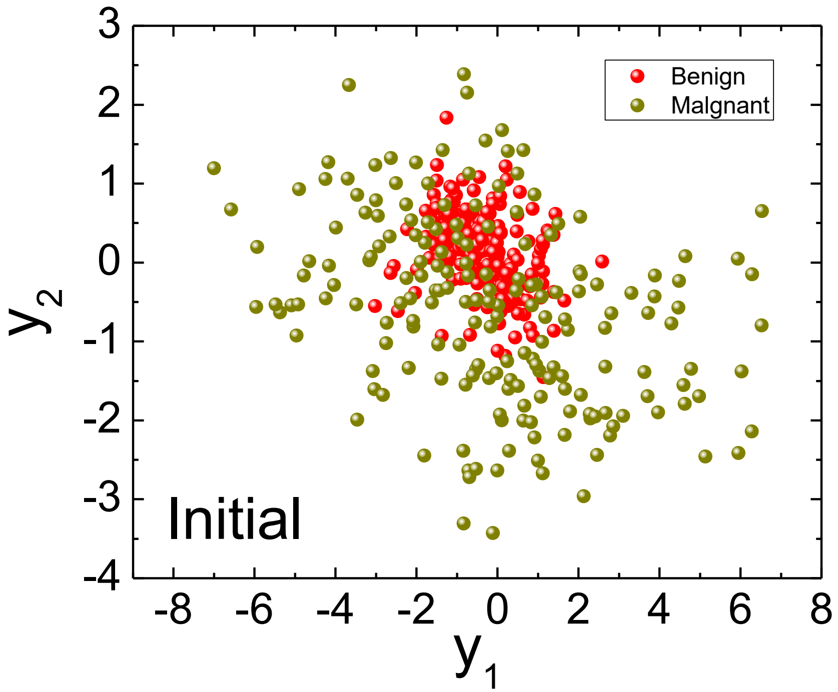
$$\Delta g_{ij} = \eta y_j (x_i - \sum_{k=1}^{j} g_{ik} y_k)$$

S. Choi, P. Sheridan, J. Shin, W. Lu, Nano Lett. 2017, 17, 3113–3118

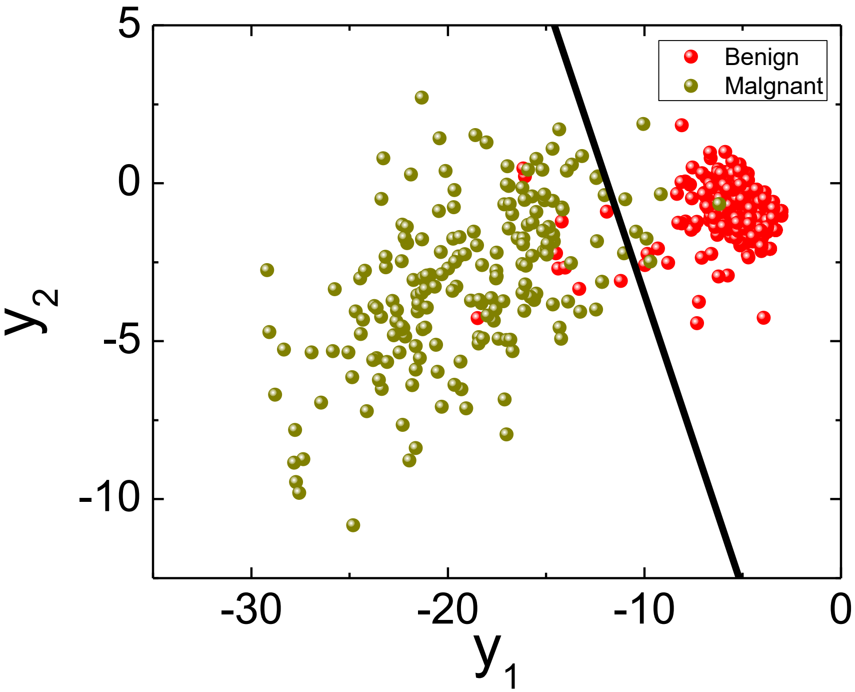
Experimental Implementation

Wisconsin Breast Cancer Data

Before training



After 100 cycles of training, experimental results

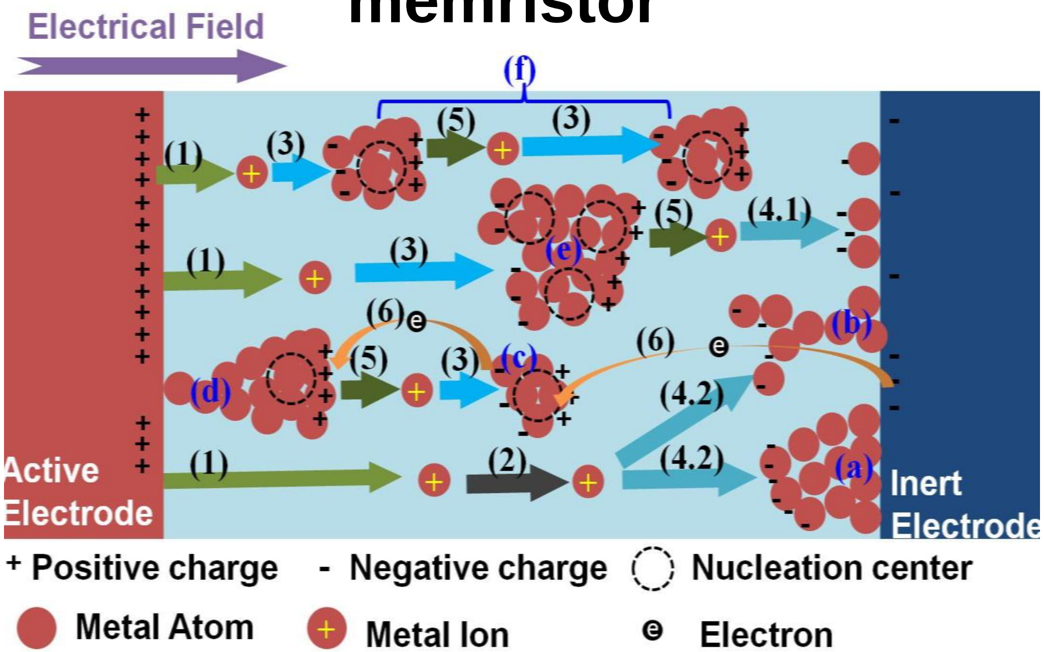


- Successful clustering obtained after unsupervised learning (without knowledge of the labels)
- Decision boundary drawn in a 2nd-step, supervised training process
- Classification accuracy ~ 97%, same as ideal software simulation

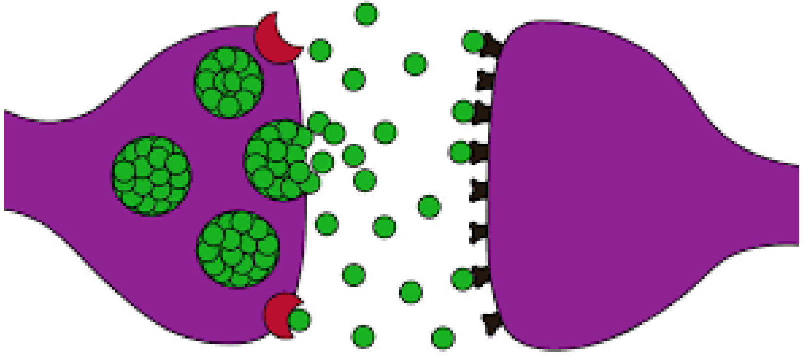
S. Choi, P. Sheridan, J. Shin, W. Lu, Nano Lett. 2017, 17, 3113–3118

Internal Dynamics at Different Time Scales

memristor



synapse



- Memristor offers interesting internal dynamics at different time scales, and can emulate synapse realistically

Microscopic physical processes during SET

(1) Ionization of metal atoms in AE (anodic dissolution)

(2) Metal ions hopping in dielectrics

(3) Metal ions attachment to existing clusters

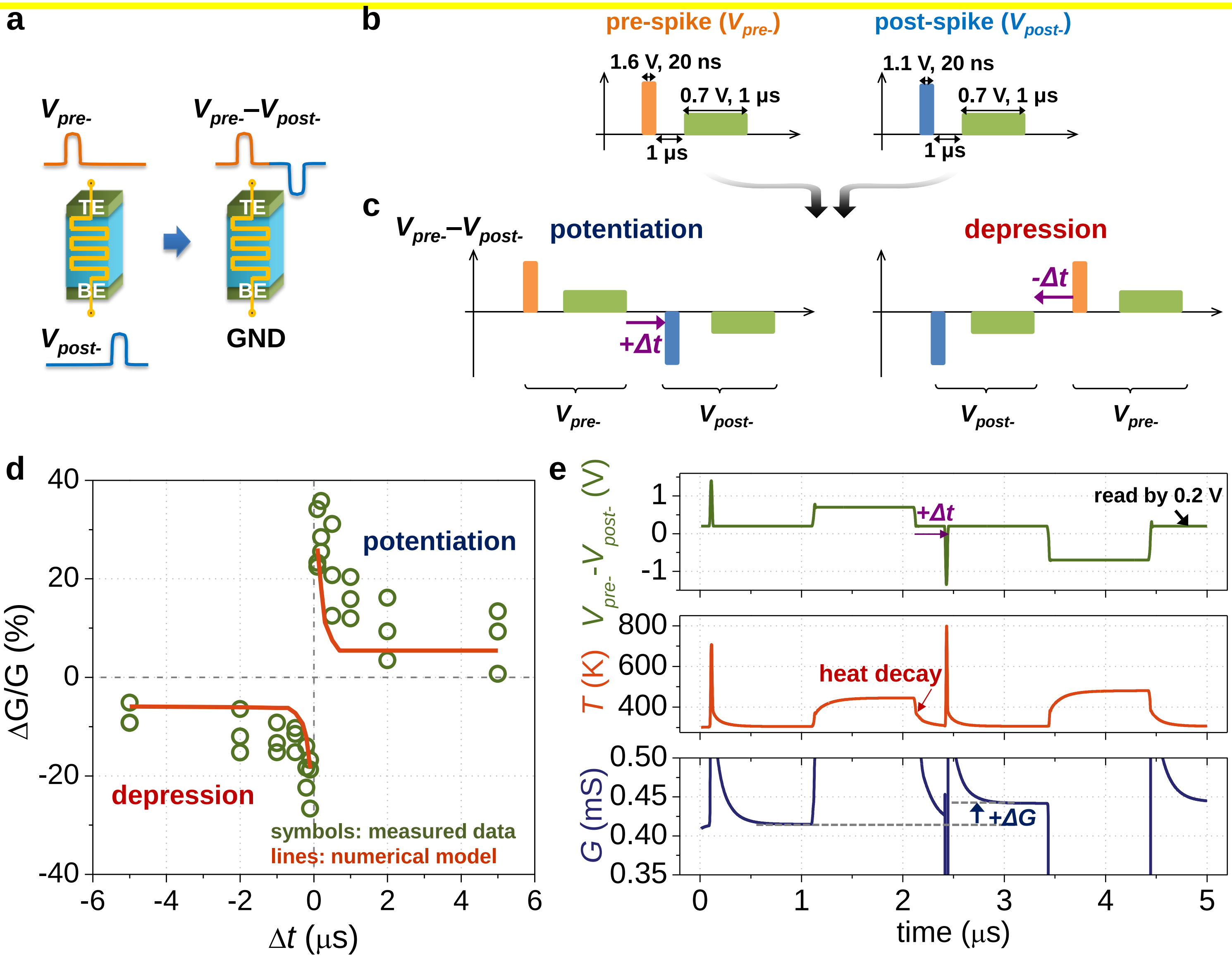
(4) Nucleation of metal ions captured by (4.1) IE and (4.2)

(5) Metal atoms in nuclei are activated to ions

(6) Electron hopping from IE to Neutralize positive charge from metal ions

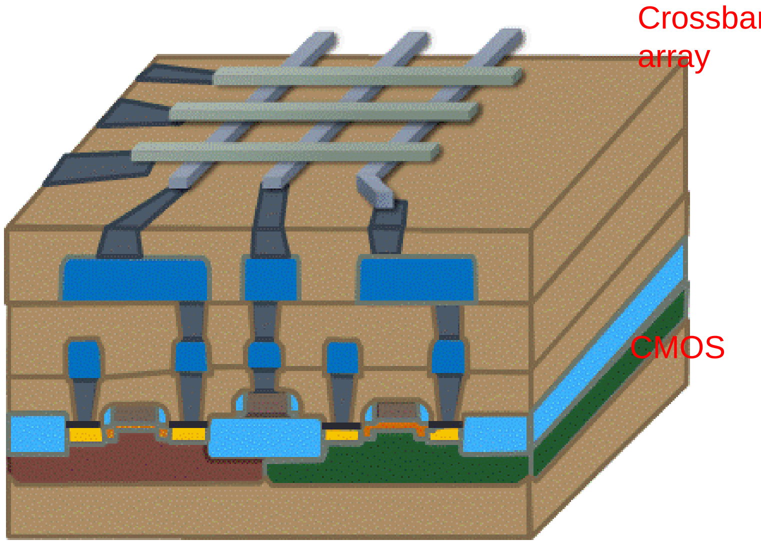
C. Du, W. Ma, T. Chang, P. Sheridan, W. D. Lu, Adv. Func. Mater., 25, 4290, (2015)
S. Kim, C. Du, P. Sheridan, W. Ma, S. Choi, W.D. Lu, Nano Lett, 15, 2203 (2015).

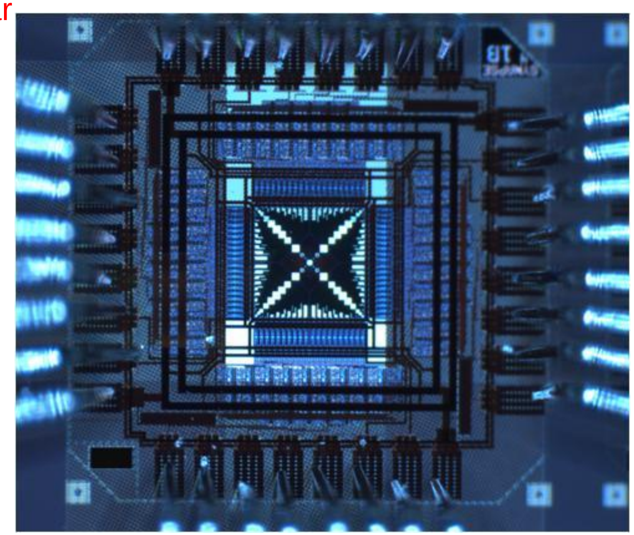
Implementing STDP (and Spiking Rate Dependent Plasticity) Naturally



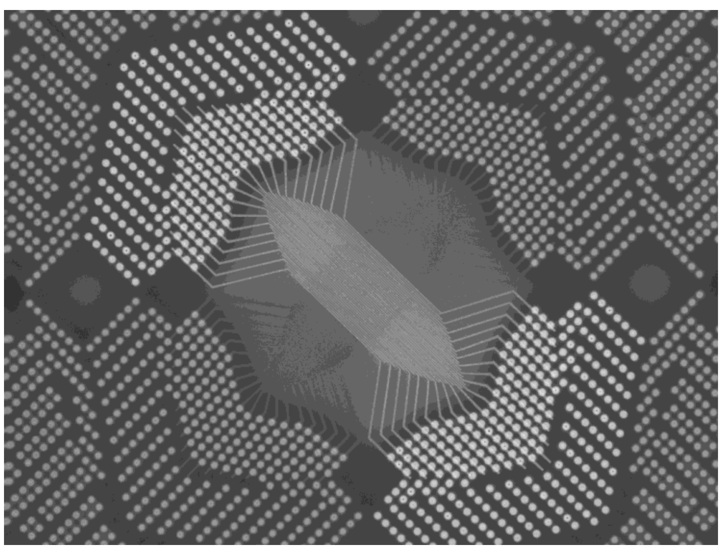
S. Kim, C. Du, P. Sheridan, W. Ma, S. Choi, W.D. Lu, Nano Lett, 15, 2203 (2015).

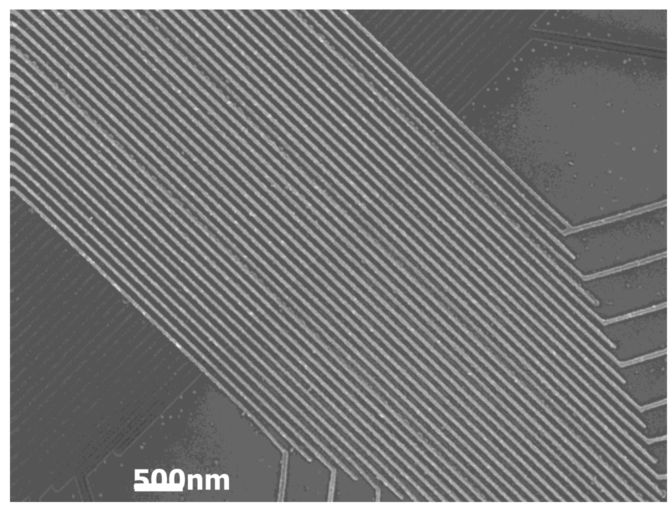
Integrated Crossbar Array/CMOS System





- Low-temperature process, RRAM array fabricated on top of CMOS
- CMOS provides address mux/demux
- RRAM array: 100nm pitch, 50nm linewidth with density of 10Gbits/cm^2
- CMOS units – larger but fewer units needed. 2n CMOS cells control n^2 memory cells



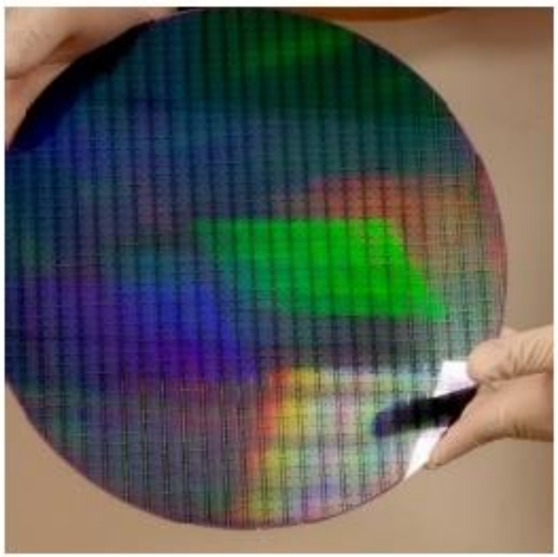


Kim, Gaba, Wheeler, Cruz-Albrecht, Srivinara, W. Lu *Nano Lett.*, 12, 389–395 (2012).

Lu Group
U. Michigan

Towards Commercialization

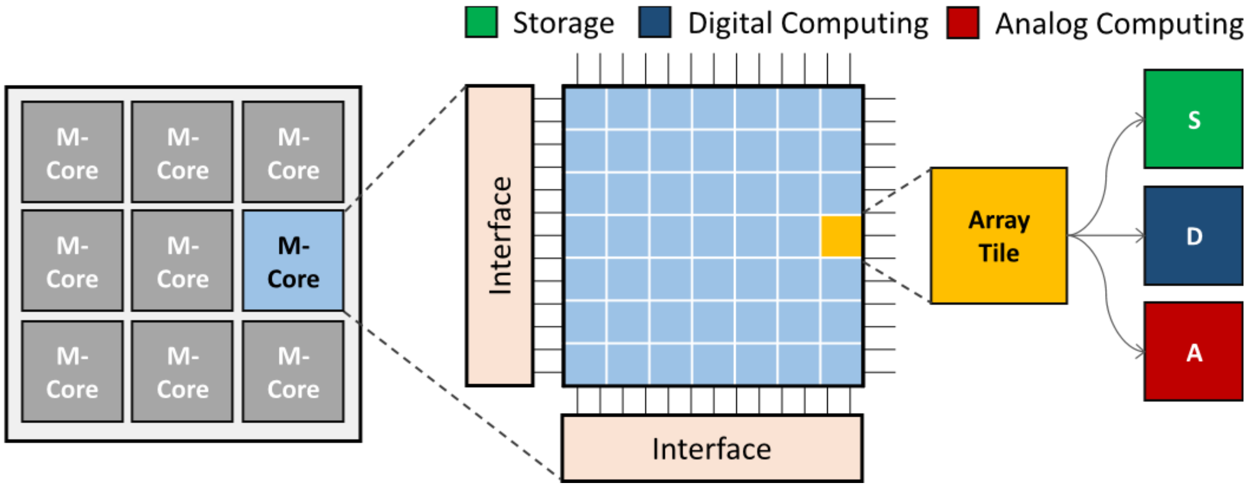
- **CMOS** Compatible
- **3D** Stackable, Scalable Architecture – Low thermal budget process
- **Architectures** proven include multiple Via schemes and Subtractive etching
- **Crossbar Inc** founded in 2010, $85M VC funding to date
- **Commercial Products** offered in 2016 based on 40nm CMOS



Crossbar

Dynamically reconfigurable Computing Fabric

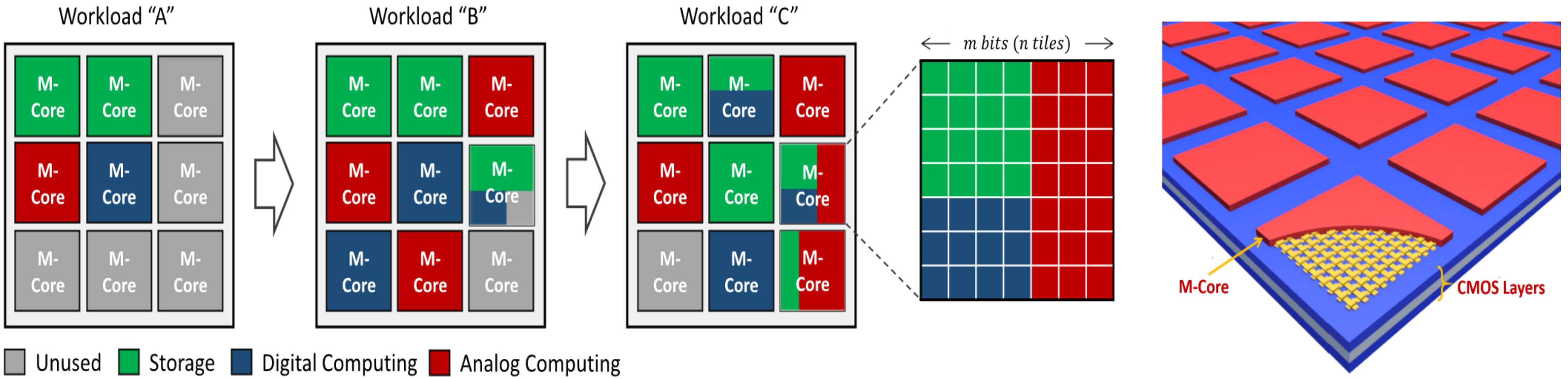
A reconfigurable hardware system with dense local connections and modular, asynchronous global connections



- Possibly FPGA-like modules, each module can be configured as a network with both feed-forward and feedback (recurrent) connections
- Spike based system with address-event coding
- Hierarchically structured interconnects: locally dense connection + globally asynchronous serial link
- “self-organized” computing modules at both fine-grained and coarse-grained levels
- Dynamically reconfigurable to adapt to the input data and the given problem (the “context”)

M. Zidan, Y. Jeong, J. H. Shin, C. Du, Z. Zhang, and W. D. Lu, IEEE Trans Multi-Scale Comp Sys, DOI 10.1109/TMSCS.2017.2721160 (2017)

Dynamically reconfigurable Computing Fabric



- "General" purpose by design: the same hardware supports different tasks – image, video, speech, …
- Dense local connection, sparse global connection
- **Run-time, dynamically reconfigurable. Function defined by software.**

M. Zidan, Y. Jeong, J. H. Shin, C. Du, Z. Zhang, and W. D. Lu, IEEE Trans Multi-Scale Comp Sys, DOI 10.1109/TMSCS.2017.2721160 (2017)

Summary

- Memristor arrays can already perform efficient image analysis and data clustering applications
- Taking advantage of the internal ionic dynamics at different time scales allow the device to more faithfully emulate biological system
- Memristor technology is already quite mature, especially for memory applications (products available)
- Towards dynamically reconfigurable circuits (i.e. software-defined chips) based on a common physical fabric

Acknowledgements

Grad students:

Jihang Lee, Wen Ma, Fuxi Cai, Yeonjoo Jeong, Jong Hong Shin, John Moon, Billy Schell, SeungHwan Lee, Qiwen Wang, Fan-Hsuan Meng

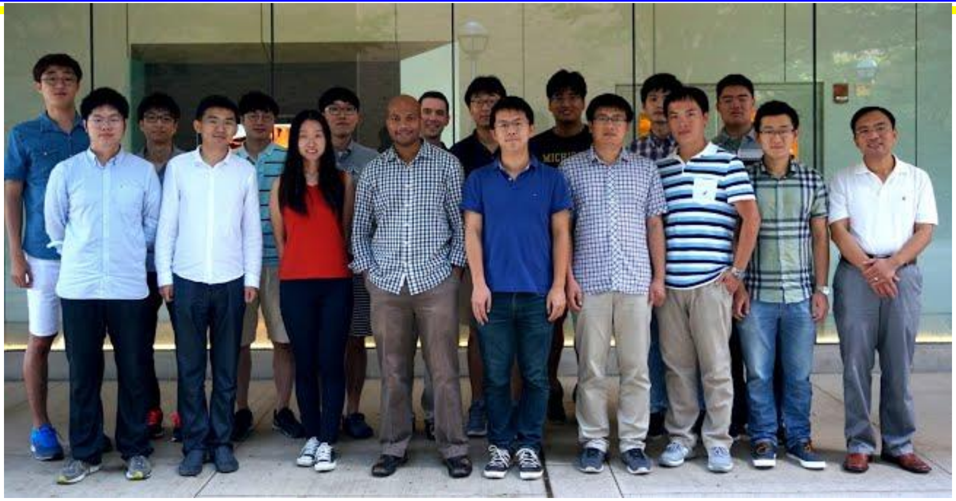
Former students

**Sung-Hyun Jo, *Kuk-Hwan Kim *Siddharth Gaba, *Ting Chang *Patrick Sheridan, *ShinHyun Choi, *Jiantao Zhou, Ugo Otuonye, *Chao Du, *Eric Dattoli *Wayne Fung, Lin Chen, *Seok-Youl Choi, *Woo Hyung Lee*

PostDocs:

Dr. Mohammed Zidan,
Dr. Xiaojian Zhu
Dr. Piotr Ogrodnik

**Dr. Yuchao Yang*
**Dr. Sungho Kim*
**Dr. Bing Chen*
**Dr. Taeho Moon*
**Dr. Zhongqing Ji*
** Dr. Qing Wan*
**Dr. L. Liu*
**Xiaojie Hao*



- Prof. Z. Zhang, Prof. M. Flynn, UM
- Dr. G. Kenyon, LANL
- Prof. C. Teuscher, PSU
- Prof. D. Strukov, UCSB,
- Prof. J. Hasler, GeorgiaTech,
- Prof. R. Li, CAS, China
- Dr. I. Valov, Prof. R. Waser
- **Crossbar Inc**

Funding:

- DARPA UPSIDE program
- National Science Foundation (ECS-0601478, CCF-0621823, ECCS-0804863, CNS-0949667, ECCS-0954621).
- DARPA SyNAPSE program
- Air Force MURI program, Air Force q-2DEG program
- Engineering Translational Research (ETR) Grant

Lu Group
U. Michigan