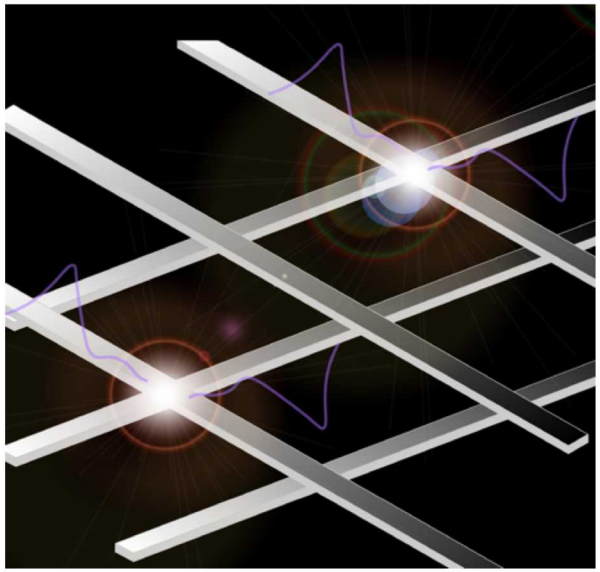
SAND2019-9702PE

Electrochemical materials for analog memory and neuromorphic computing

PRESENTED BY

Yiyang Li



Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

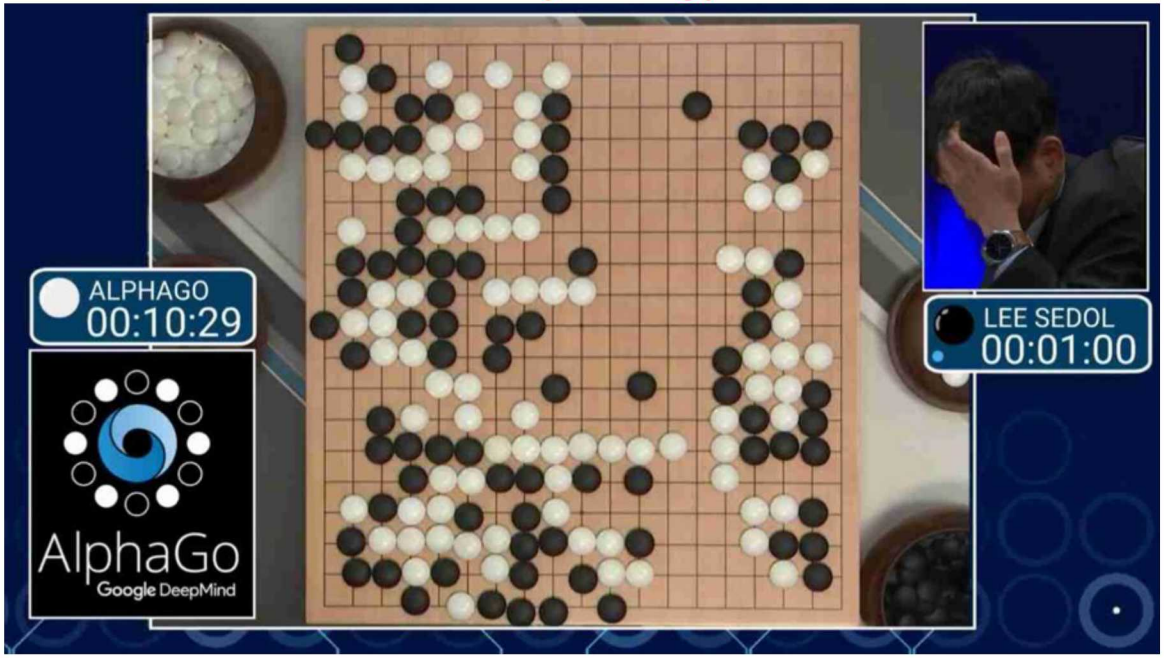
Computers vs the Human Brain

10^{12} FLOPS

~1 FLOPS

Machine learning can beat humans in certain tasks,
but are extremely energy intensive

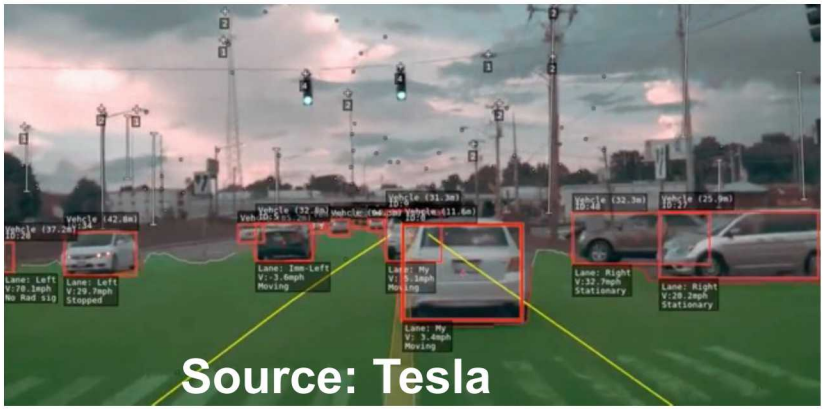
10^{6} Watt



20 Watt

Example: autonomous driving

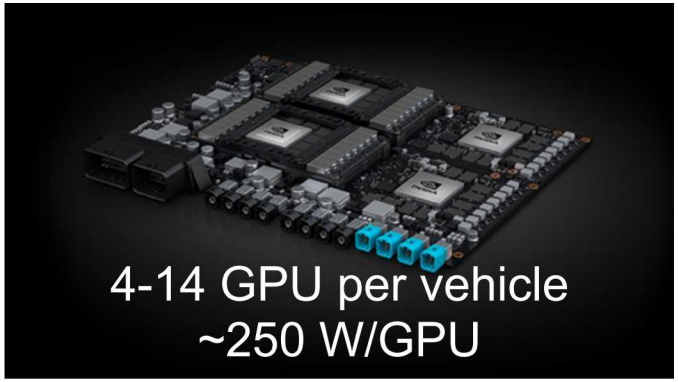
Chevy Bolt



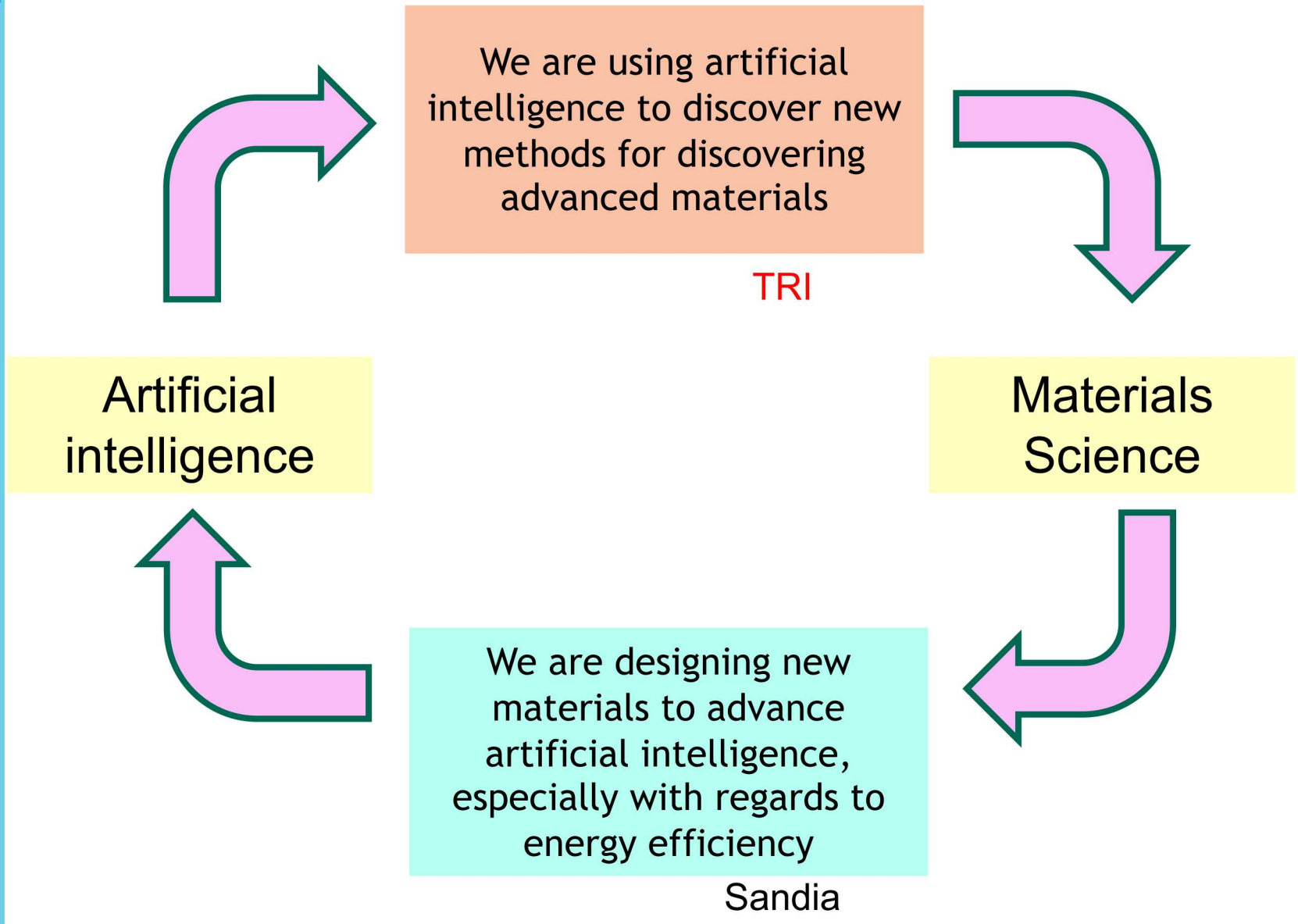
~100 GB/sec

Human brain: 10 Watt

Computing: 1-4 kW



Nvidia


We are using artificial intelligence to discover new methods for discovering advanced materials
TRI
Materials Science
We are designing new materials to advance artificial intelligence, especially with regards to energy efficiency
Sandia
Artificial intelligence

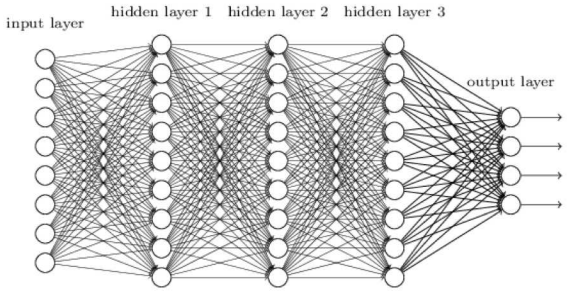
Artificial neural networks

Which one of these images is a cat?

Image recognition
Autonomous driving
Natural language processing

Artificial neural networks: use training examples and error backpropagation to find the matrix weights that correctly maps the input **x** onto the desired output **y**

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,n} \\ \vdots & \ddots & \vdots \\ w_{m,1} & \cdots & w_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$



Inference and training are very energy expensive

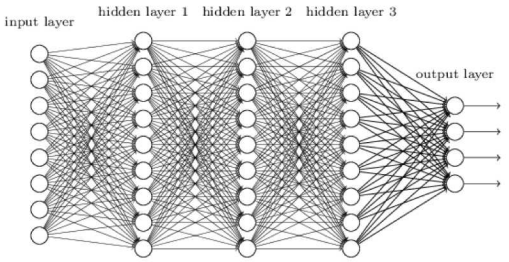
n, m > 1000

Andrew Ng, Coursera
Nawrocki et al. *IEEE Elec. Dev.* 2016

Digital and analog implementations of neural networks

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,n} \\ \vdots & \ddots & \vdots \\ w_{m,1} & \cdots & w_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

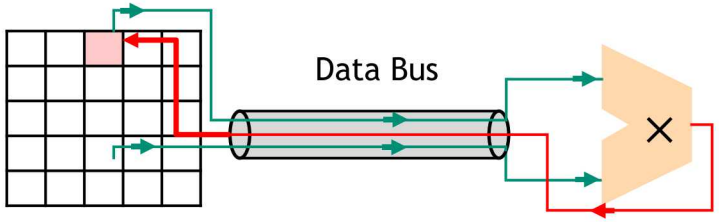
n, m > 1000



Von Neumann Digital
Separate logic and memory structures

SRAM to store the weights

Arithmetic logic unit for multiplication

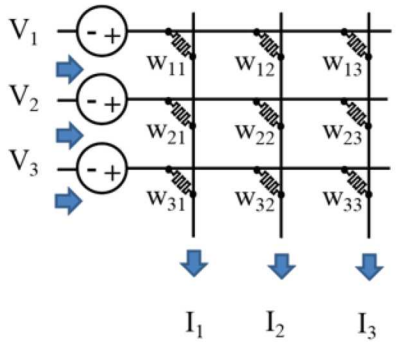


Uses established CMOS technology
Data bus results in latency and power

In-memory Parallel Analog
Use non-volatile memory

Crossbar for matrix multiplication

Conductance of each element can be changed in a predictable manner



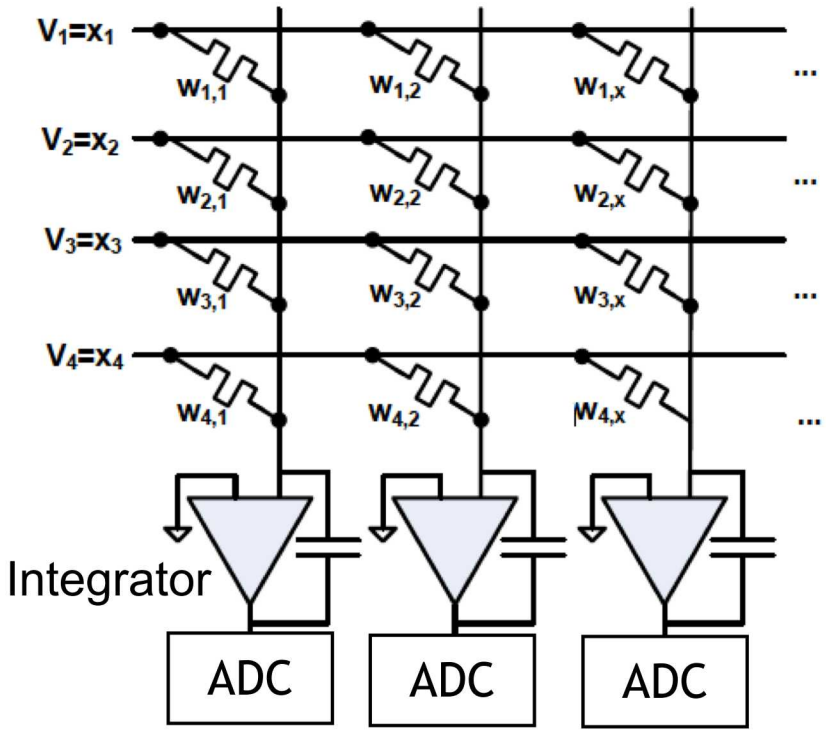
$I_1 = V_1W_{11} + V_2W_{21} + V_3W_{31}$

Simultaneous logic and memory
3 orders of magnitude less power

M. Marinella, *IEEE Circuits and Systems*, 8, 86-101, 2018

Energy consumption: analog vs digital

Analog core: doing $O(n^2)$ operations using $O(n)$ inputs and outputs



M. Marinella, *IEEE Circuits and Systems*, 8, 86-101, 2018

Energy per operation per array,
1024X1024 synapses
1 Million multiply-and-accumulate

	Analog	Digital
Vector matrix multiply	13 nJ	2600 nJ
Outer product update	2 nJ	3600 nJ

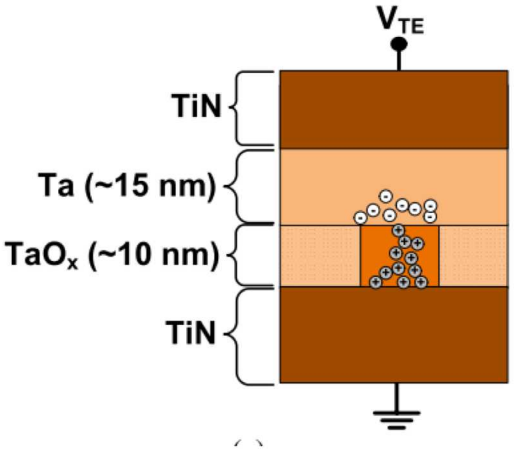
Analog energy breakdown:
Read: 0.4 nJ
Write: 1.66 nJ
Integrator: 2.8 nJ
ADC: 9.4 nJ

Analog: ~200 Tera-OPS/watt
Digital: 1-5 Tera-OPS/watt

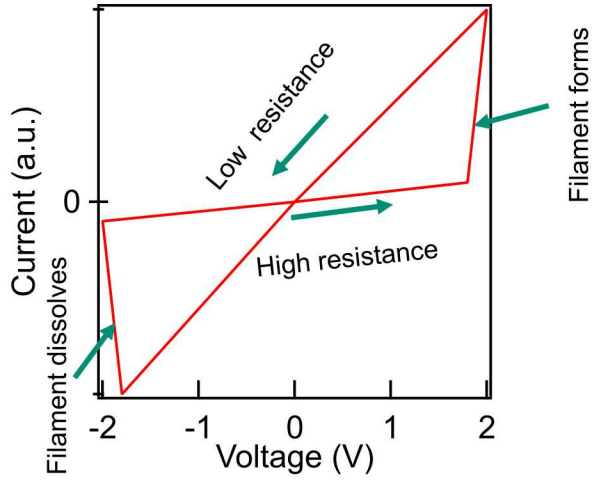
Challenge: analog non-volatile memory that meets these requirements do not exist

Two-terminal resistive memory

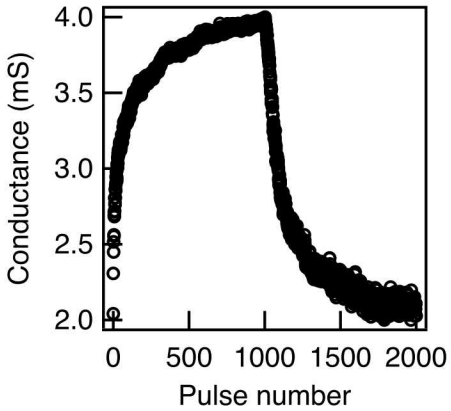
TaO_X resistive memory cell



Resistive memory switch between high resistance and low resistance states

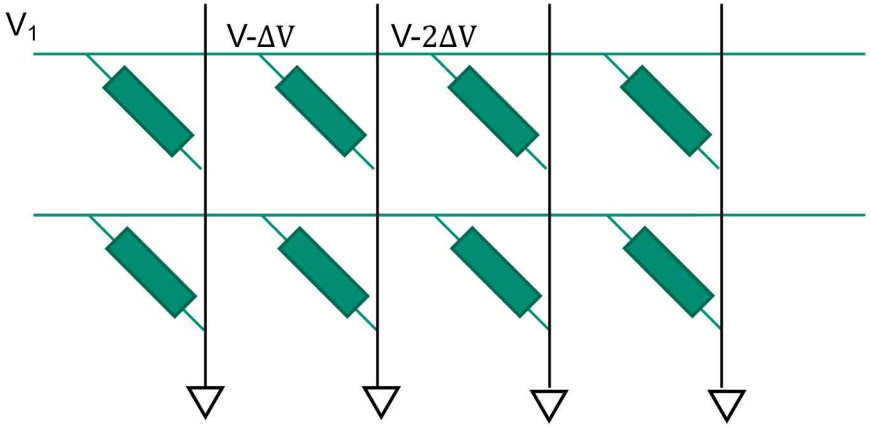


Analog resistive memory is nonlinear and unpredictable



R. Jacobs-Gedrim et al. *ICRC*, 2017
Q. Xia & J. Yang, *Nat. Mater.*, 18, 303 (2019)

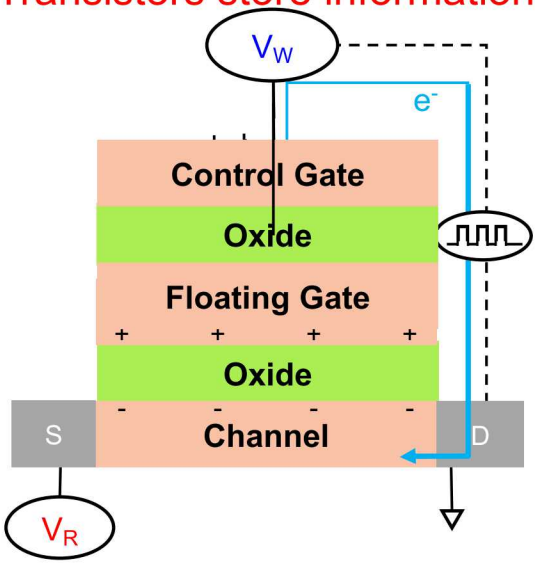
$\Delta V \sim R_{wire}/R_{device}$



Target: $R_{device} = 10^7\text{-}10^9\ \Omega$

Three-terminal memory

Transistors store information by moving charge between the gate and the channel



$$\tau_{retention} = C_{gate}\, R_{OFF}$$

$$C_{gate} = \epsilon_0 \epsilon_r \frac{A}{t_{ox}} \approx 10^{-6}\ \mathrm{F\ cm^{-2}} = 10^{-16}\ \mathrm{F}\ \ (A = 100\ \mathrm{nm^2})$$

Option 1: transistor switch: $R_{OFF} \sim 10^{15}\ \Omega$

$\tau_{retention} < 1$ s (Resistive Processing Unit, dynamic RAM)

S. Ambrogio, G. Burr, et al. Nature, **558**, 60 (2018)

Option 2: floating gate oxide switch, $R_{OFF} \sim \infty$

$\tau_{retention} > 10$ yr (Flash memory)

10V, high power, limited endurance (10^5)

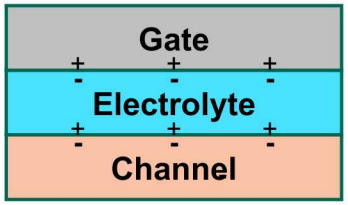
Improving retention by increasing C_{gate}? $C = \frac{\Delta Q}{\Delta V}$

Electrochemical double-layer transistors

Electrons stored at the **interface**

$C_{gate} \sim 10^{-4}\ \mathrm{F\ cm^{-2}}$

$t_{"ox"} < 1\ nm$



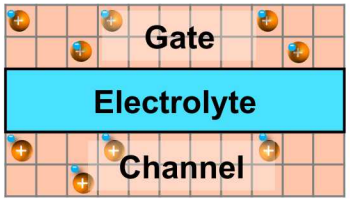
P. Gkoupidenis, G. Malliaras, Adv Mater, 27, 7176 (2015)

Redox transistors

Electrons (and ions) are stored in the **bulk** of the material (pseudo-capacitor)

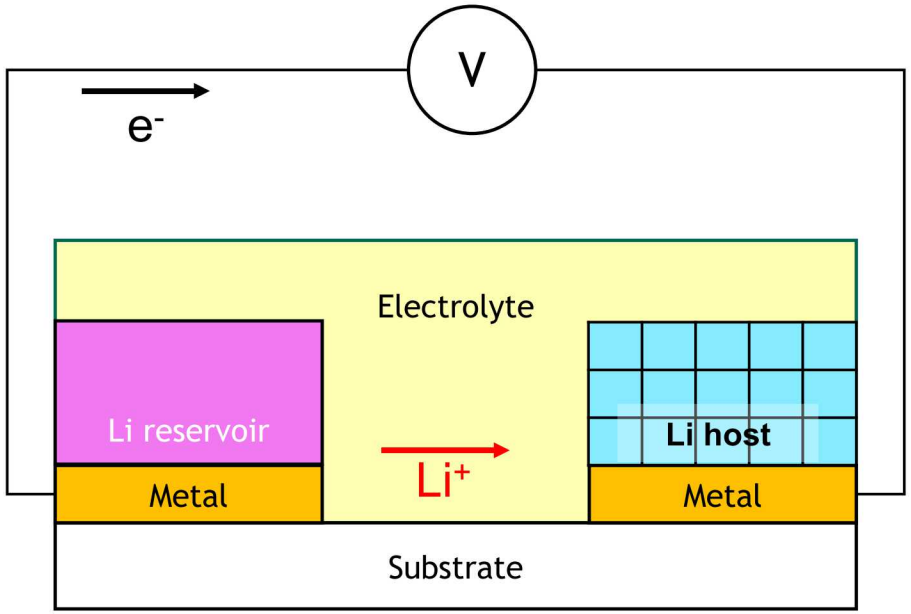
Electron/ion neutral ambipolar pair enables ΔQ without electrostatic charging

$C_{gate} \sim 10^{-1}\ \mathrm{F\ cm^{-2}}$ (100-nm film)



E. Fuller, Adv Mater. 29, 1604310 (2017)
Y. van de Burgt, Nat. Mater., 16, 414 (2017)

Redox transistors based on electrochemical ion insertion



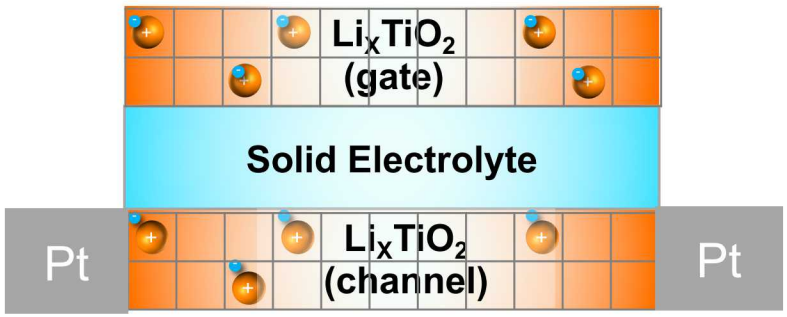
Dynamic control of the lithium doping level using current and voltage

In batteries, intercalation provides a high density of energy storage;
In redox transistors, intercalation provides a high density of information storage

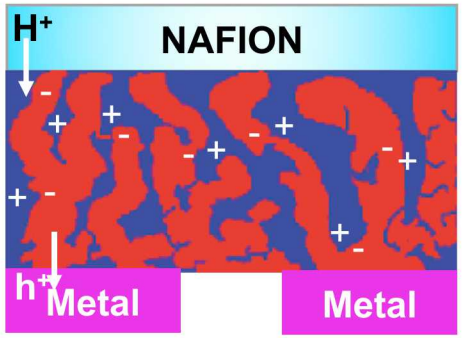
Y. Li, W. Chueh. Ann. Rev. Mater. Res. 48, 137 (2018)

Outline

Part 1: Inorganic redox transistor
Lithium insertion into TiO_2

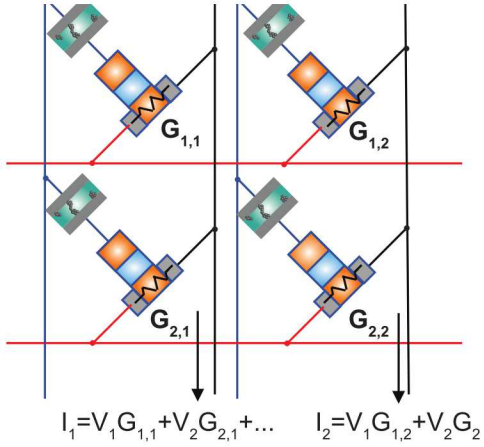


Part 2: Organic redox transistor
Proton insertion into PEDOT:PSS

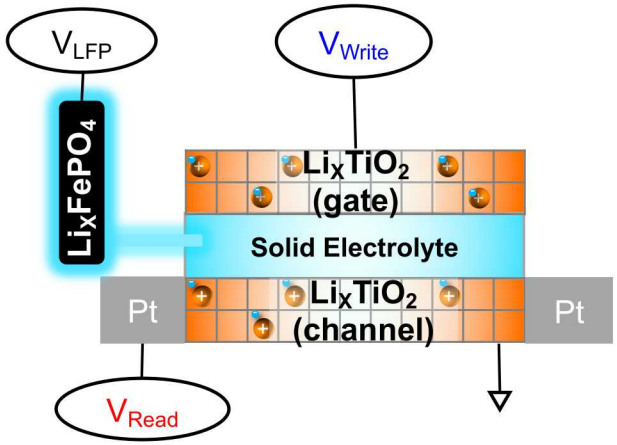


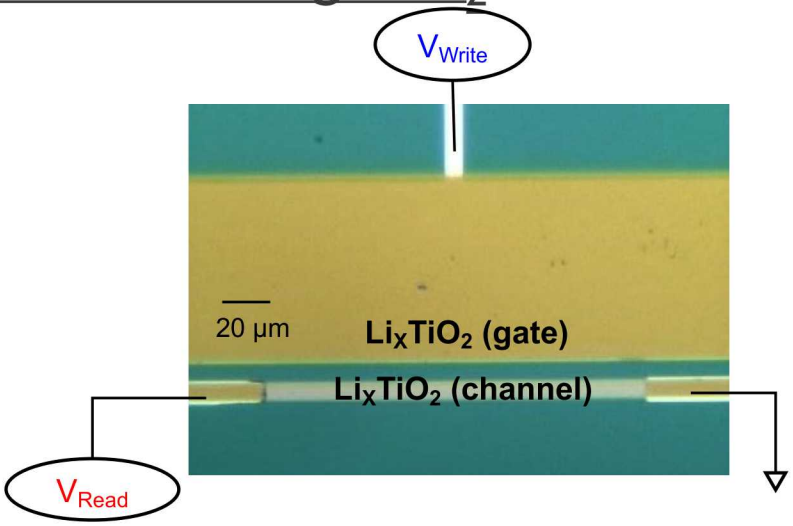
H^+: proton; h^+: hole
PEDOT:PSS

Part 3: Crossbar Array integration



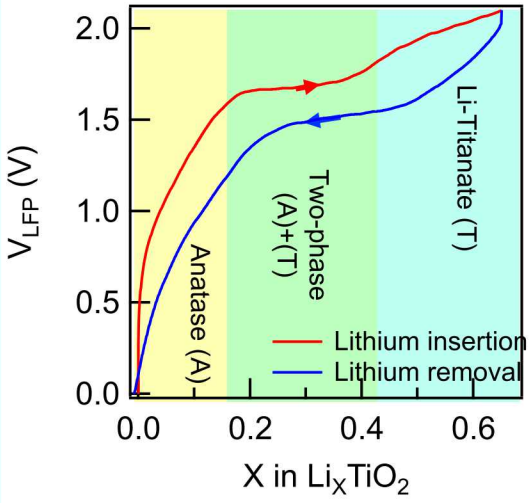
Symmetric redox transistor using TiO_2

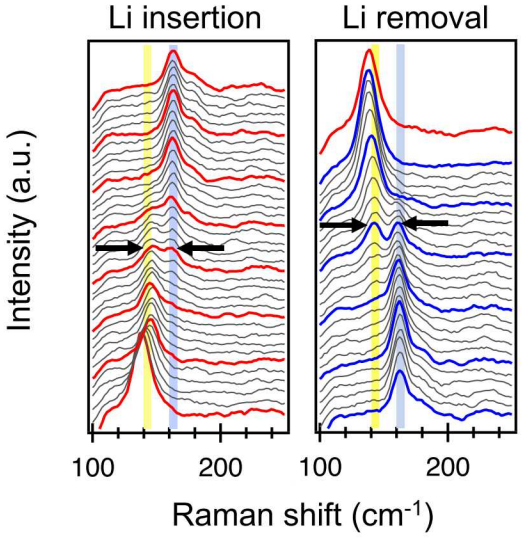




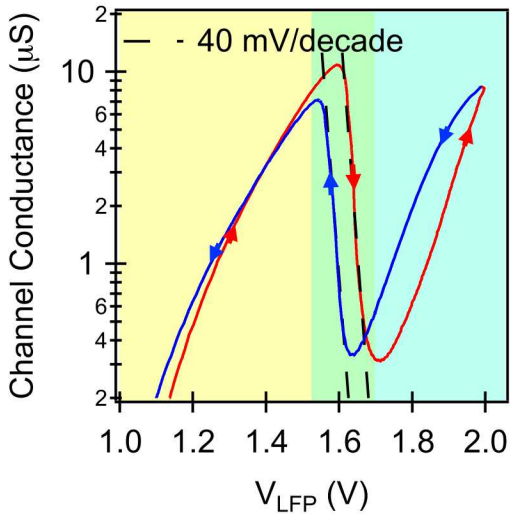
A. Talin

E. Fuller

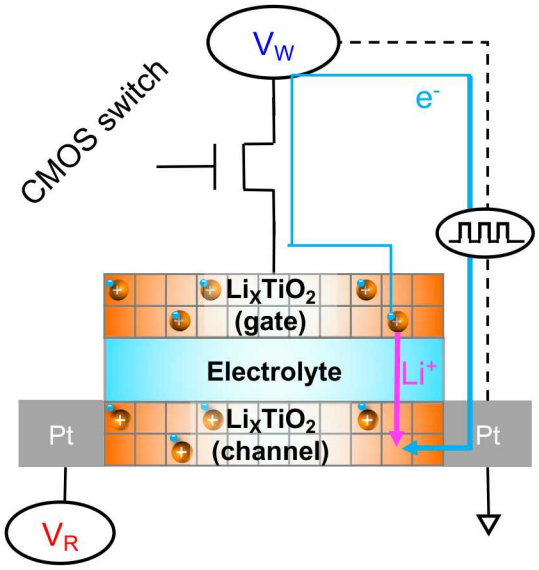




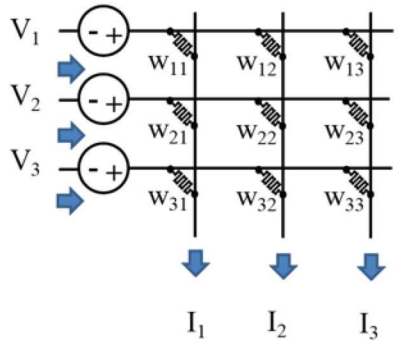
$$Li^+ + e^- + Ti_{Ti}^X \rightarrow Li_i^\circ + \boldsymbol{Ti'_{Ti}}$$

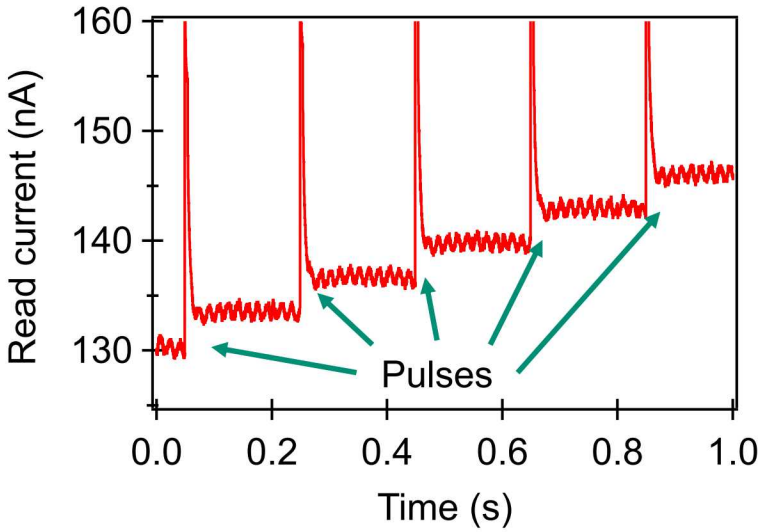


Redox transistor using Li_XTiO_2 (anatase)



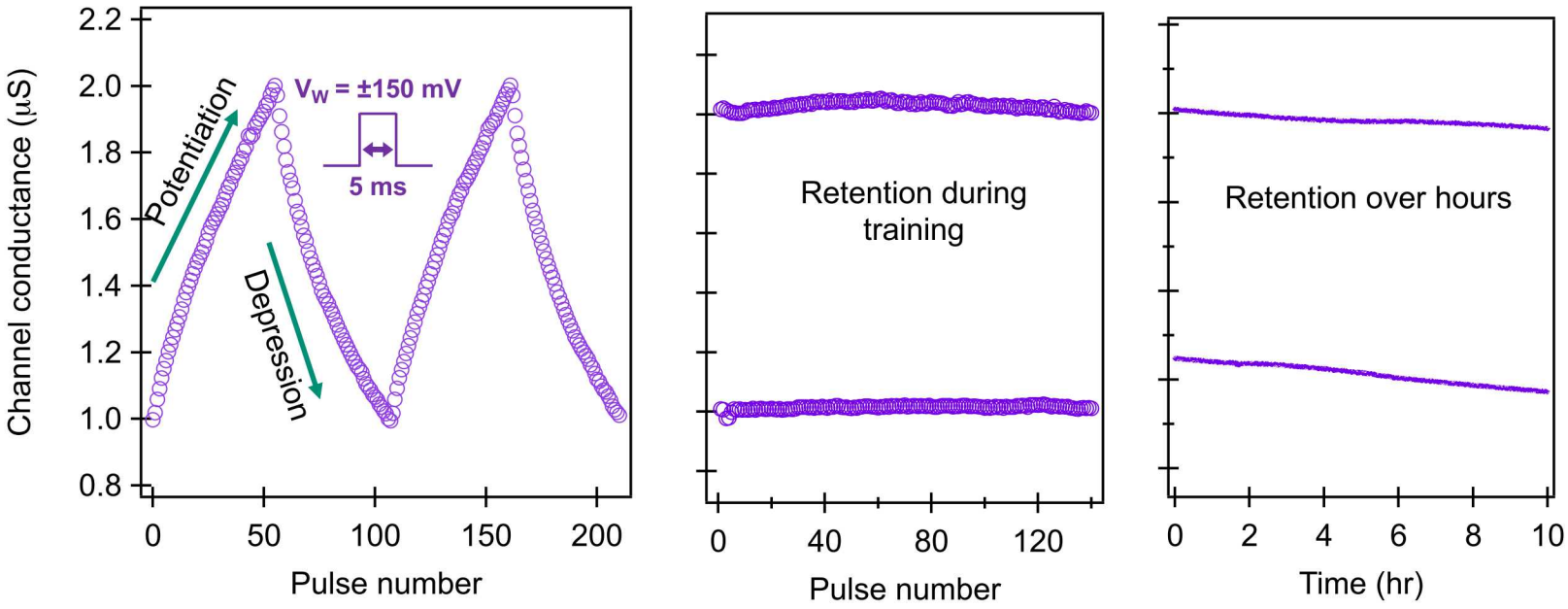
V_W = ±150 mV





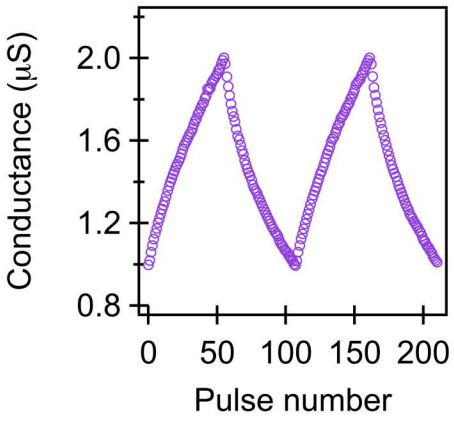
- Linear and symmetric programming
  - 150-mV write pulses
  - Analog, non-volatile states

Redox transistor using Li_XTiO_2 (anatase)

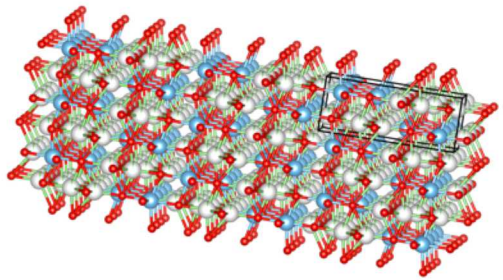


Linear programming and high accuracy

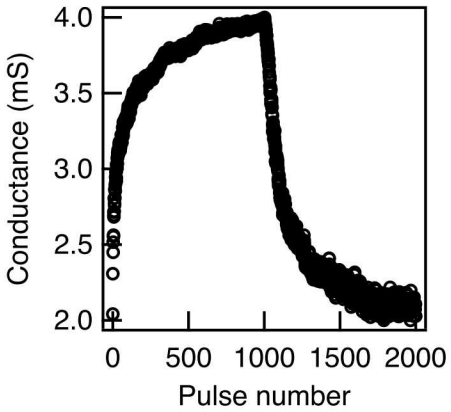
Li_XTiO_2 transistor
Voltage: ± 0.15 V



Redox transistors store information continuously as dopants in a crystal

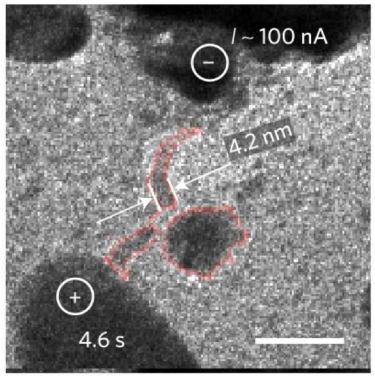


TaO_x Memristor
Voltage: ±1 V



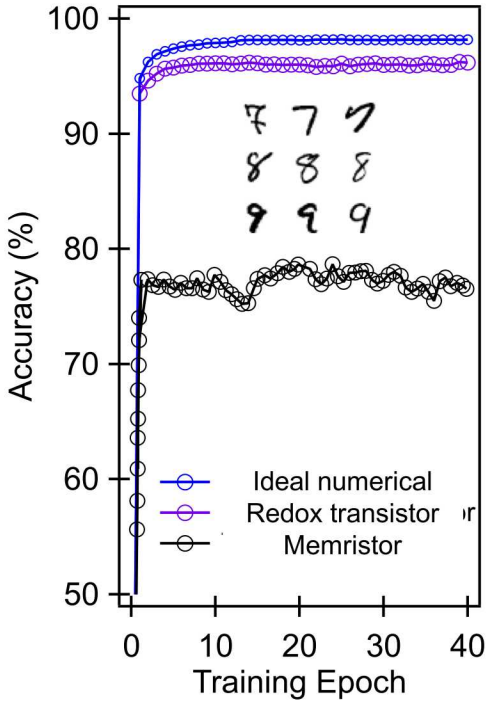
R. Jacobs-Gedrim et al. *ICRC*, 2017

Memristors store information at filaments



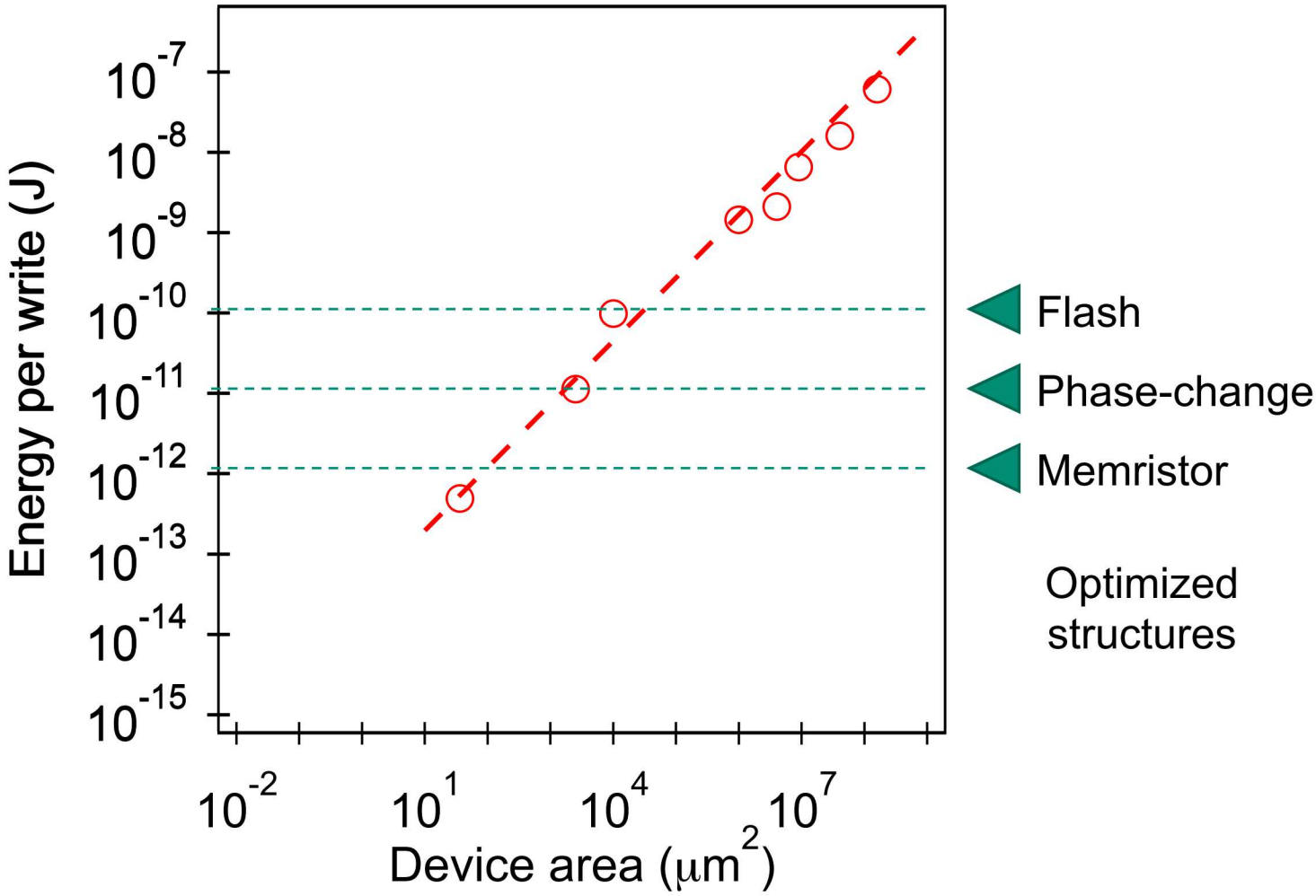
Wang et al. *Nat Mater.* 2017

Linearity is essential to train an accurate neural network



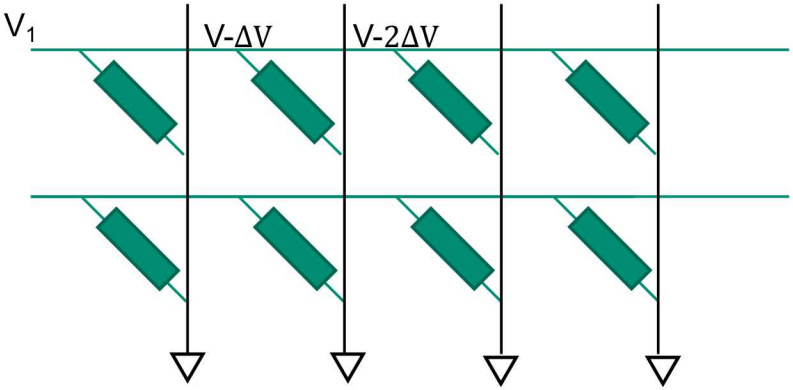
Crossbar Simulator

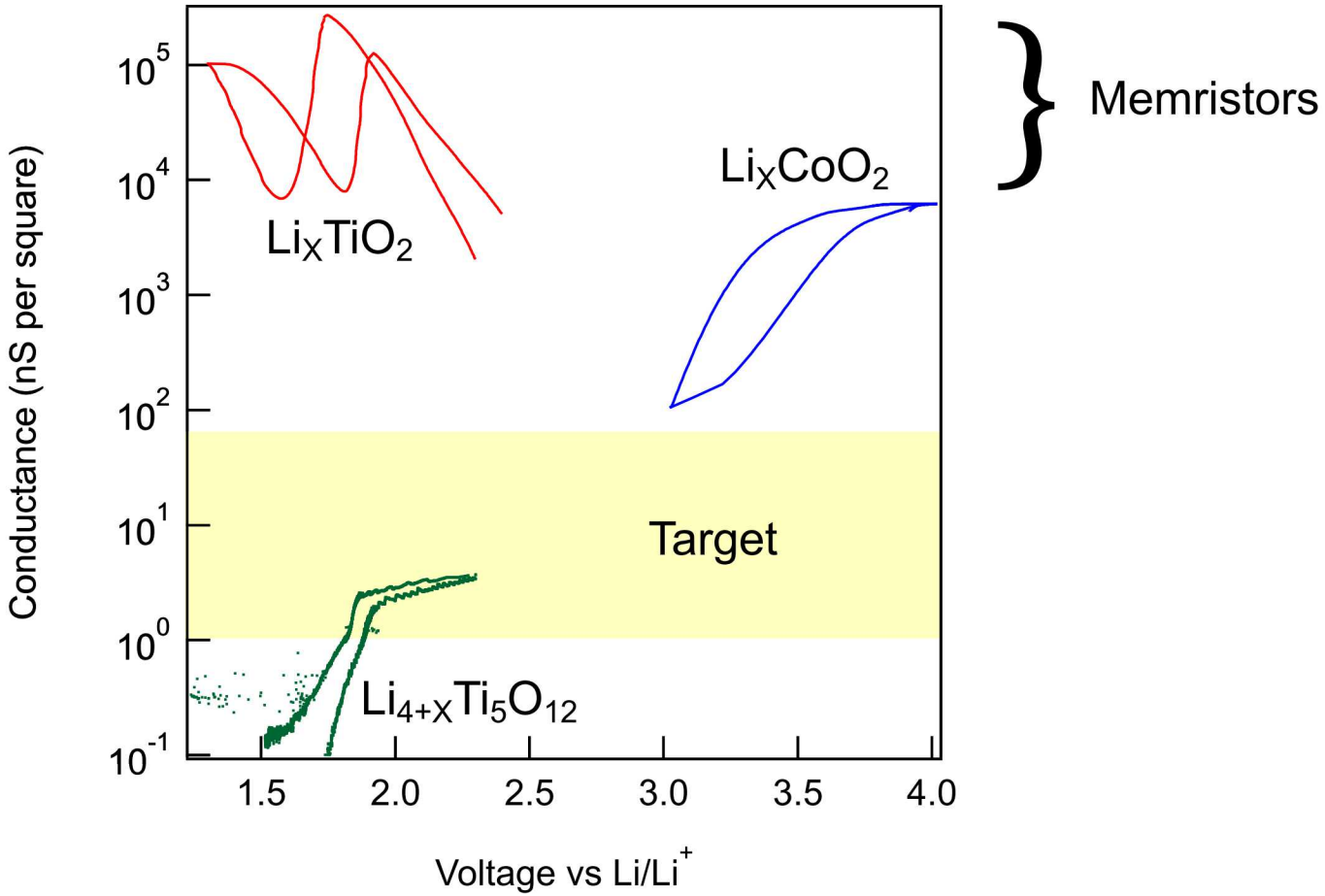
Low write energies



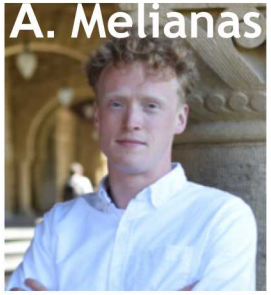
Y. van de Burgt, *Nat Mater*, 16, 414(2017)
M. Sharbati *Adv. Mater.* 30, 1802353 (2018)
J. Yang et al. *Nat. Nano.* 8, 13(2012)

High resistance devices

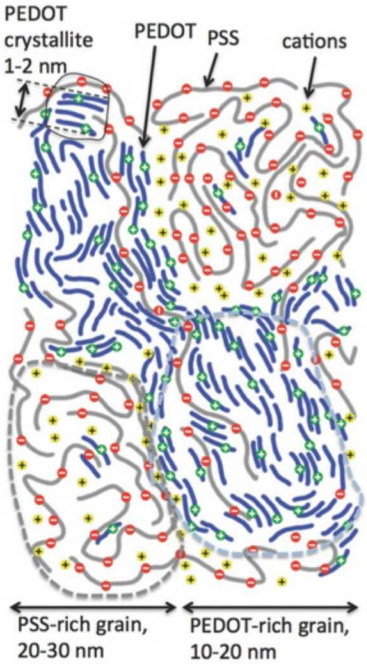




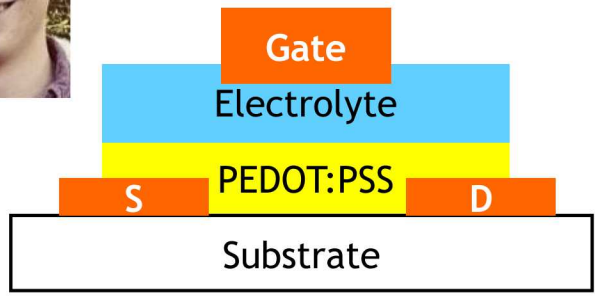
Part 2: Organic redox transistor based on PEDOT:PSS



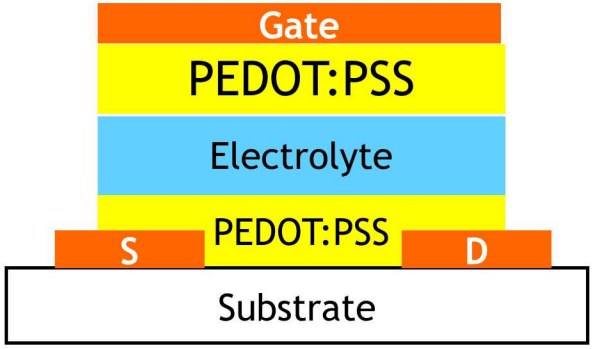
PEDOT (hole conductor)
PSS (proton conductor)



A. Volkov, et al. *Adv. Funct. Mater.* 27, 1700329 (2017)



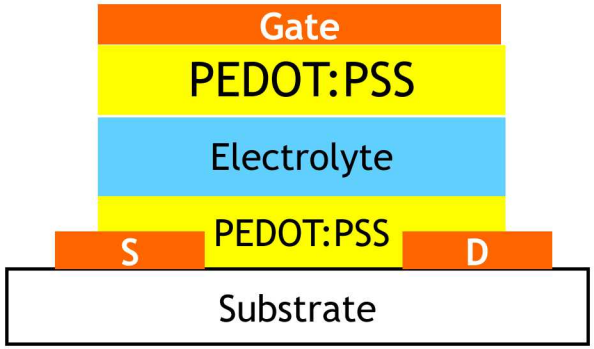
Metal/electrolyte double-layer capacitance ~ 20 μF/cm^2

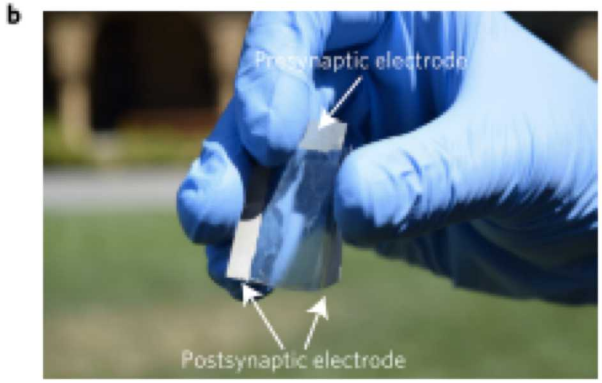


Volumetric charge density: 40 F/cm^3
~ 400 μF/cm^2 (100-nm film)

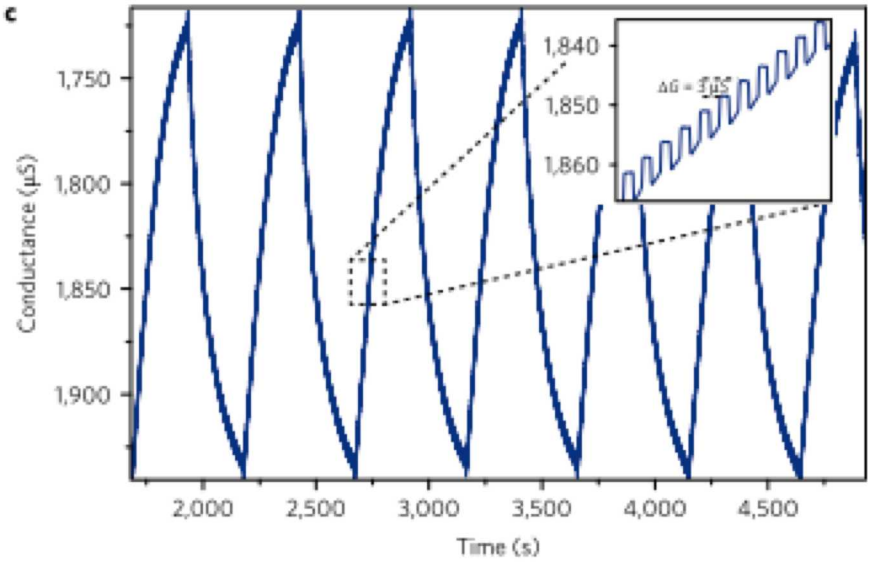
PEDOT:PSS Redox Transistor

Flexible device



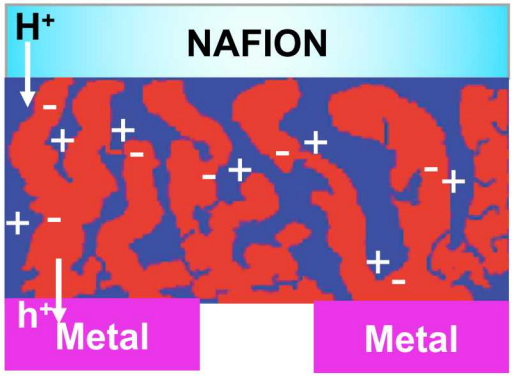


Linear weight updates



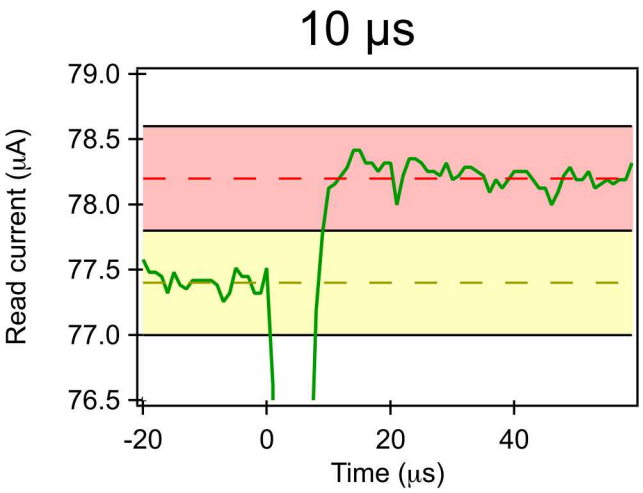
Y. van de Burgt, et al. *Nat Mater*, **16**, 414(2017)

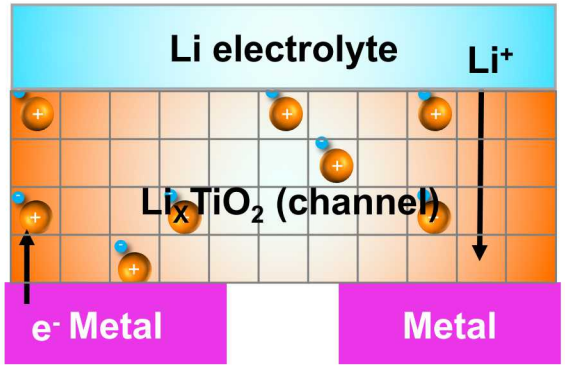
Super-capacitative vs intercalation redox transistor



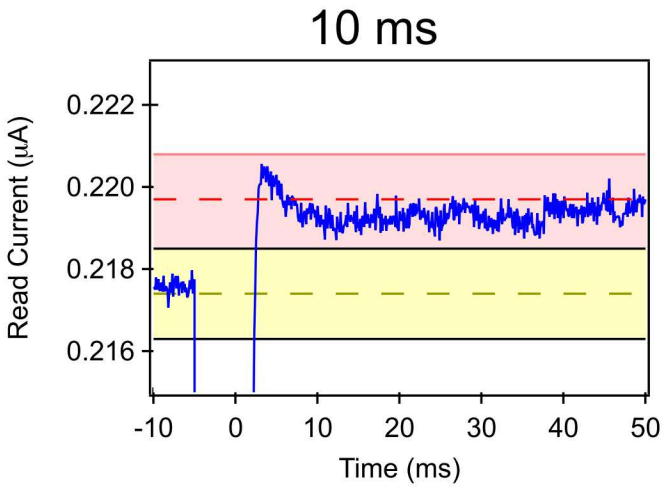
H^+: proton; h^+: hole
PEDOT:PSS

No charge transfer
Fast single-specie diffusion
Low capacitance: 40 F cm^{-3}



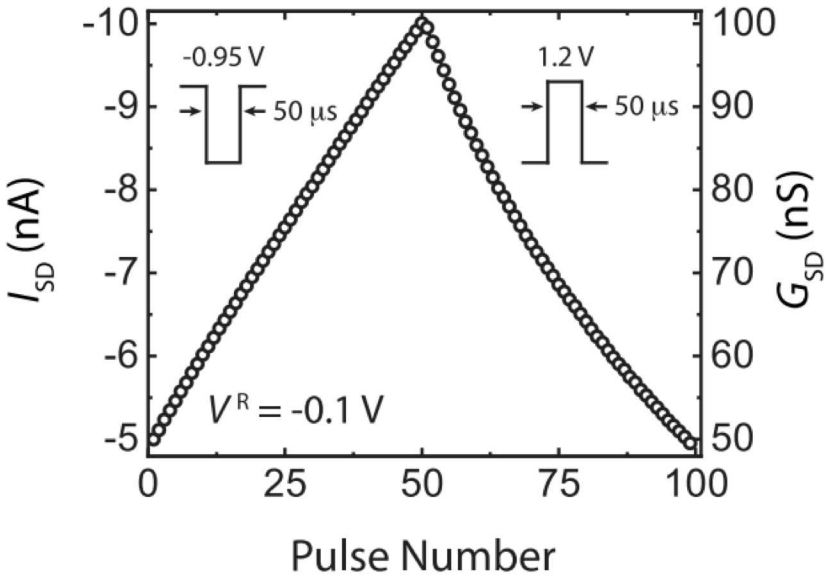


Interfacial charge transfer
Slow chemical diffusion
High capacitance: 5000 F cm^{-3}



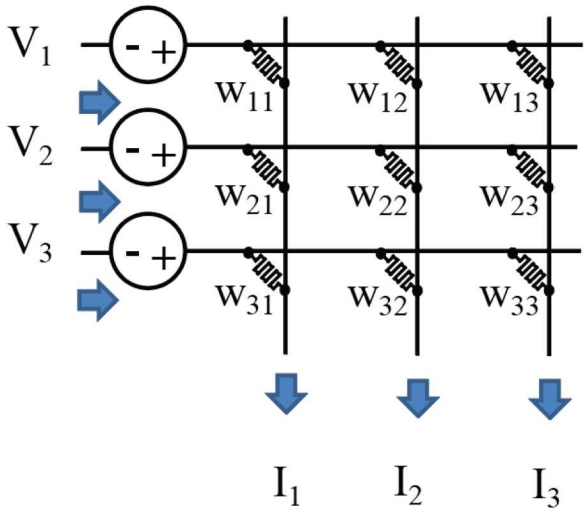
High resistance devices

10^7 Ohm achieved by doping with filler PEI

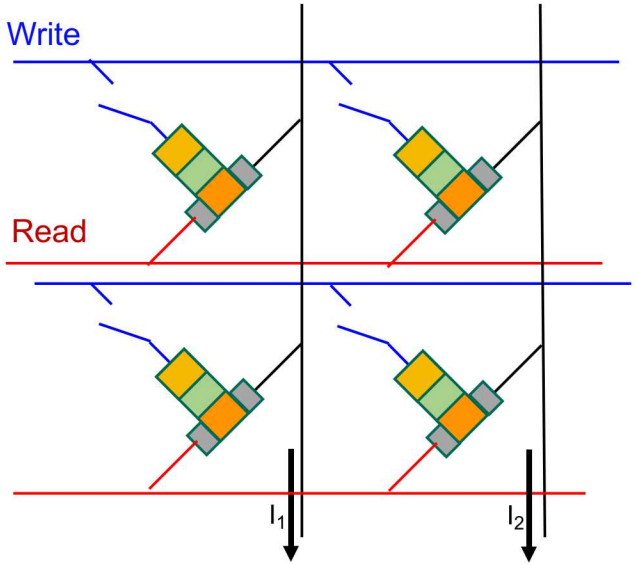


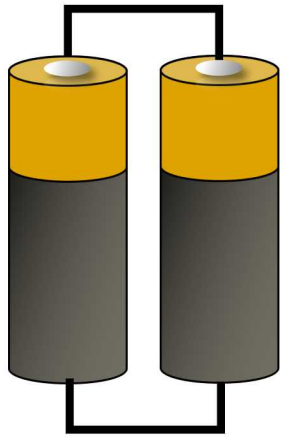
Part 3: Crossbar array integration

Matrix multiplication crossbar



Gates and channels are shorted to each other





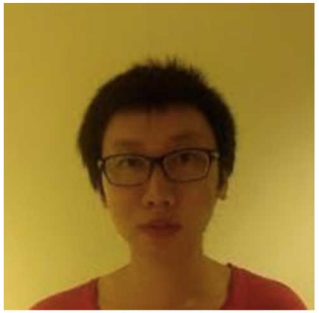
Challenge: find a highly nonlinear switch with a high ON/OFF ratio at low voltages

$$\tau_{retention} \sim R_{switch} * C_{device}$$

Diffusive memristor

Two-terminal switch

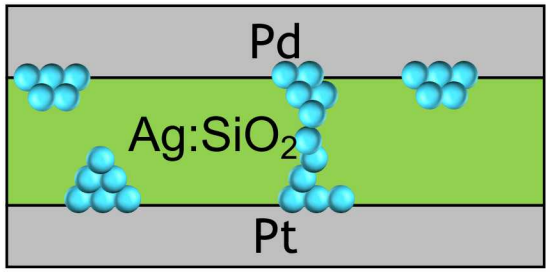
J. Yang

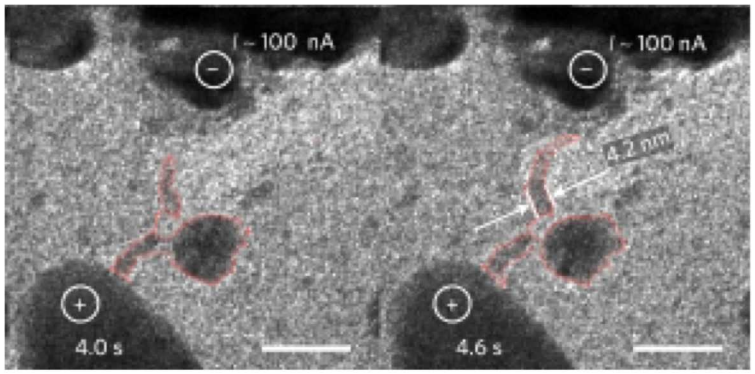


Z. Wang

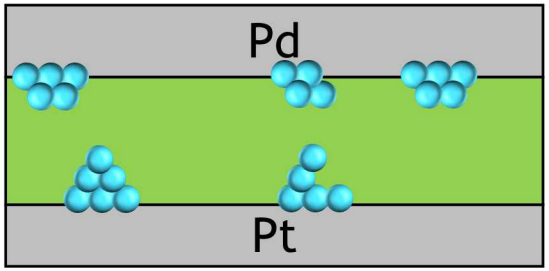
S. Asapu

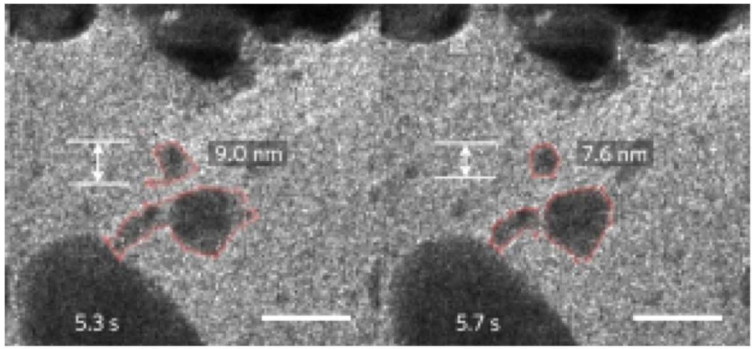
ON





OFF

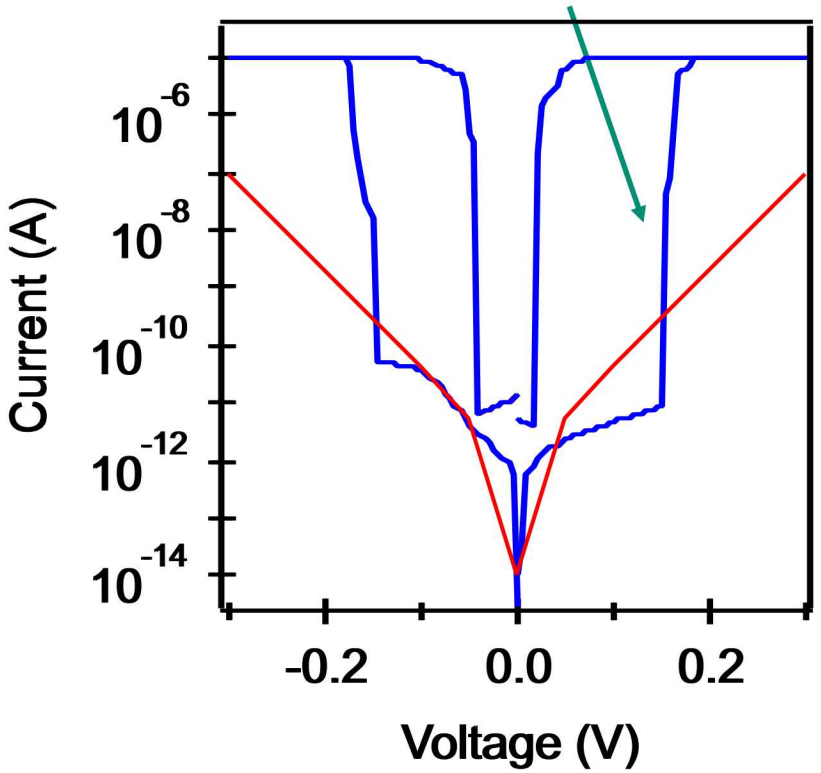




Z. Wang et al. Nat. Mater. 16, 101 (2017)

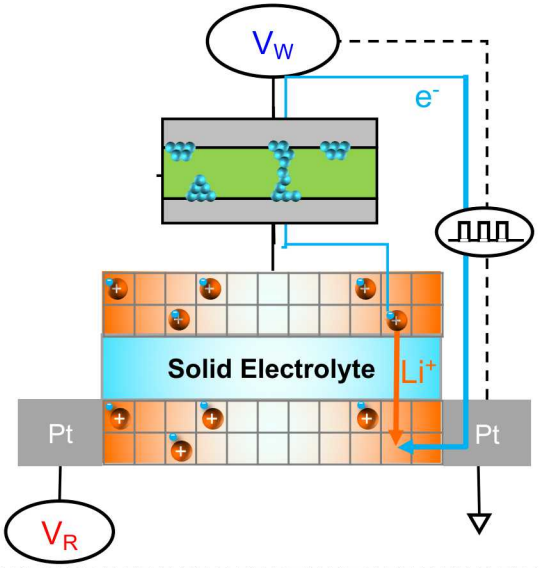
High nonlinearity at low voltages

Diffusive Memristor: ~1 mV/decade,
10^7 ON/OFF



Ideal diode: 60 mV/decade
Need much higher voltages
for same ON/OFF ratio

Low-voltage, Si-free electrochemical memory



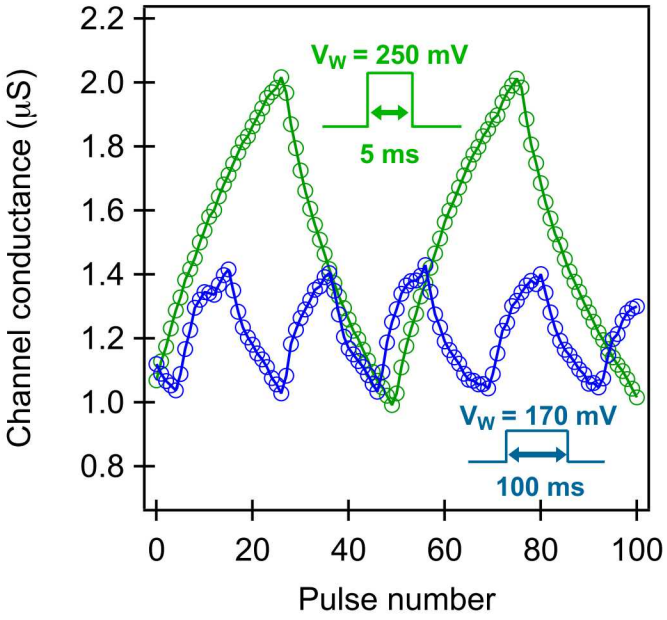
Diffusive memristor (Ag in SiO_X): high ON/OFF ratio

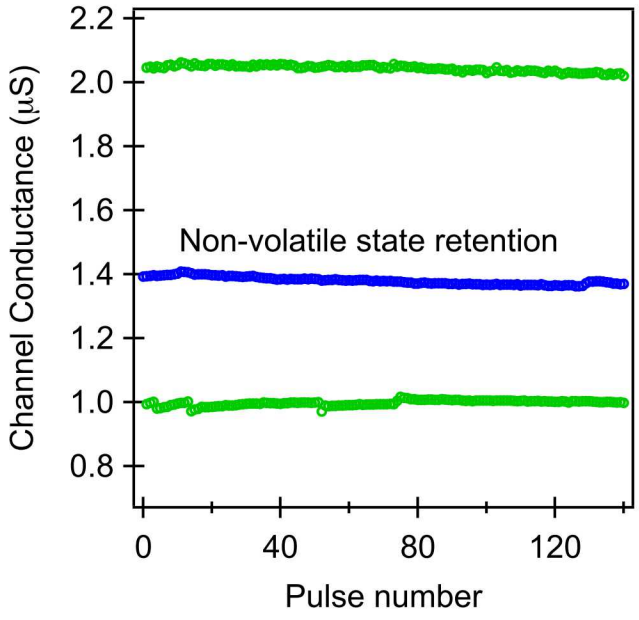
Redox transistor: high charge density via bulk storage

Both: Low switching voltages

Wang et al. Nat. Mater. 2017

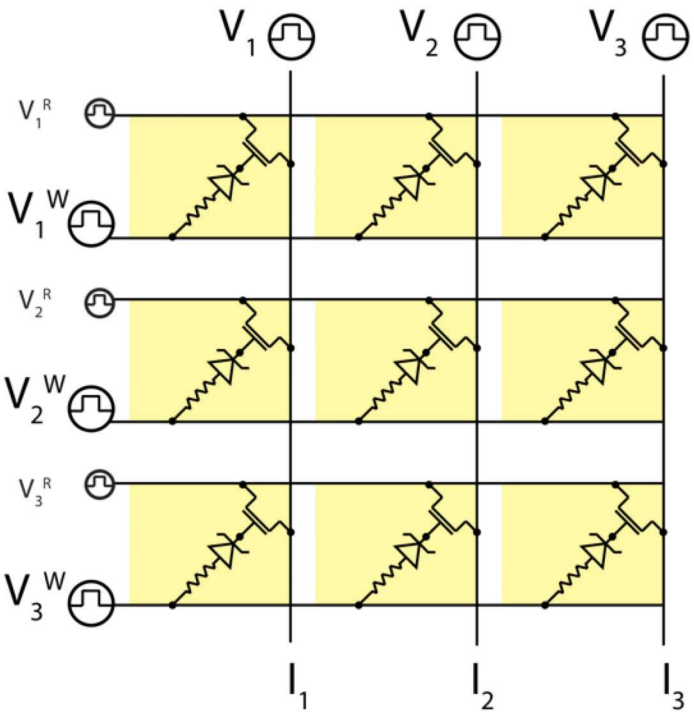
Non-volatile memory that switches at just 6 times the thermal voltage

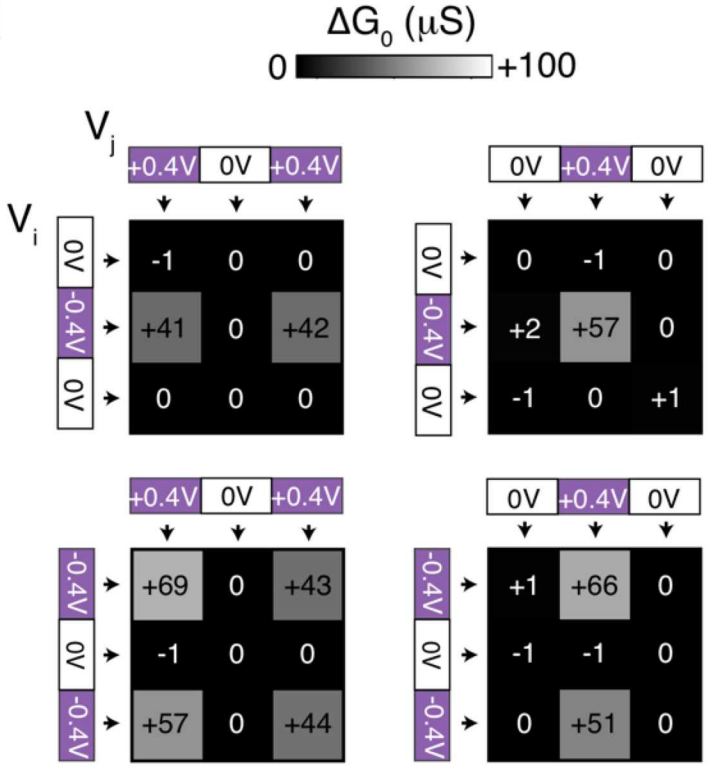




Parallel weight updates in crossbar arrays

Speed of a network ~ speed of a device * # of devices switched in parallel

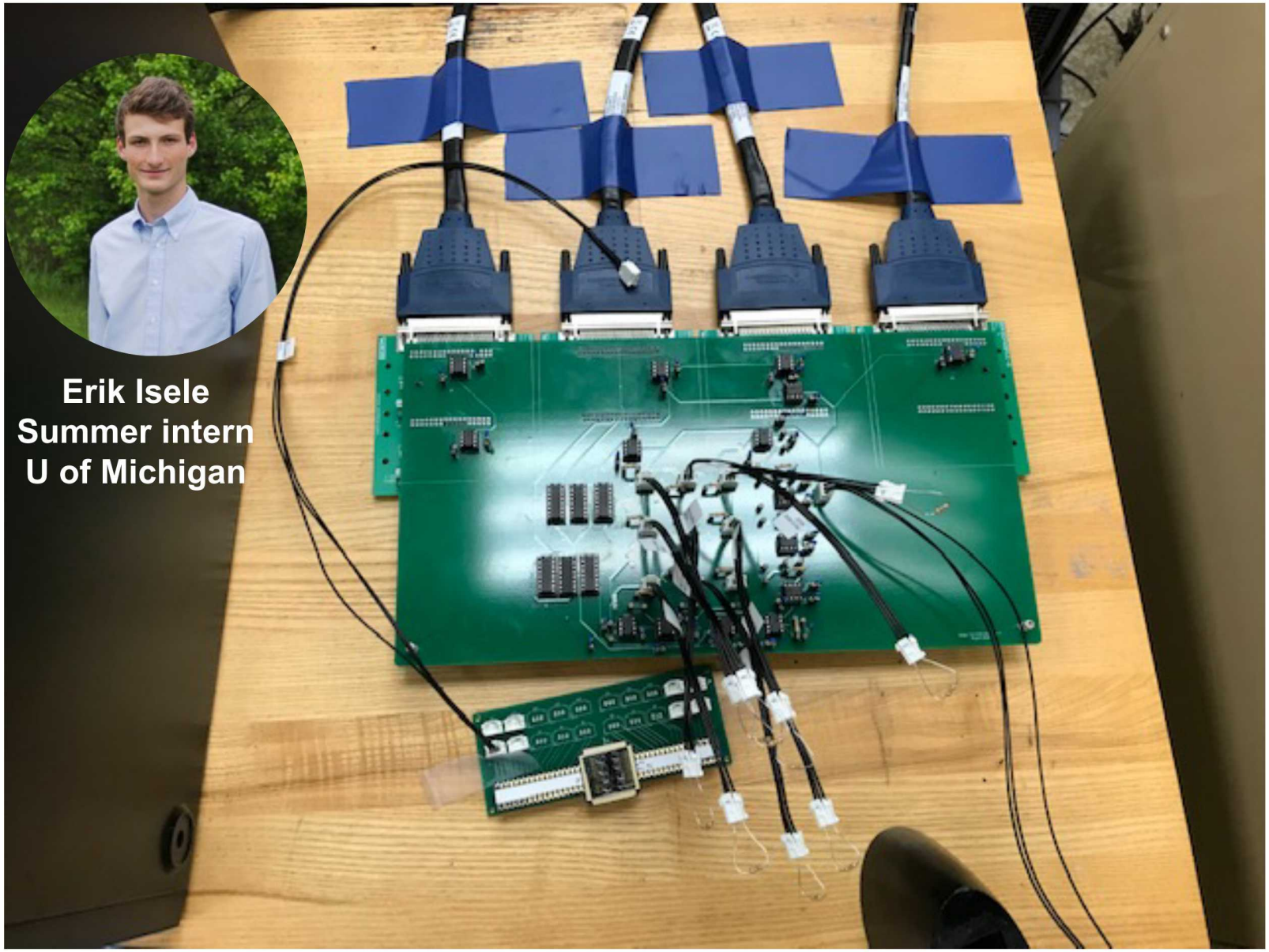




We can select the rows and columns in this crossbar to update the weights of many synapses in a highly parallel fashion in accordance with an outer product update

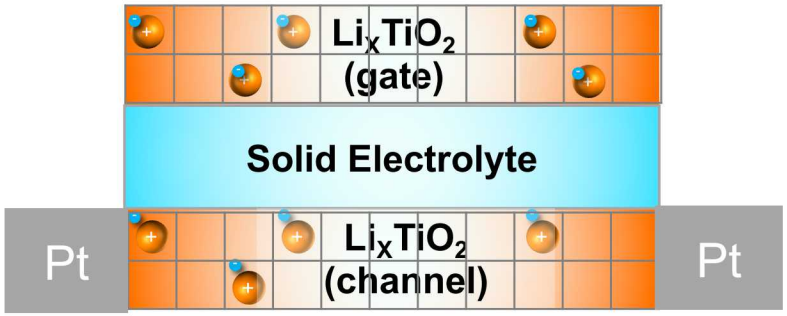
E. Fuller, et al. *Science*, 364, 570 (2019)

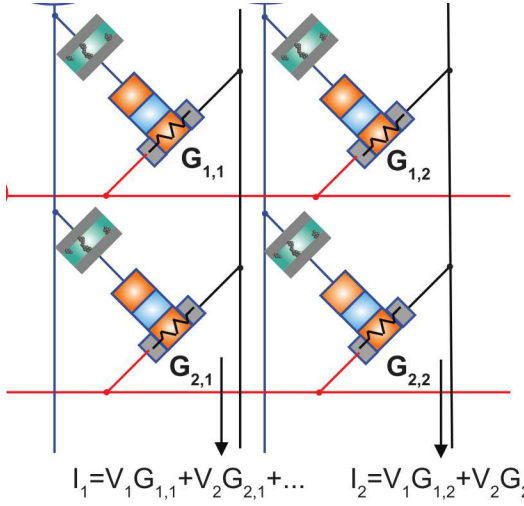
Next step: learning demonstration



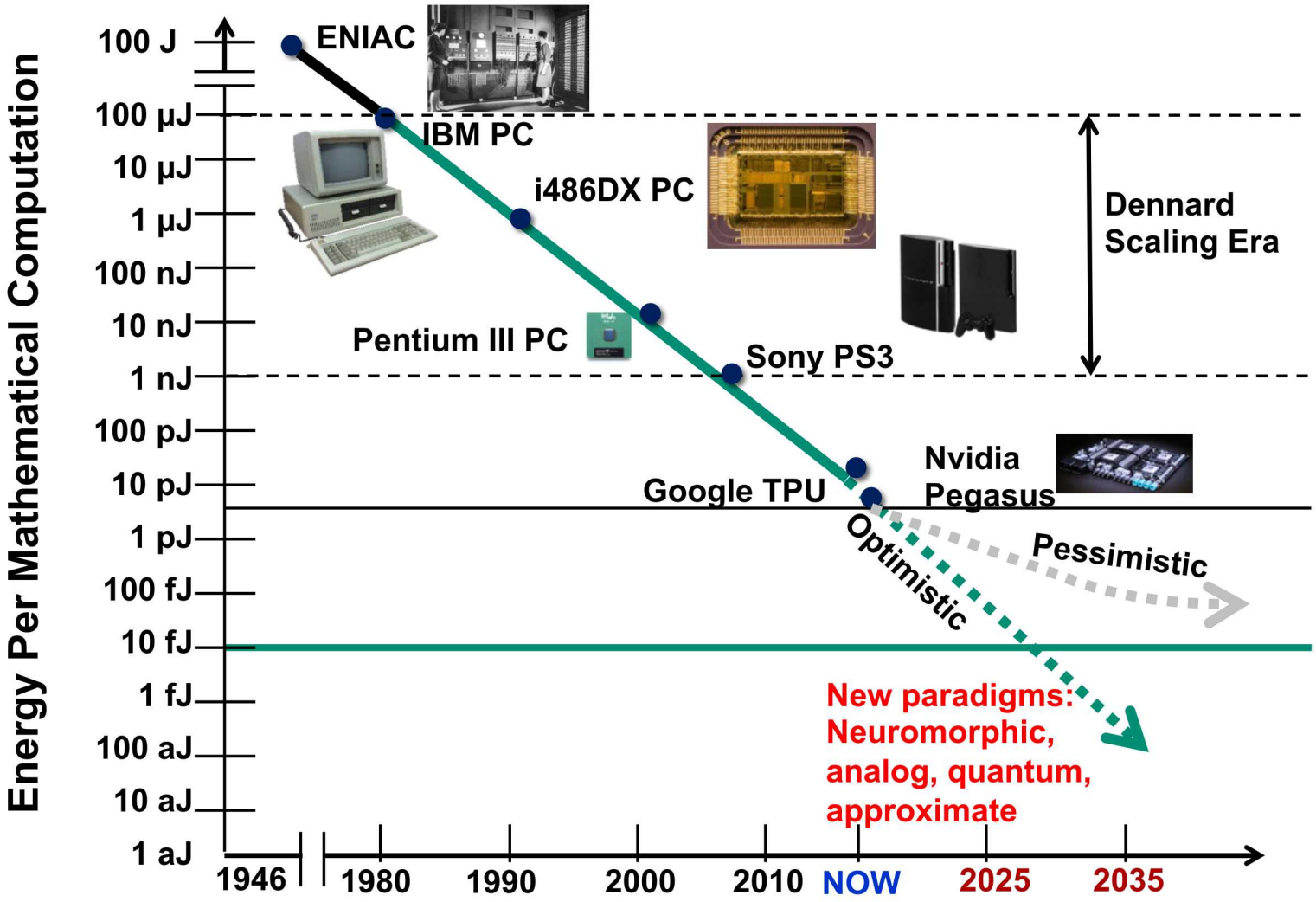
Summary

- In-memory analog computing can be much more energy efficient than digital computing, but there exists significant materials challenges.
- Redox transistors based on electrochemical ion insertion fulfill the energy and accuracy needs for in-memory computing.
- Super-capacitative redox transistors attain sufficient speeds for most neural applications
- Crossbar arrays enable parallel weight updates at low energy.





Evolution of Computing Machinery



Just as artificial intelligence can be used to advance materials science, materials science can be used to advance artificial intelligence

29

Acknowledgements

Sandia

Alec Talin

Elliot Fuller

Zhiyong Li

Sapan Agarwal

Erik Isele

Univ. of Mass.

J. Joshua Yang

Shiva Asapu

Zhongrui Wang

Stanford

Alberto Salleo

Scott Keene

Armantas Melianas

This work is supported by the Sandia National Laboratories Truman Fellowship Program, which is funded by the Laboratory Directed Research and Development (LDRD) Program. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.