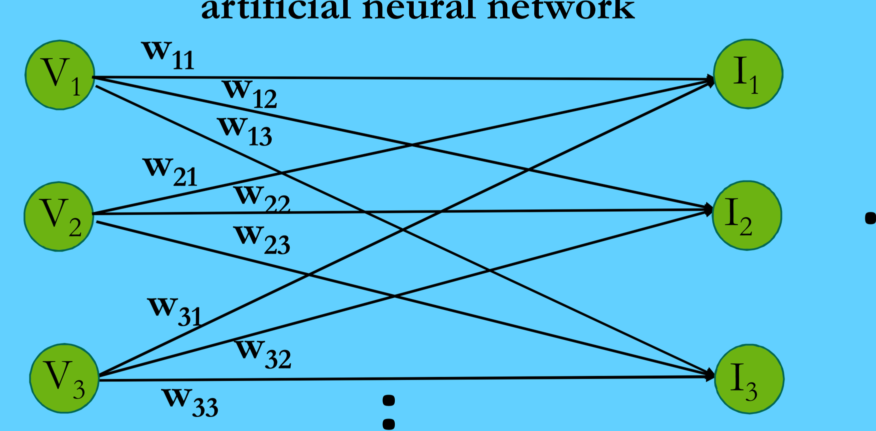
SAND2018-1825C

Ionic floating-gate memory for neuromorphic computing

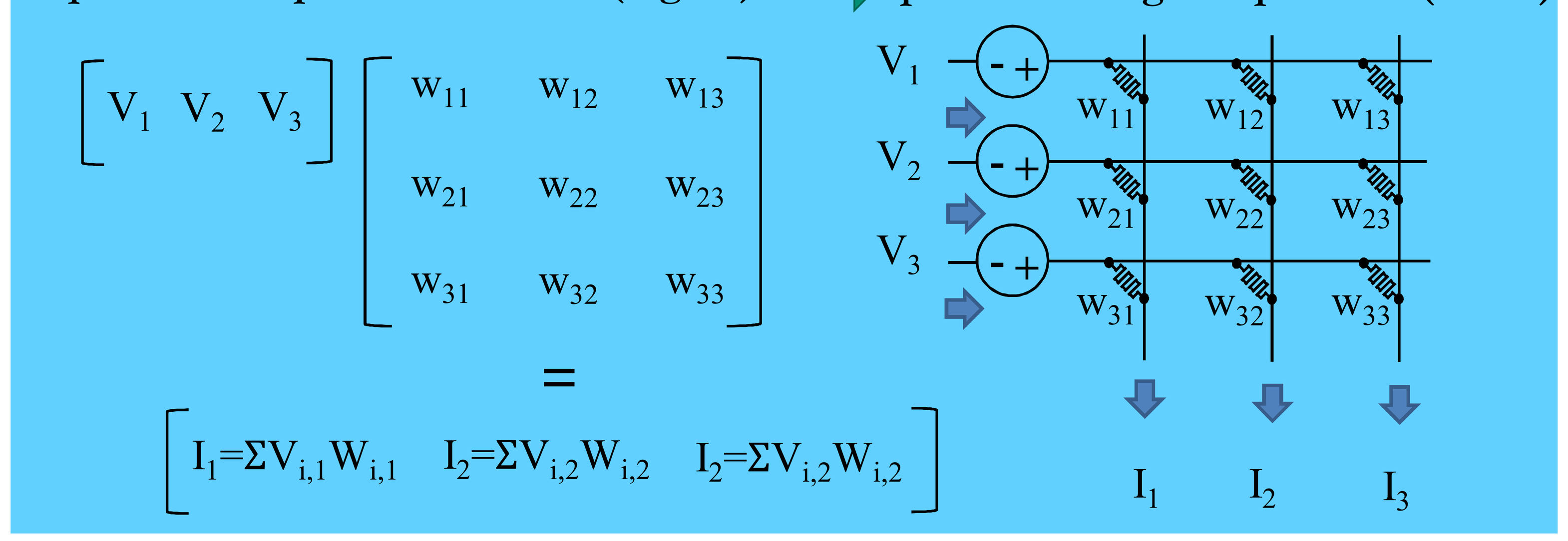
Elliot J. Fuller[1], Scott T. Keene[2], Zhongrui Wang[3], Sapan Agarwal[1], Robin B. Jacobs-Gedrim[4], John Niroula[4], Chris Bayley[1], Umar Sohi[1], Yaakov Tuchman[2], Armantas Melianas[2], Conrad D. James[4], Matthew J. Marinella[4], J. Joshua Yang[3], Alberto Salleo[2] and A. Alec Talin[1]* Affiliations: [1]Sandia National Laboratories, Livermore, CA; [2]Department of Materials Science and Engineering, Stanford University, Stanford, CA; [3]Department of Computer Science and Electrical Engineering, University of Massachusetts Amherst, Amherst, MA; USA [4]Sandia National Laboratories, Albuquerque, NM

Neuromorphic computers can overcome efficiency bottlenecks inherent to digital computers by using analog memory to both process and store synaptic weights.

artificial neural network

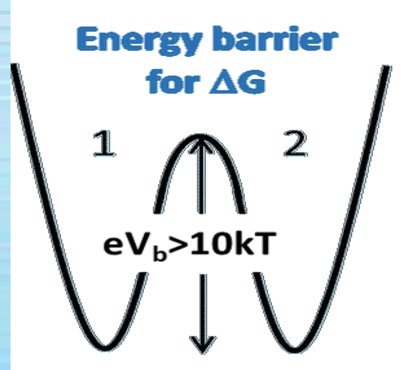


sequential computation of V^TW (digital) → **parallel analog computation (neuro)**



Resistive random access memory (RRAM) that is proposed for neuromorphic computers suffers from large voltages and currents that prevent scaling to large arrays.

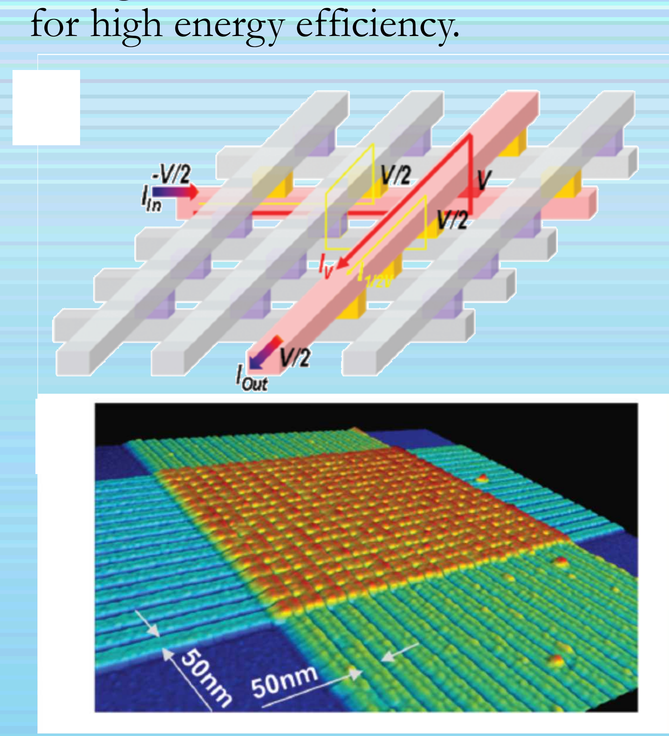
RRAM time-voltage dilemma: For two-terminal devices, a large energy barrier between conductance states is needed for retention but leads to large voltages and currents.



Schroeder et. al. *Journal of Applied Physics* (2010)

Large voltages an currents lead to parasitic losses within an array that inhibit programming.

Parasitic losses limit working neuromorphic arrays to ~100×100. Arrays > 1000×1000 are needed to average out cost of circuit overheads for high energy efficiency.



Yang et al., Applied Physics Letters, 100, 113501 (2012)

Ionic floating-gate memory (IFG) can meet important requirements for a neuromorphic computer with 10^5 energy efficiency gain over existing CPUs and 10^3 gain over GPUs.

- low voltages (<700mV) and low currents (<5nA) for scaling to arrays > 1000×1000
- linear programmability required for massively parallel "blind" updates (outer-product update)
- array level selectivity and retention

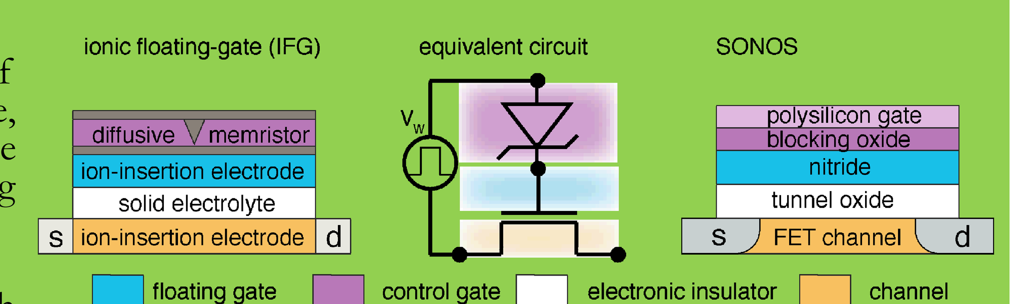
A redox transistor composed of ion-insertion electrodes (blue, orange) and a solid electrolyte (white) is used to store the analog states

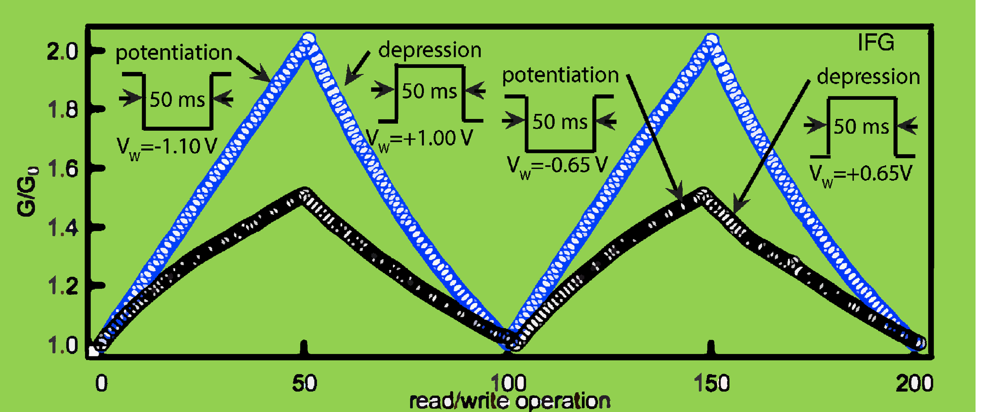
A threshold device (purple), such as a diffusive memristor is used as a control gate.

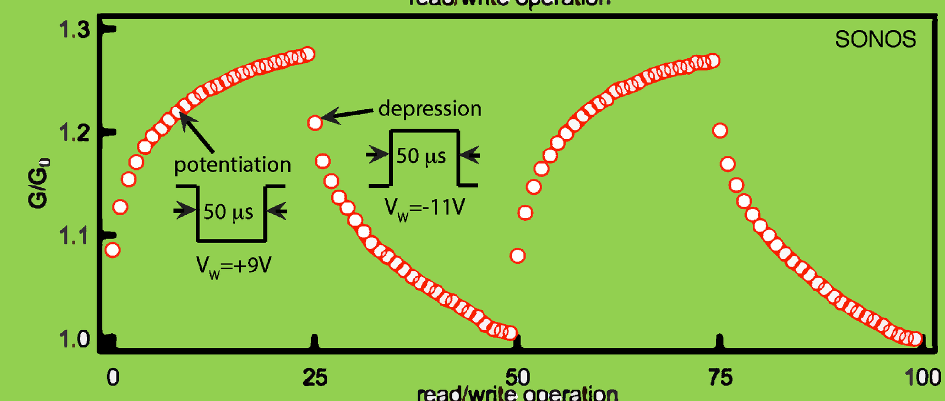
Operational similarity to CMOS floating gate memory (e.g. SONOS, FLASH) but with an order of magnitude lower programming voltage.

RRAM time-voltage dilemma is circumvented with a third terminal used for programming and memory read out from source and drain terminals. This allows low voltages and currents with long retention times.

Linear Tuning: IFG has linear programmable levels while SONOS and RRAM do not.

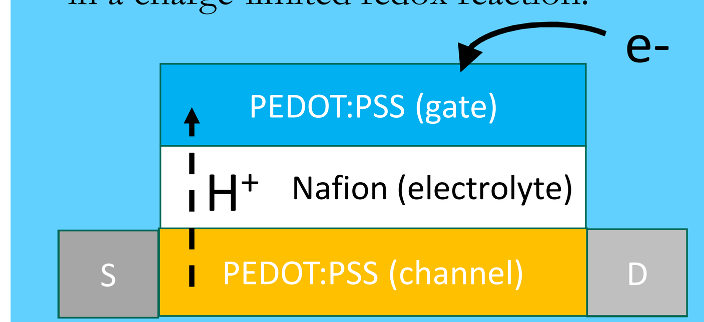






IFG uses entirely ion-based switching mechanisms in order to achieve low voltages and high performance.

Redox transistor to store weights: In order to switch the channel resistance, charge is needed to drive ion exchange between the gate and channel in a charge-limited redox reaction.



PEDOT:PSS is a redox-active polymer blend that undergoes a charge compensation reaction after proton insertion that generates hole carriers.

Nafion is a solid electrolyte with a high proton diffusivity.

Low voltages (<100mV) can switch a redox transistor state. Floating gate accrues minimal voltage (<50mV) due to high accumulation capacitance.

Large arrays >1000 × 1000 can be realized through doping PEDOT:PSS channels to >100 MΩ.

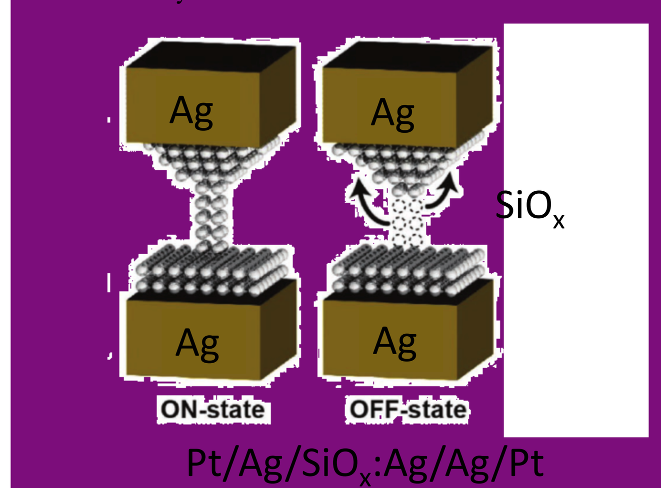
Leakage requirements are relaxed due to 10^3 higher charge density per analog level for ion-insertion electrodes compared to an FET channel.

Material versatility: redox transistors can be made with transition metal oxides and ceramic electrolytes

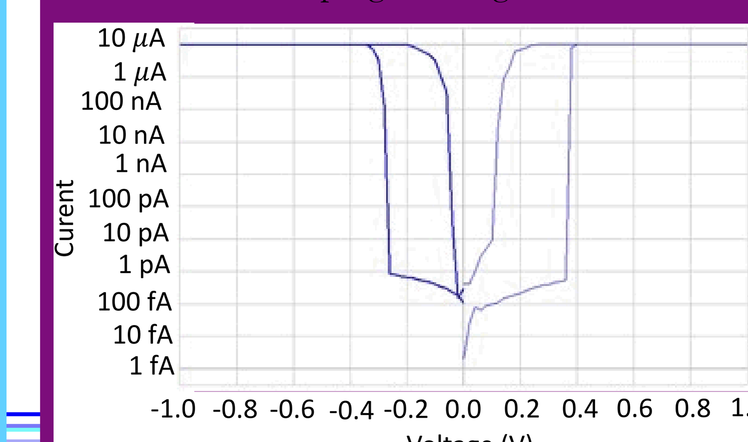
van de Burgt et. al., *Nature Materials*, 16(4), 2017

Fuller et. al., *Advanced Materials*, 29(4), 2017

Diffusive memristor as control gate: This volatile threshold switch is normally in the OFF state to prevent charge from entering the redox transistor gate in order to preserve the memory state.

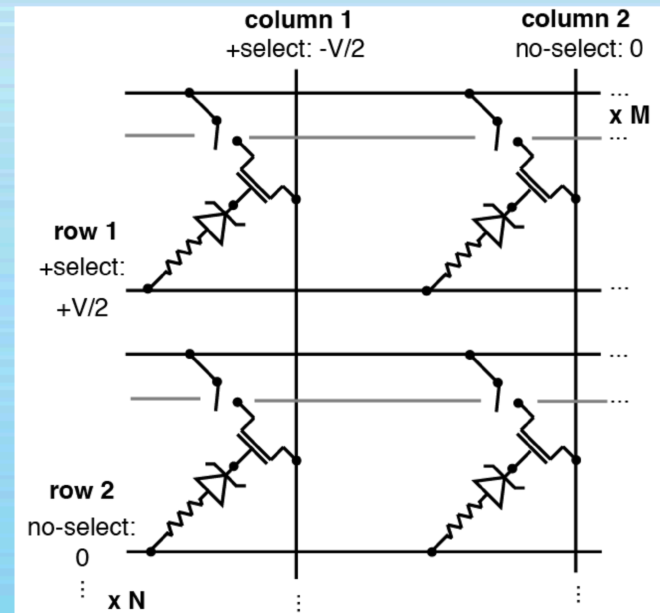


Once a small voltage threshold (±400 mV) is reached, the memristor turns to the ON state allowing the redox transistor to be programmed. A large ON/OFF ratio ensures long retention times. A fast filament dissolution time (150 ns) can support fast redox transistor programming times.

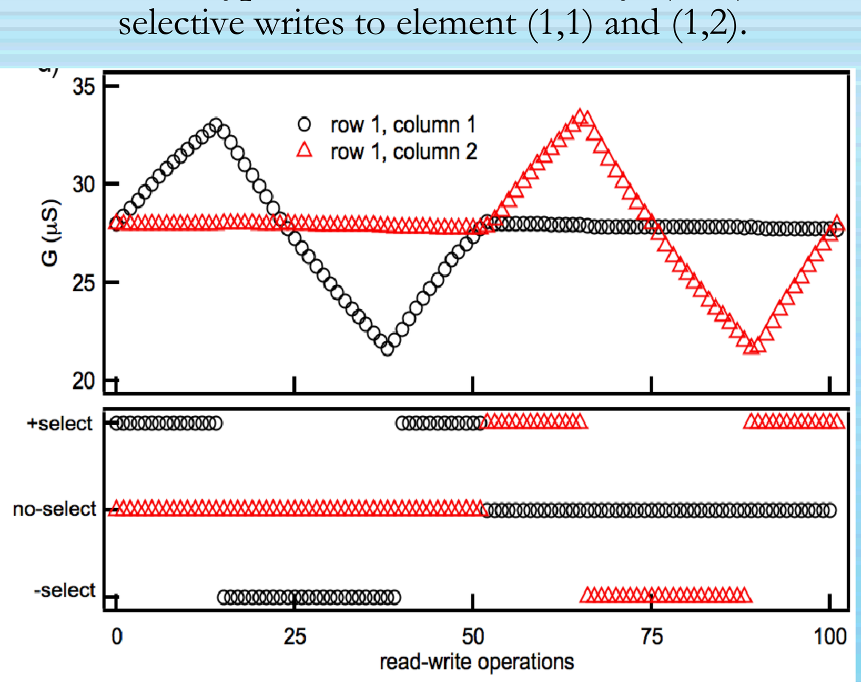


Midya et. al., *Advanced Materials*, 29(12), 2017

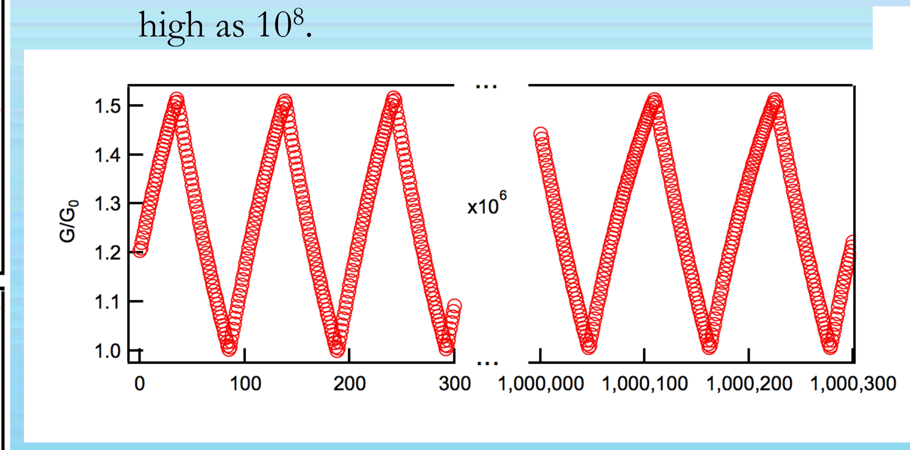
IFG crossbar circuit-level schematic is shown with outer-product operation.



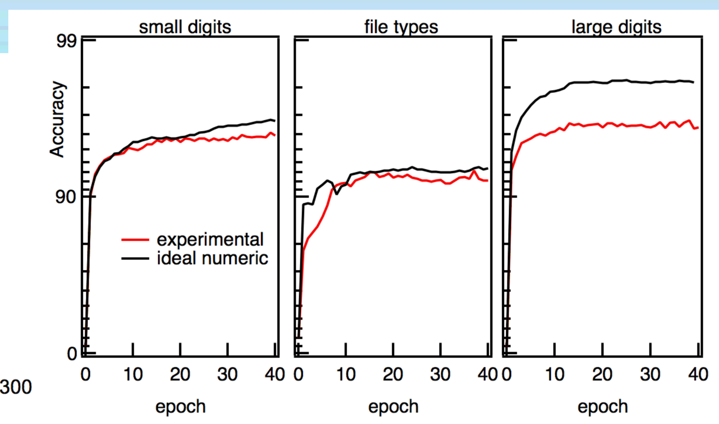
Prototype IFG crossbar array (1×2) with selective writes to element (1,1) and (1,2).



Endurance of a redox transistor reaches 10^6 write operations without degradation. Diffusive memristors have been demonstrated with endurance as high as 10^8.



Neural network simulation of IFG crossbar achieves ideal accuracy in backpropagation.



Future work: faster switching speeds >1 MHz, endurance 10^{13}, and CMOS integration must be demonstrated.