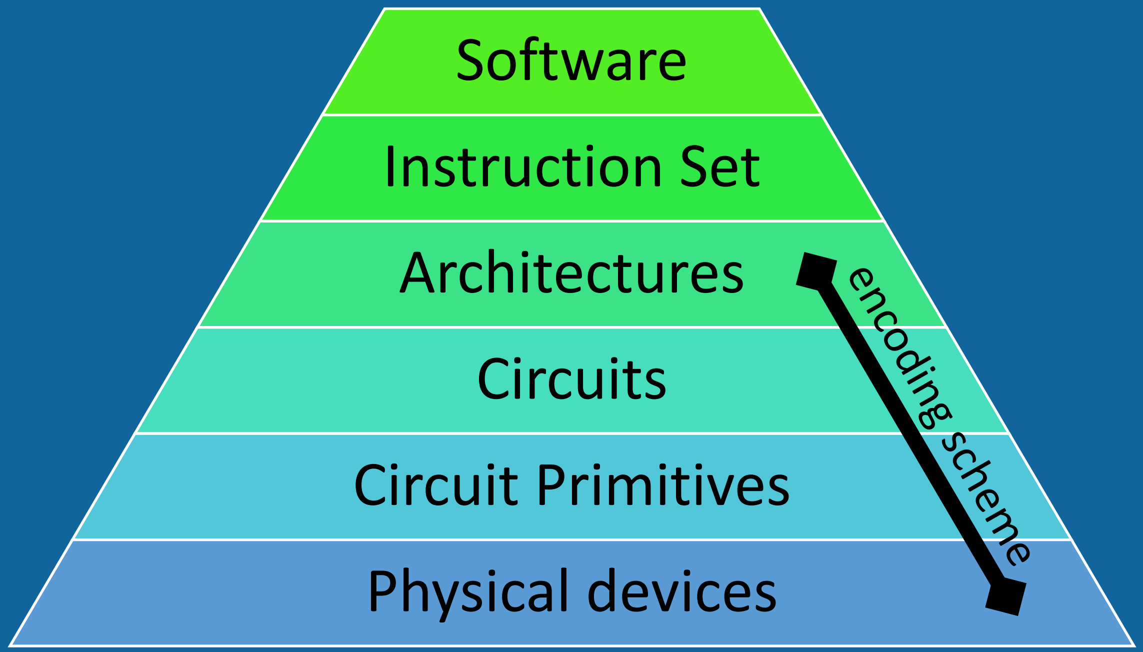
Unconventional computing approaches using time, stochasticity, and physics

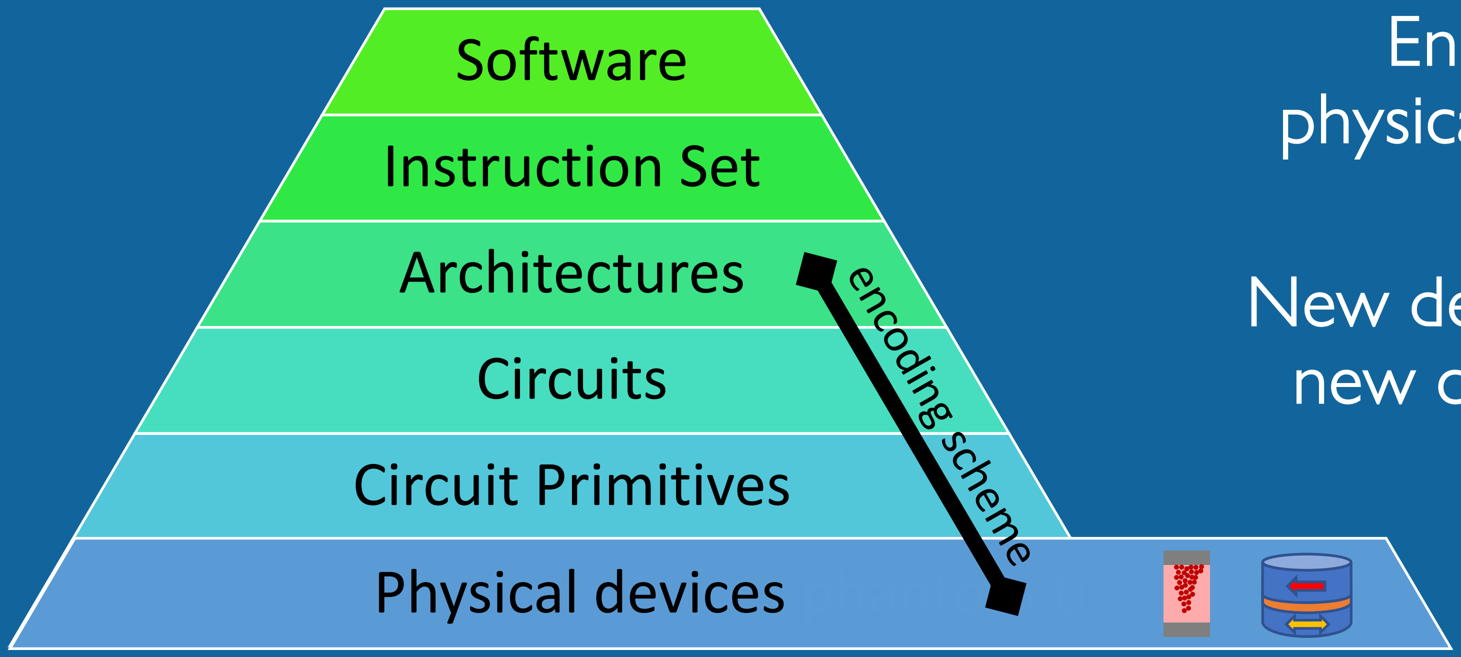


Encoding schemes map physical state to data types

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Matthew W. Daniels
Alternative Computing Group
Physical Measurement Laboratory

Unconventional computing approaches using time, stochasticity, and physics



New devices integrated BEOL : CMOS+X

Encoding schemes map physical state to data types

New device physics leads to new circuits, architectures, and features

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Matthew W. Daniels
Alternative Computing Group
Physical Measurement Laboratory

Collaborators

NIST PHYSICAL MEASUREMENT LABORATORY

UNIVERSITY OF MARYLAND 1856

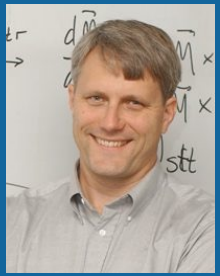
THE GEORGE WASHINGTON UNIVERSITY WASHINGTON, DC

Western Digital

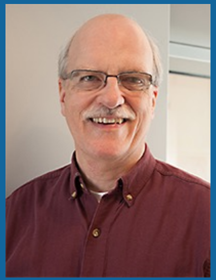
spintec SPIN IN ELECTRONICS

Brian Hoskins

William “Drew” Borders



Mark Stiles



Jabez McClelland

Advait Madhavan

Sidra Gibeault

Liam Pocher

Temitayo Adeyeye

Gina Adam

Osama Yousuf

Patrick Braganca

Martin Lueker-Boden

Philippe Talatchian

Lucile Soumah

Part I: Temporal Computing

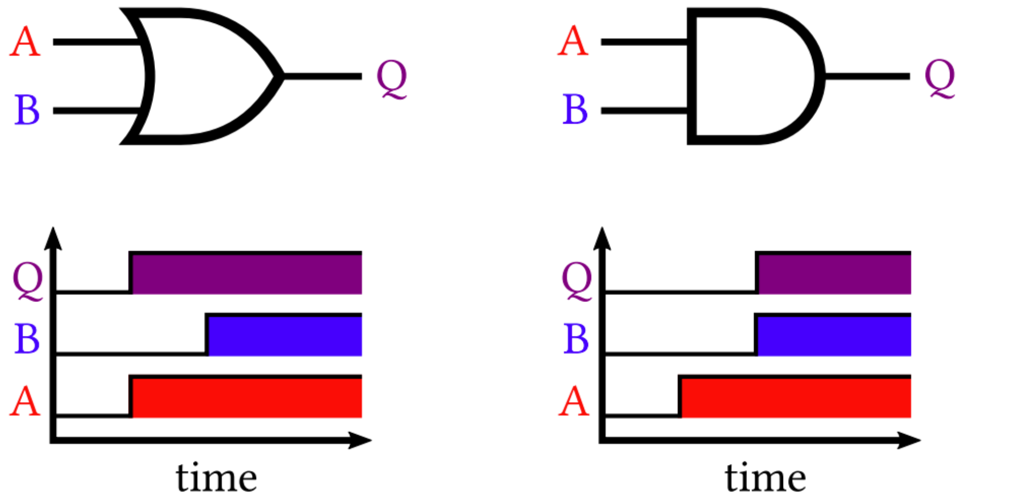
Computing with the arrival times of rising edges

Logic gates have natural temporal interpretations.

Let a wire that switched from low to high voltage at time *t* encode the value *t*.

Circuit primitive	Temporal operation
logical AND	max
logical OR	min
RC circuit delay	+ *RC*

$$A \vee B \iff \min(t_A, t_B) \qquad A \wedge B \iff \max(t_A, t_B)$$



Mathematicians have long studied max-plus/min-plus algebras.

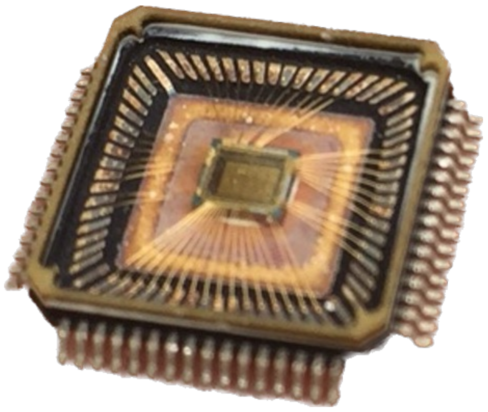
Max-plus (tropical) semiring:

$$a \oplus b \coloneqq \max(a, b)$$

$$a \otimes b \coloneqq a + b$$

✓ Associativity
✓ Commutativity
✓ Distributivity

✓ Additive identity
✓ Multiplicative identity
✗ Additive inverses
✓ Multiplicative inverses

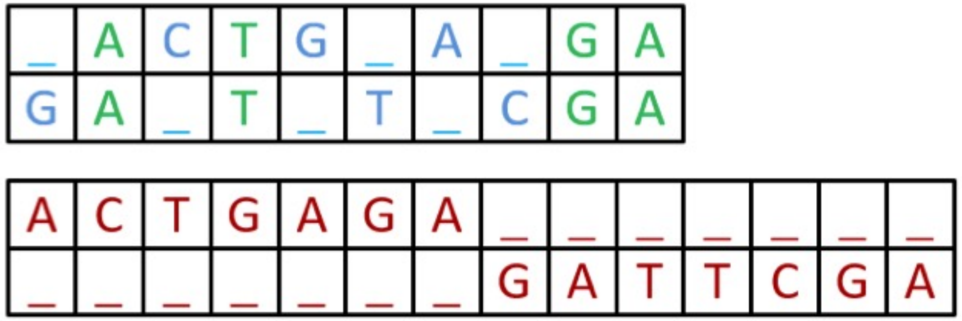


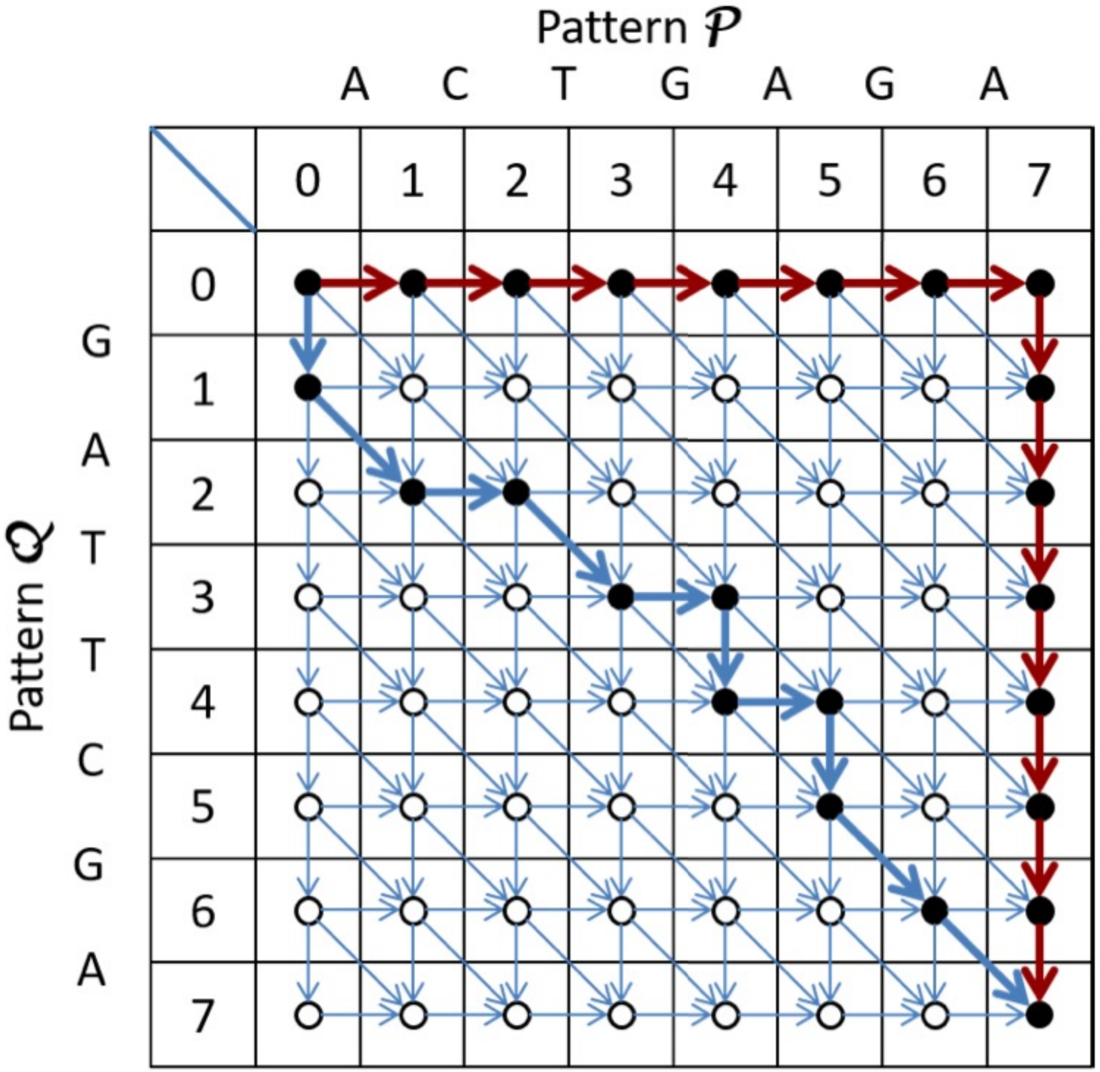
DNA sequence alignment chip (2017):
Madhavan et al., 2017 IEEE Custom Integrated Circuits Conference (CICC), 1-4. IEEE.

Theory of temporal state machines (2021): Madhavan, Daniels, Stiles, ACM Journal on Emerging Technologies in Computing Systems (JETC), 17 (3), 1-27.

Certain problems can be solved by “races” in graphs.

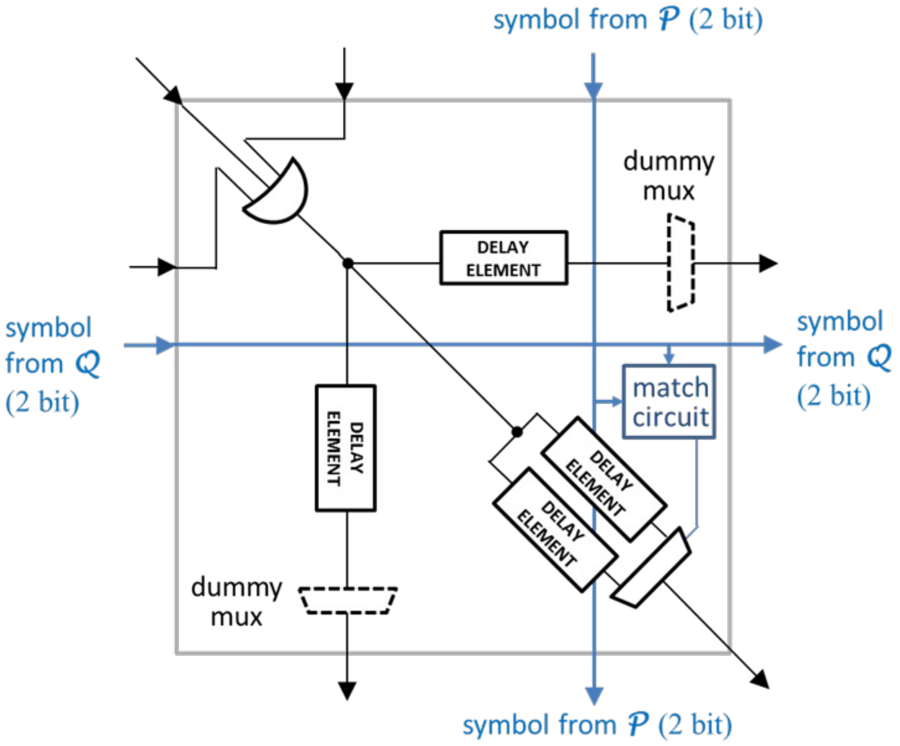
Early work on race logic showed that problems with optimal substructure (i.e. dynamic programming problems) map neatly into temporal computing.





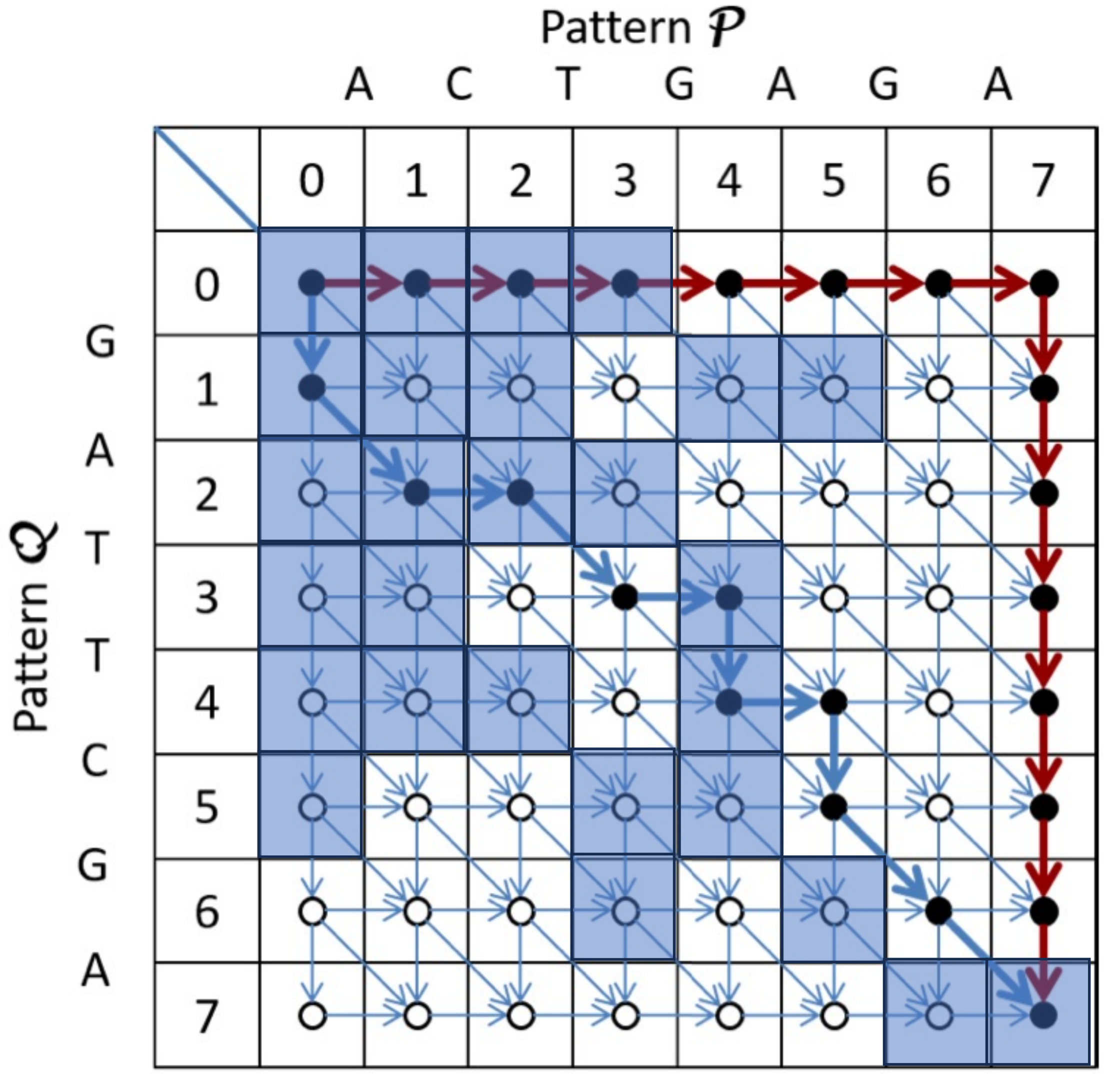
Madhavan et al., 2017 IEEE Custom Integrated Circuits Conference (CICC), 1-4. IEEE.

Certain problems can be solved by “races” in graphs.



_	A	C	T	G	_	A	_	G	A
G	A	_	T	_	T	_	C	G	A

A	C	T	G	A	G	A	_	_	_	_	_	_	_
_	_	_	_	_	_	_	G	A	T	T	C	G	A



Madhavan et al., 2017 IEEE Custom Integrated Circuits Conference (CICC), 1-4. IEEE.

How can we generalize this?

- Building graph problems directly into CMOS can be difficult for non-planar graphs
- We don’t want to hardcode every single problem into an ASIC.
- We would like to be free of certain physical constraints.

Physics restricts what a single temporal "race" can compute.

Restrictions on pure race logic:

- Causality: if the output of a function depends on one of the inputs, then the output's time must be later than that input's time

- Time-translational invariance: the function must behave the same regardless of when it's used

$$f(t_1 + \delta, \ldots, t_n + \delta) = f(t_1, \ldots, t_n) + \delta$$

$$f(\delta \otimes t_1, \ldots, \delta \otimes t_n) = \delta \otimes f(t_1, \ldots, t_n)$$

You can't even add temporal signals together!

- Time-translational invariance: the function must behave the same regardless of when it's used

$$f(t_1 + \delta, \ldots, t_n + \delta) = f(t_1, \ldots, t_n) + \delta$$

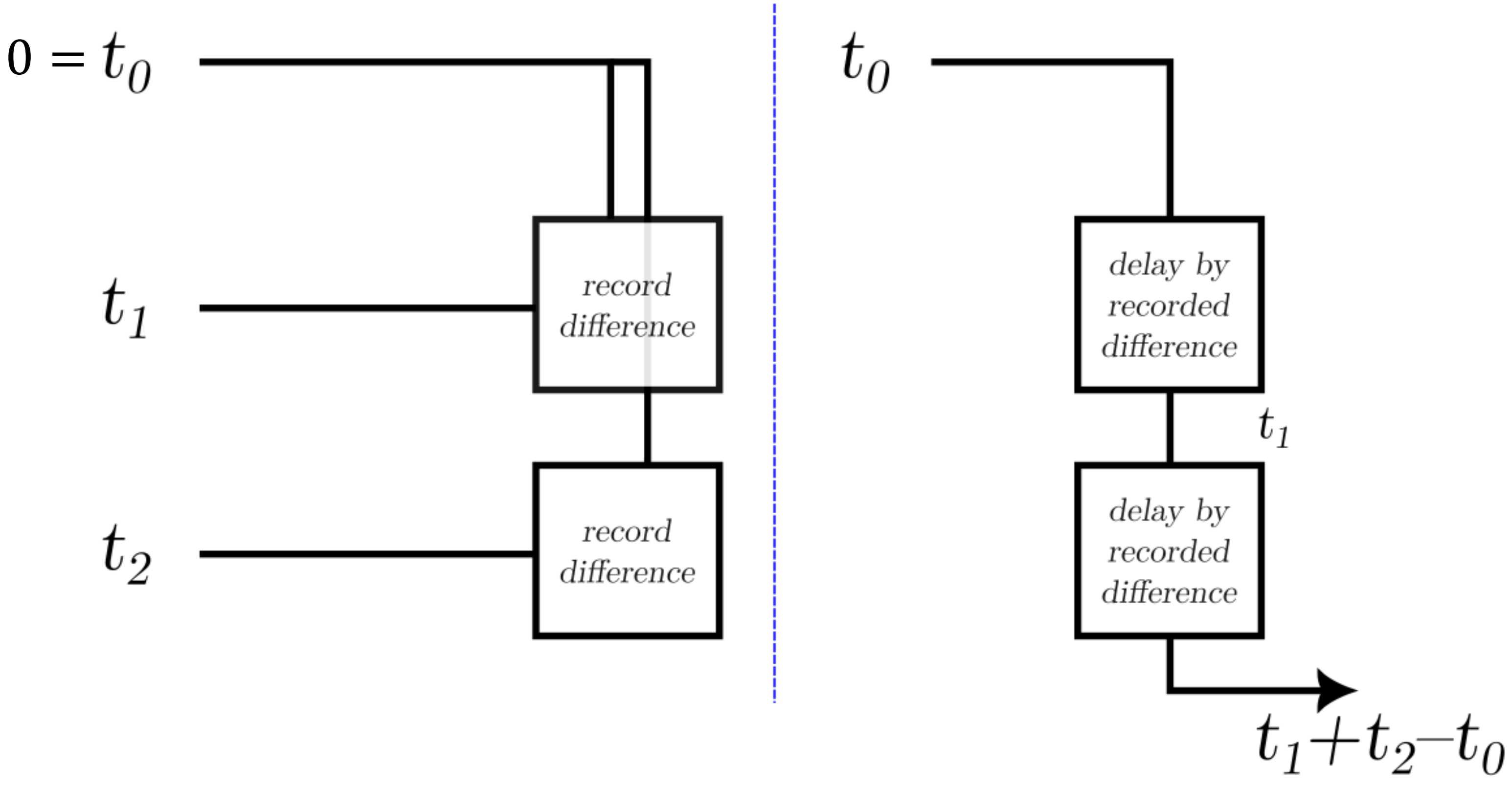
- For example, cannot compute

$$f(t_1, t_2) = t_1 + t_2.$$

We can recover addition and other ops via state machines.

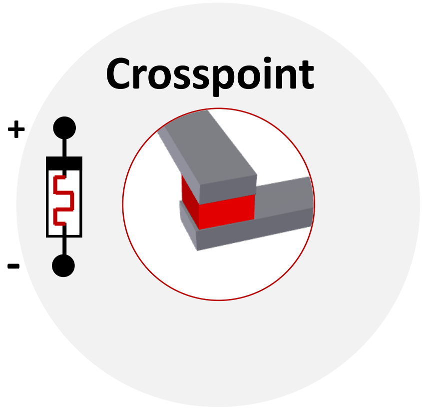
How can we break invariance?

1. Mathematically, by designating a privileged point in time (e.g. the temporal origin)
2. Physically, by introducing a changeable, hidden state

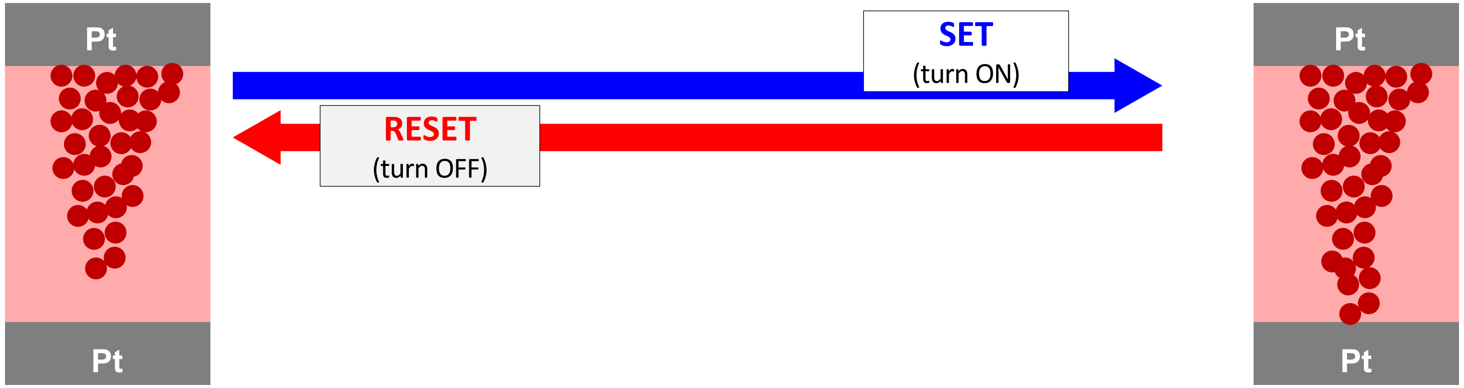


$= t_1 + t_2$

Memrisitve devices can save times as resistances.



- Two-terminal metal / oxide / metal structure
- Filamentary switching (physics not yet fully understood!)

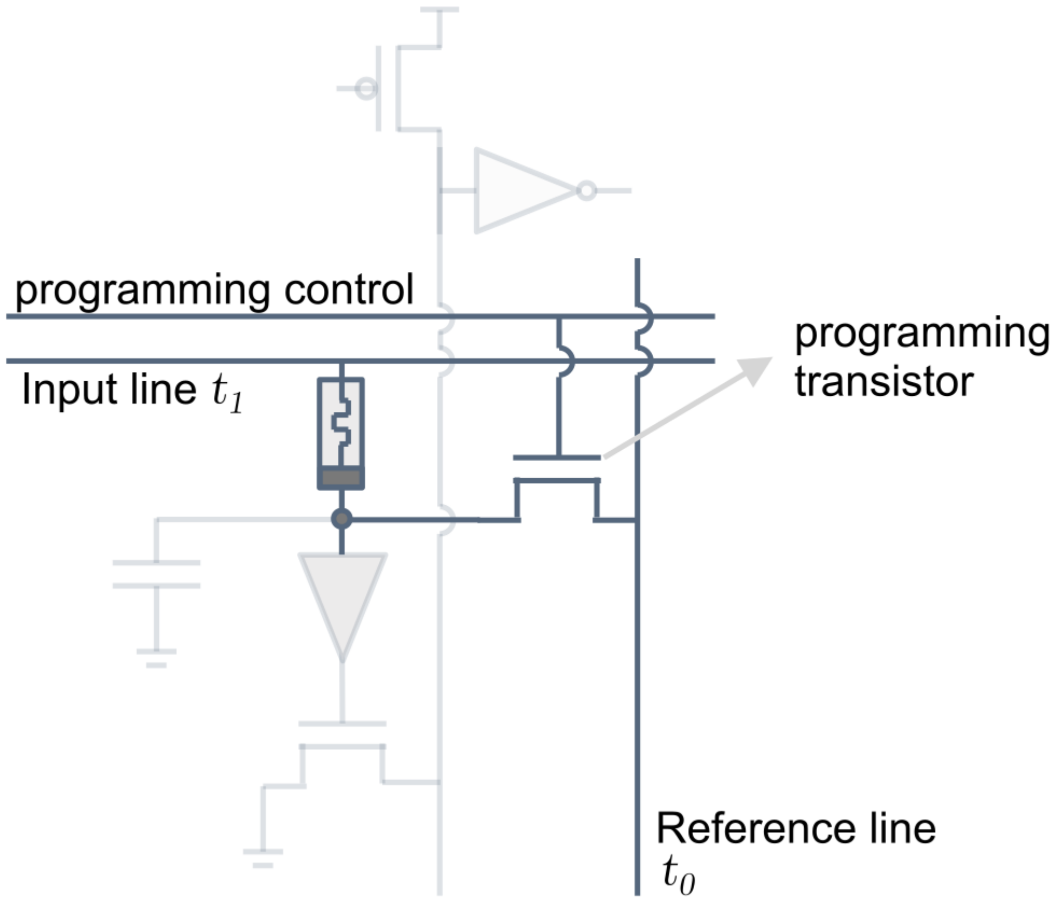


Filament shape is retained in absence of applied voltage

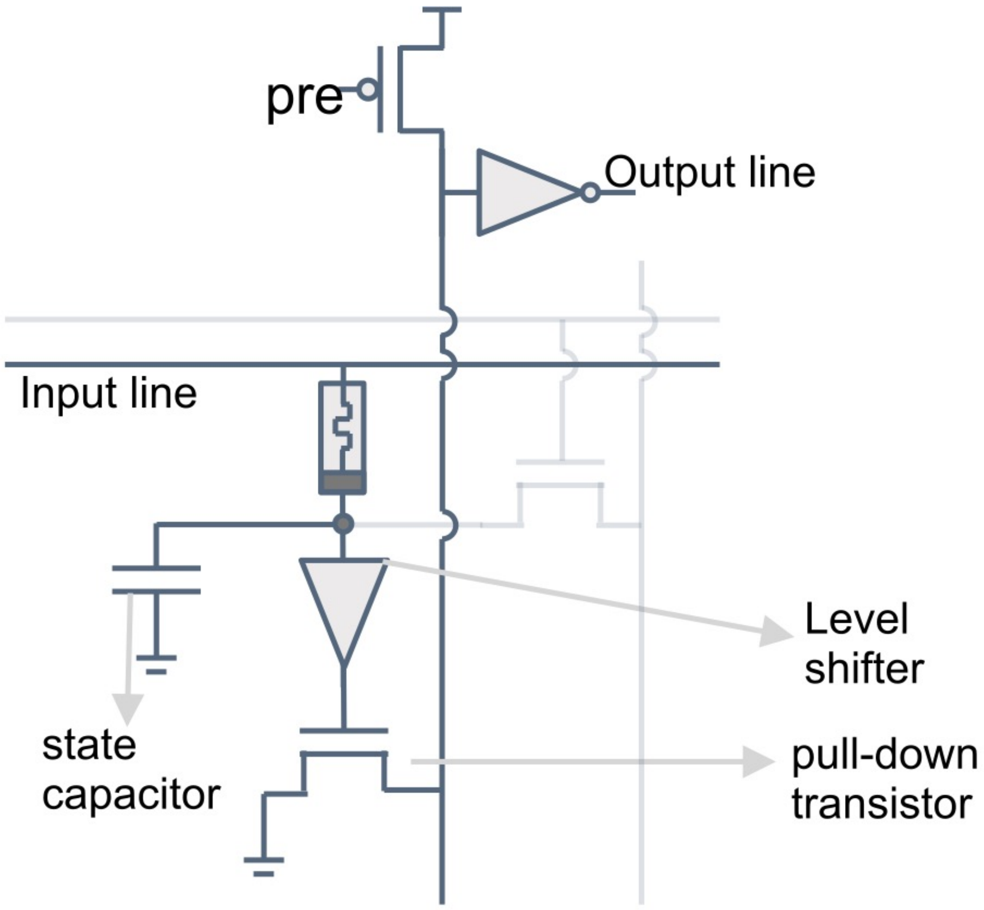
Adapted from *Waser R. et al (Nanoelectronics & Information Technology., 2012)*

A single cell can be used to read and write temporal signals from memristor devices.

Write mode



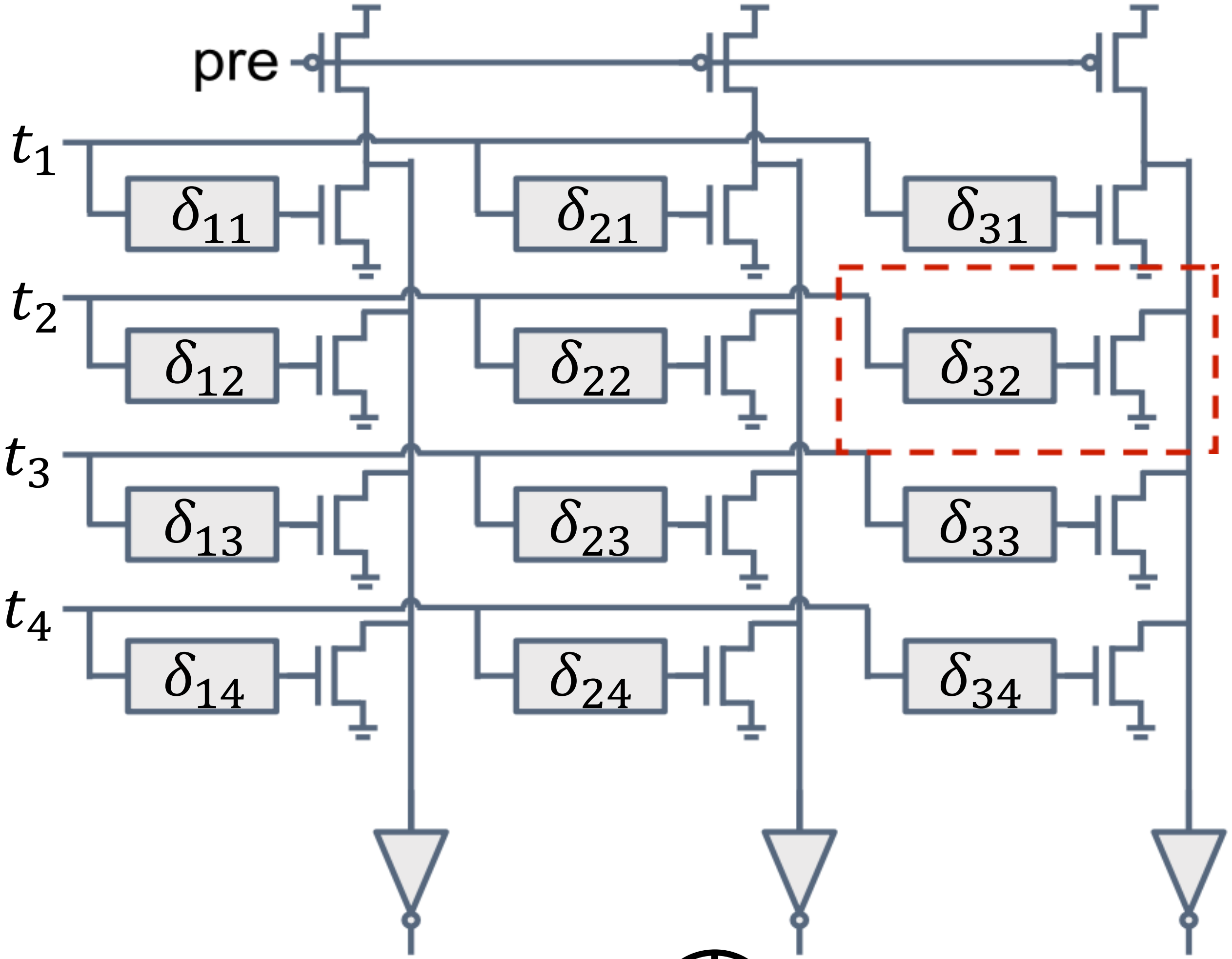
Read mode



A naturally parallel operation emerges from a crossbar memory of such delay elements.

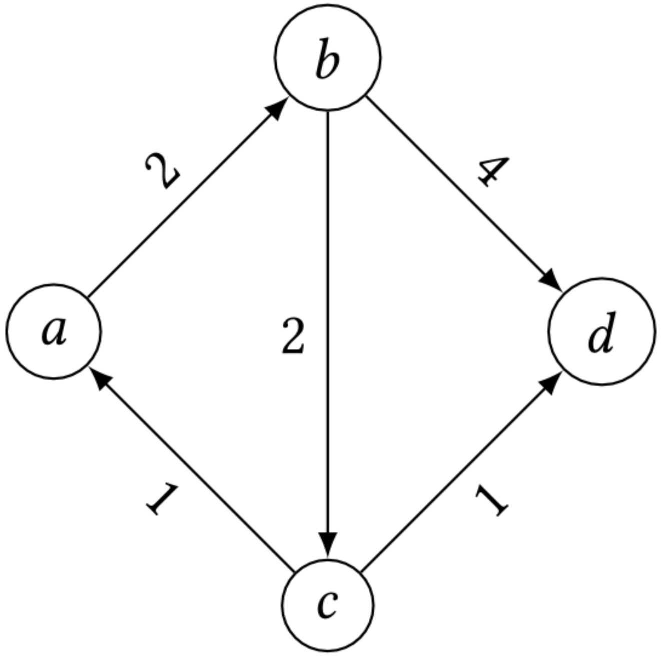
Tiling this cell gives rise to a natural ***temporal analog of vector-matrix multiplication***.

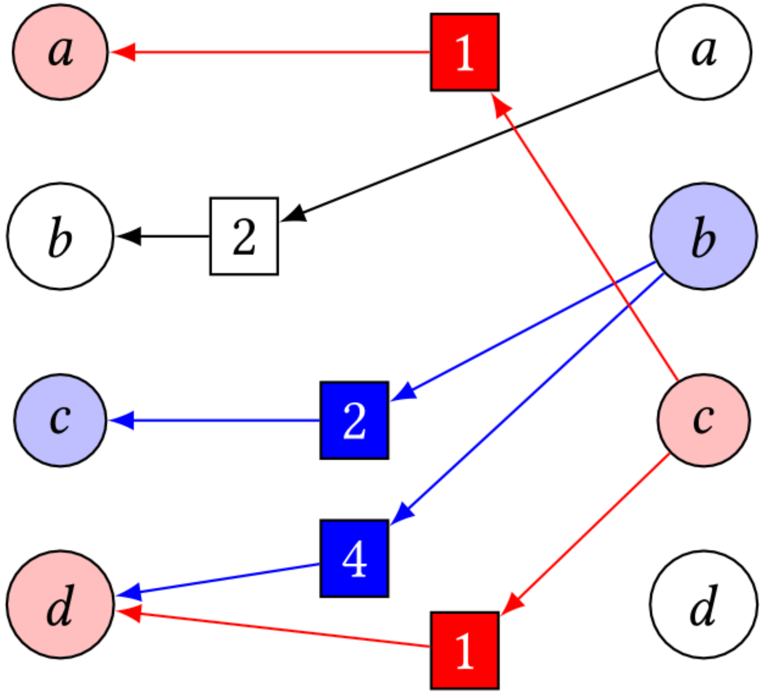
This "tropical" min-plus version of linear algebra naturally expresses ***graph traversal***.



$$t_i^* = \min_j(\delta_{ij} + t_j) = \bigoplus_j (\delta_{ij} \otimes t_j)$$

Tropical matrix multiplication builds the minimal spanning tree in a graph.





$$\begin{bmatrix} \infty & \infty & 1 & \infty \\ 2 & \infty & \infty & \infty \\ \infty & 2 & \infty & \infty \\ \infty & 4 & 1 & \infty \end{bmatrix}$$

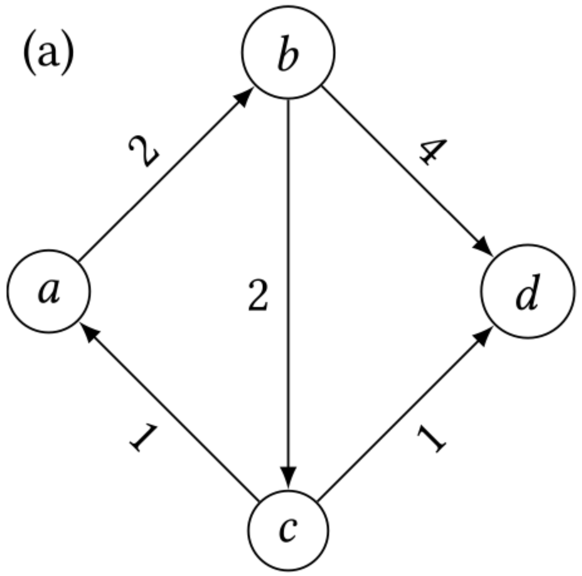
$$\begin{bmatrix} 1 \\ \infty \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \infty & \infty & 1 & \infty \\ 2 & \infty & \infty & \infty \\ \infty & 2 & \infty & \infty \\ \infty & 4 & 1 & \infty \end{bmatrix} \begin{bmatrix} \infty \\ 0 \\ 0 \\ \infty \end{bmatrix}$$

ALGORITHM 3: Pseudocode for Temporal Dijkstra's Algorithm

```
Input: graph G, source node s
// Variable initializations
d⃗ := 0_s;        // distances to unvisited nodes (tropical one-hot labels source)
v⃗ := ∞;                                        // visited nodes (tropical zero vector)
P̂ := ∞;                                        // parent matrix (tropical zero matrix)
Â := adjacency-matrix(G);                          // adjacency matrix of the graph
while (⊕_j d_j < ∞) do
    n⃗ := argmin(d⃗);                                      // choose node to visit
    // Examine neighbors
    e⃗ := Â ⊗ n⃗;                              // VMM examine neighbors of current node
    f⃗ := d⃗ ⊣ e⃗;                               // keep only newly found shortest paths
    // Update records for the next iteration
    v⃗ := v⃗ ⊕ n⃗;                                // record the current node as visited
    d⃗′ := d⃗ ⊕ f⃗;                             // construct new record of shortest paths
    d⃗ :≅ v⃗ ⊣ d⃗′;                            // update global unvisited distance vector
    // Parent vector update process
    f⃗* :≅ binarize(f⃗);                                 // vector indices of found nodes
    P̂ := f⃗* ⊣ P̂;      // delete row data of previously recorded parents for found
    nodes
    P⃗_n⃗ := f⃗;          // record in column n⃗ distances f⃗ from n⃗ to the found nodes
end
return P̂;  // adjacency matrix of the minimal spanning (from s) subgraph of G
```

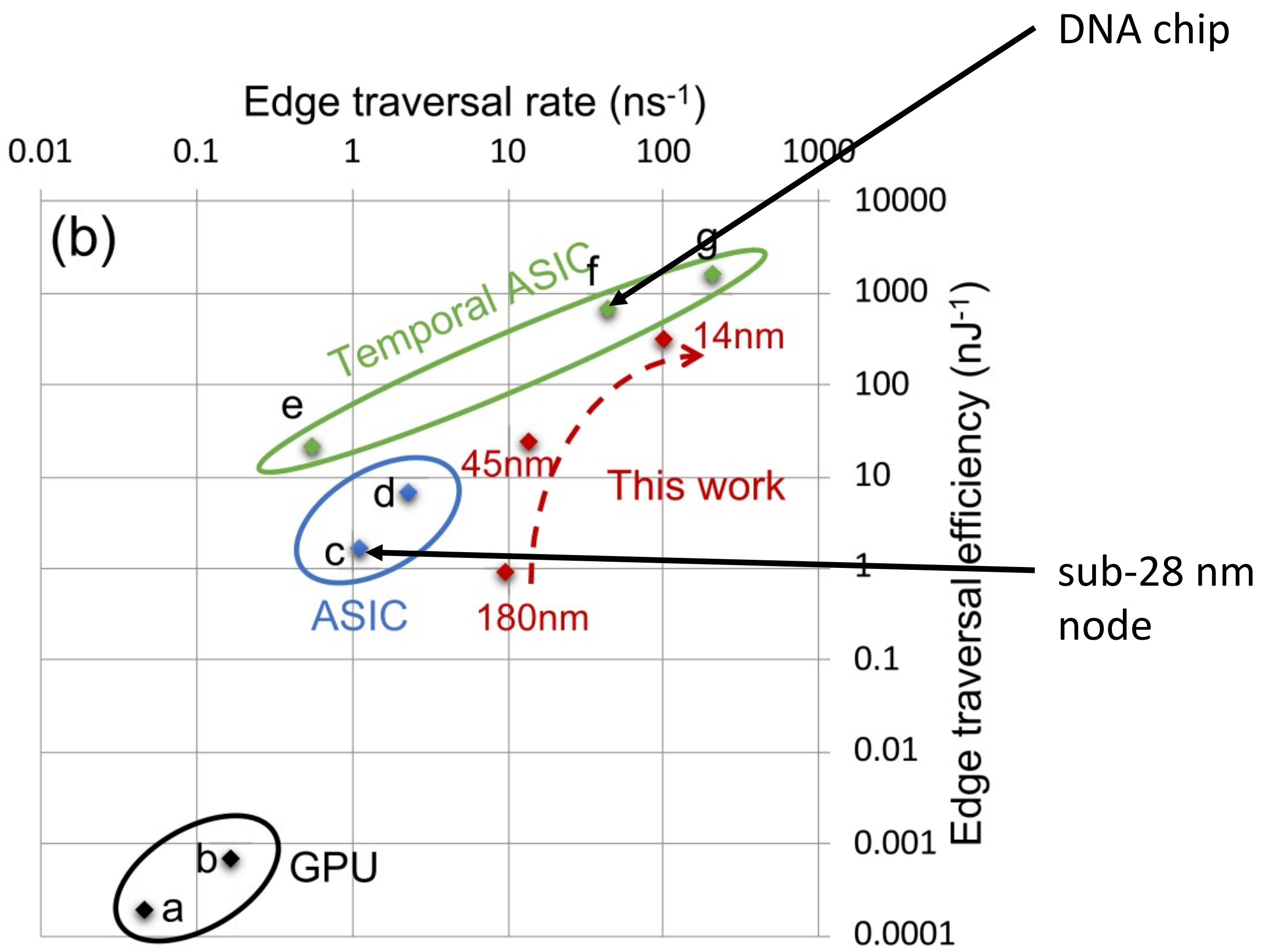
Highly parallel edge traversal

Edge traversal is almost as good as a temporal ASIC, and much better than non-temporal ASICs.



(b)

$$\begin{bmatrix} \infty & \infty & 1 & \infty \\ 2 & \infty & \infty & \infty \\ \infty & 2 & \infty & \infty \\ \infty & 4 & 1 & \infty \end{bmatrix}$$

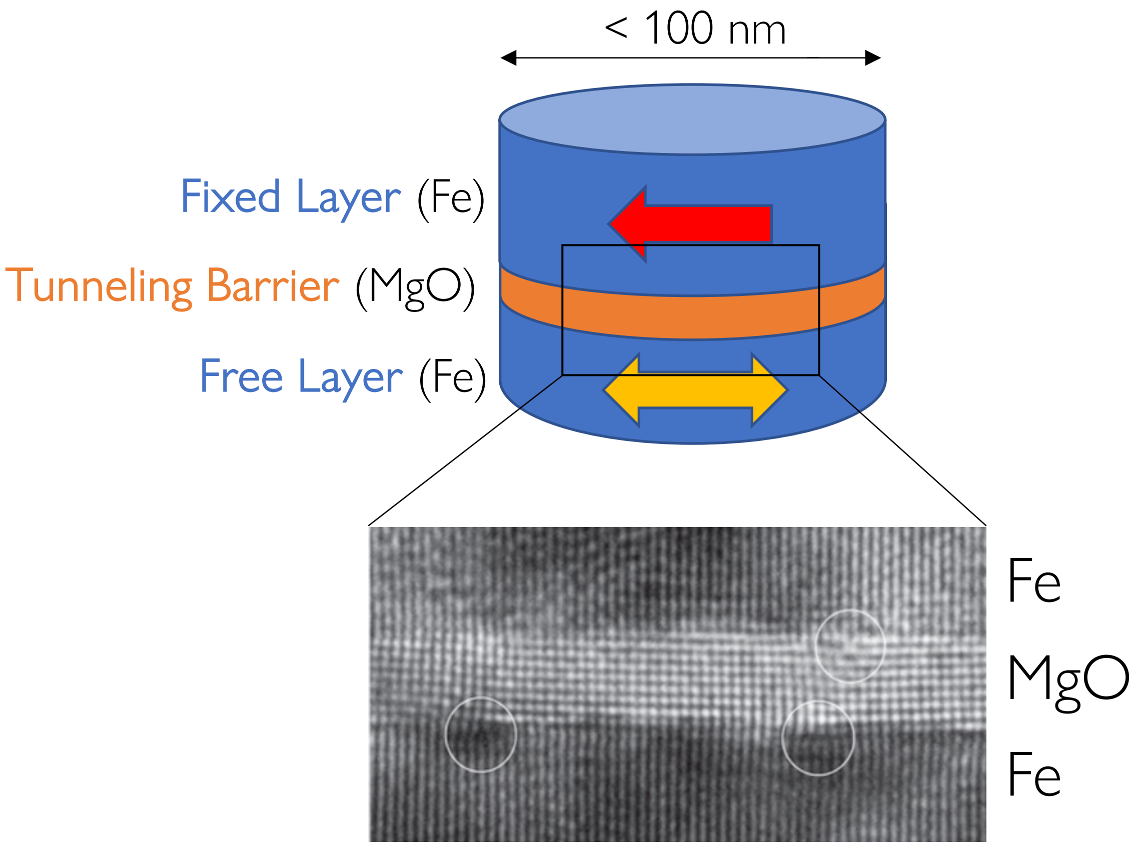


Madhavan, Daniels, and Stiles, ACM J. Emerg. Tech. in Comp. Sys. 17, 3 (2021)

II: Stochastic Computation

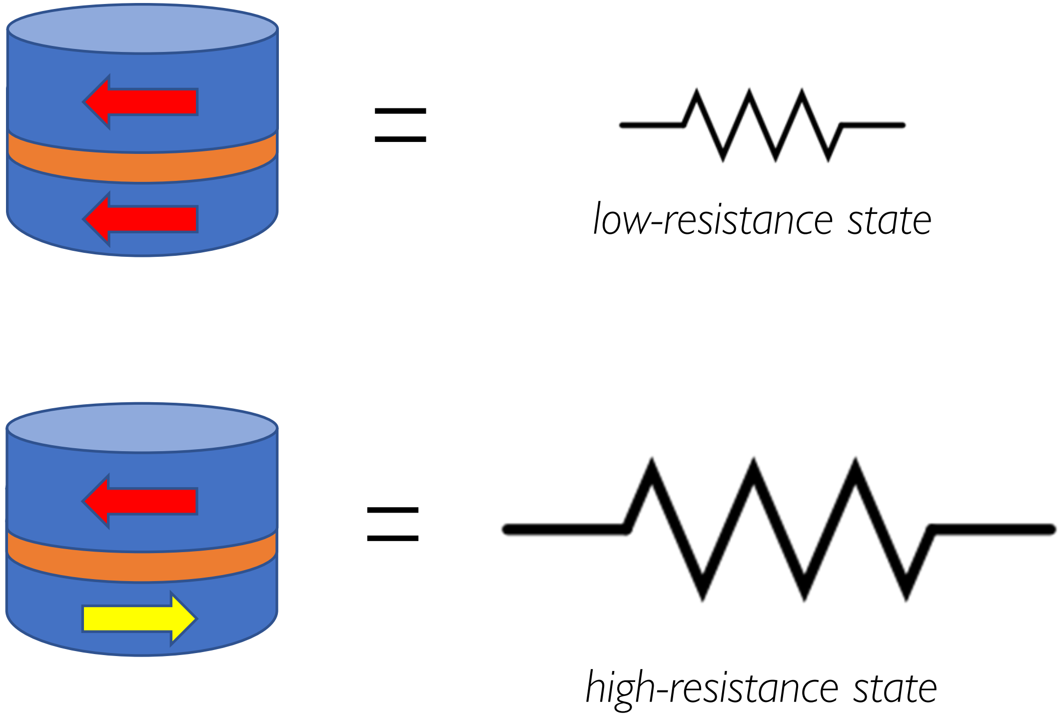
Computing using randomness

MTJs bridge magnetism and electronics.



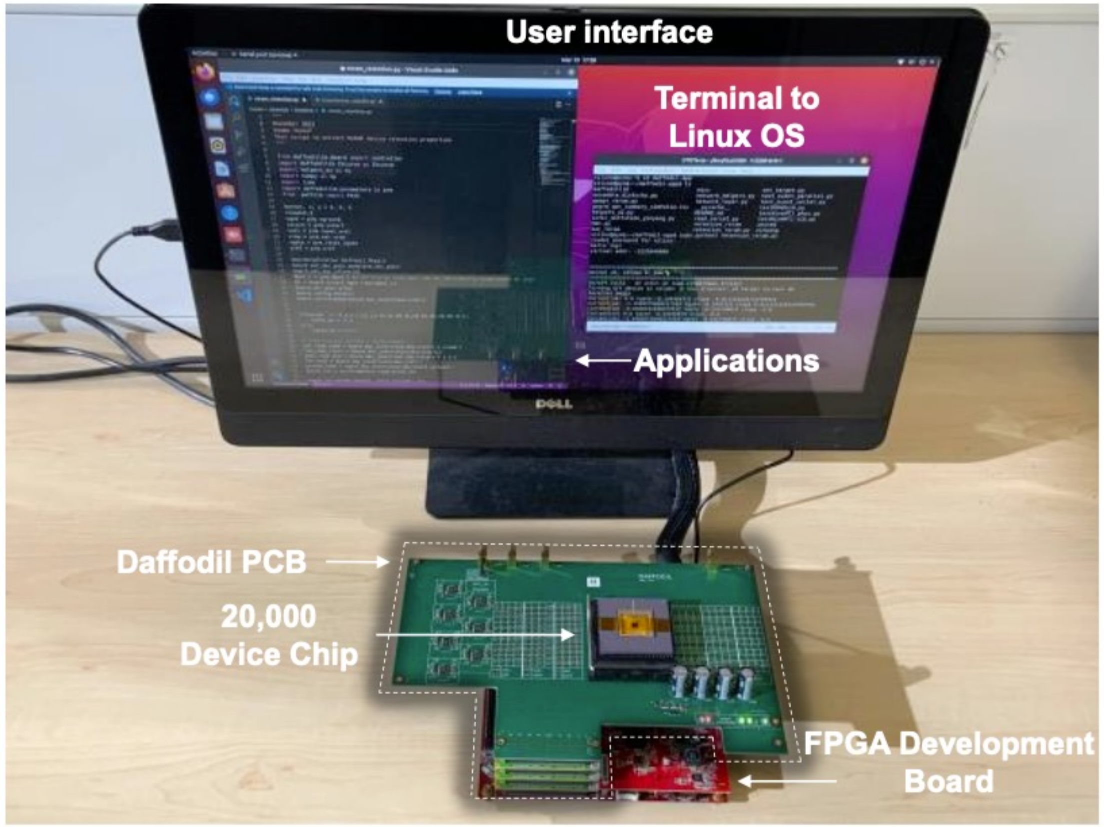
TEM image: Yuasa, Nagahama, Fukushima, Suzuki & Ando, Nat. Mater. 3, 868 (2004)

Magnetic tunnel junctions (MTJs) associate magnetic state with resistive state



20

We use MTJs (and memristors) as weights in hardware neural networks.



Integrated MTJ crossbar neural network experiment:

Borders, Madhavan, Daniels, et al., arXiv:2312.06446, Phys. Rev. Appl. (accepted; in press, 2024)

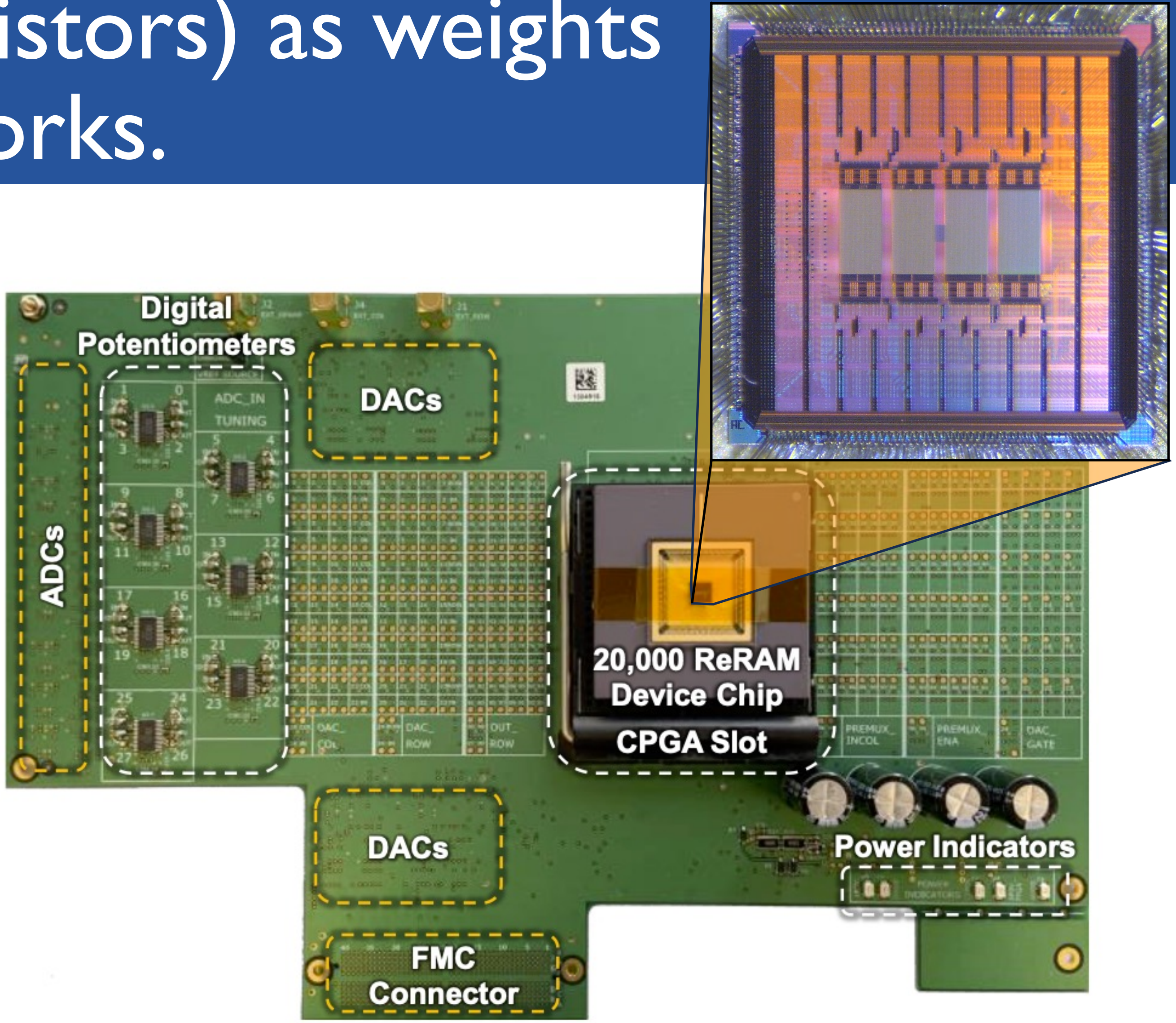


Figure above (from first experiment using Daffodil):
Yousuf et al., arXiv:2404.15621 (2024)

***contact*: nanotechnologyxccelerator@nist.gov**

Grown differently, MTJs can be used to generate high-quality random bits.

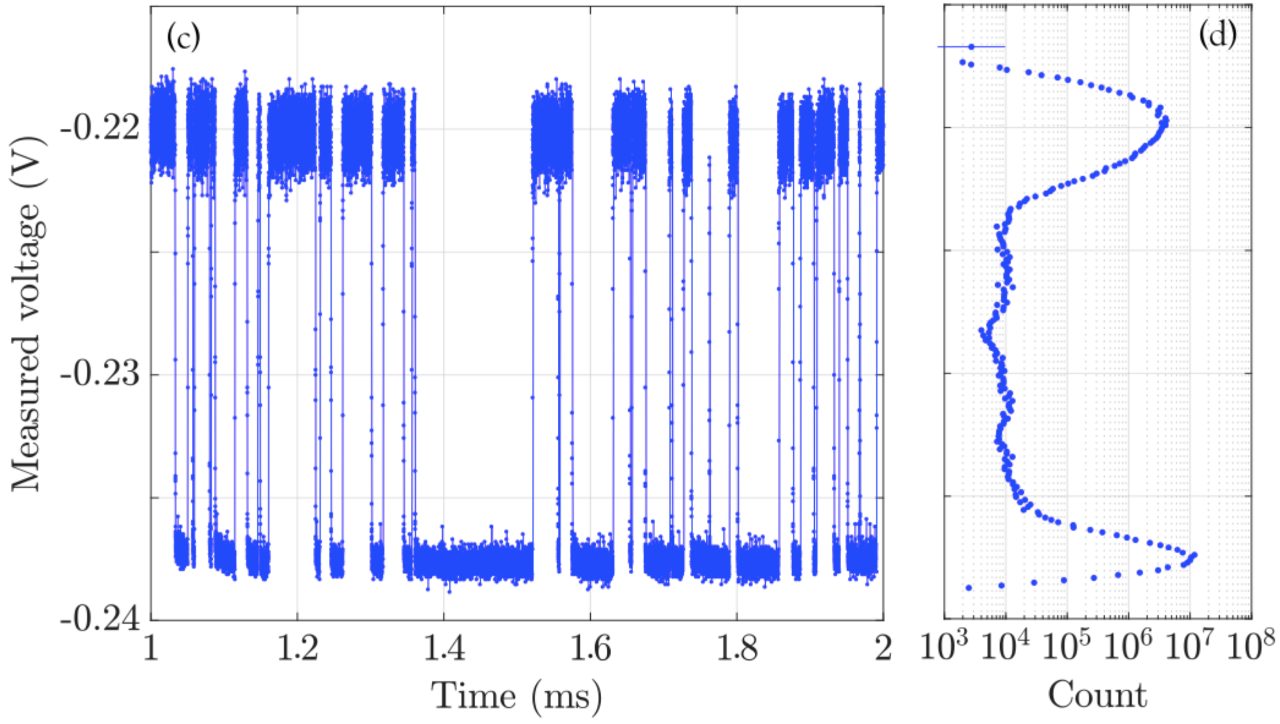
Nonvolatile memory for storage (MRAM)

Energy barrier

Retention Time

60 kT — years

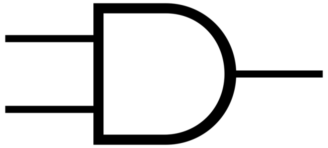
< 7 kT — μs



Thermally driven, superparamagnetic tunnel junctions (SMTJs)

Talatchian, Daniels, et al., Phys. Rev. B 104 (5), 054427, 2021

Simple logic gates get new interpretations in every new encoding—AND/OR gates make a neuron.



AND gate (multiply)

Example

1001010100 (0.4)

x 1010001110 (0.5)

= 1000000100 (0.2)

$$p_{\text{AND}} = p_a p_b$$

OR gate (nonlinear add)

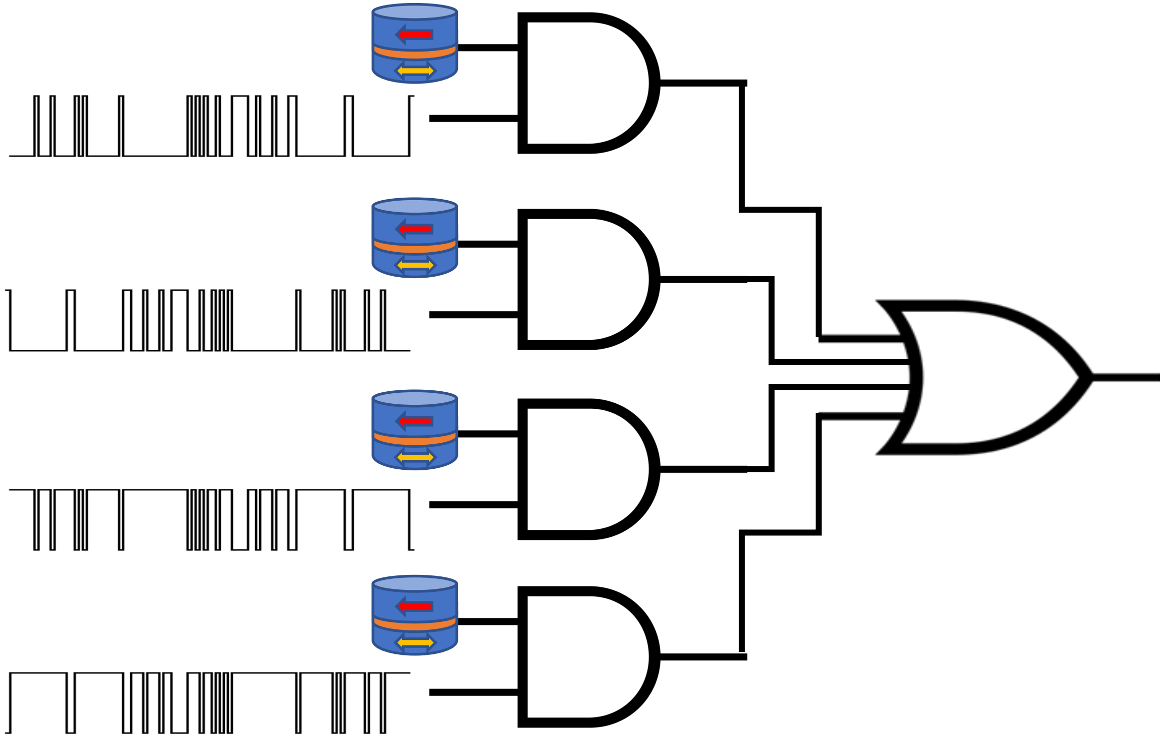
Example

1001010100 (0.4)

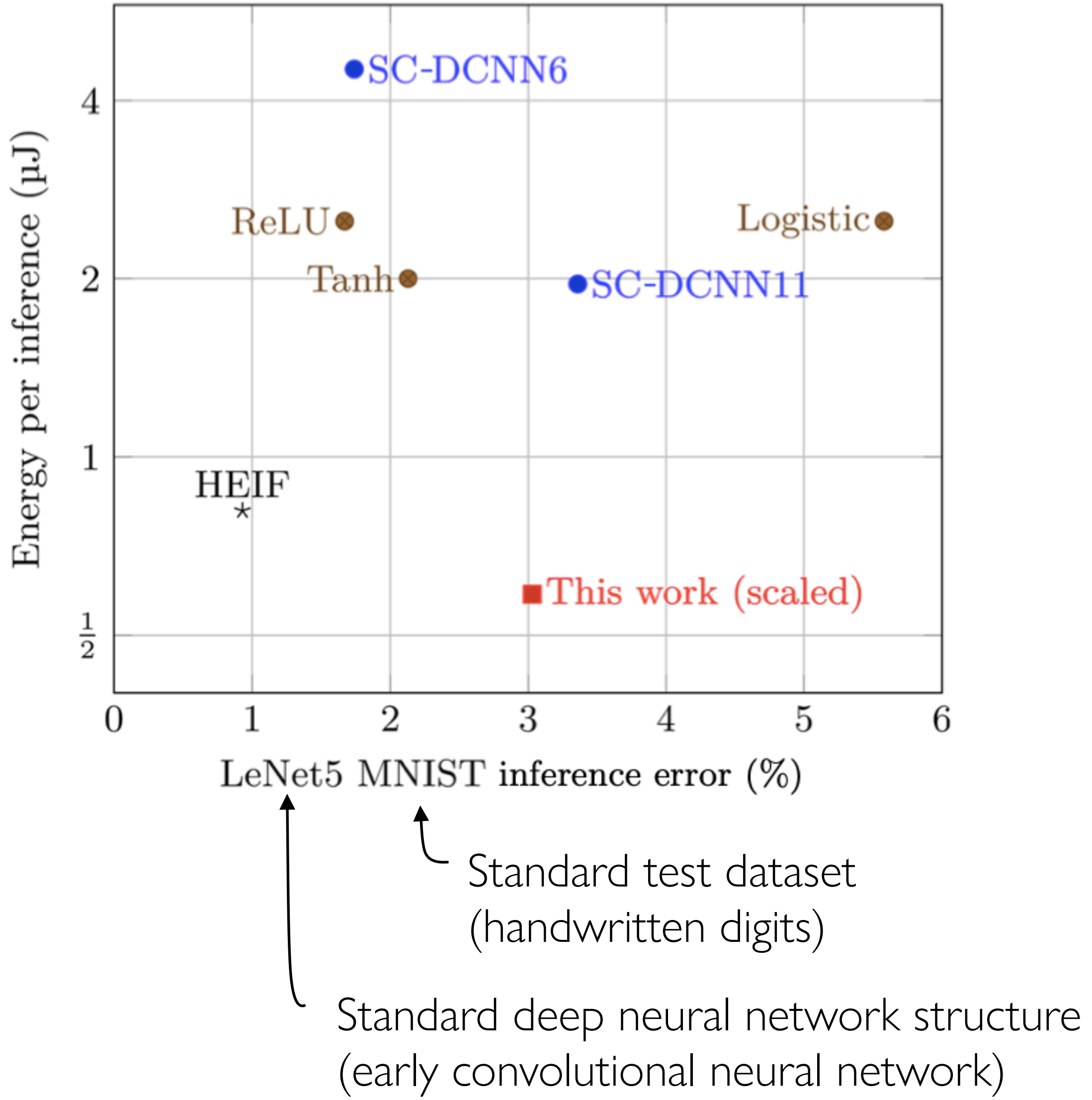
\+ 1010001110 (0.5)

= 1011011110 (0.7)

$$p_{\text{OR}} = p_a + p_b - p_a p_b$$



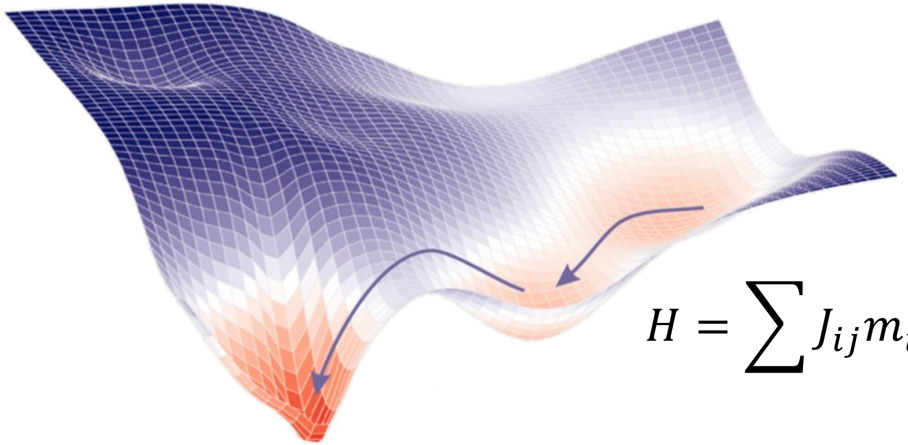
Energy efficiency gains arise when you get high quality randomness for free.



- 2–8x better energy efficiency than contemporary stochastic computing networks on same task, same technology node
- Accuracy lower than state-of-the-art, could be worse – more algorithmic work needed.

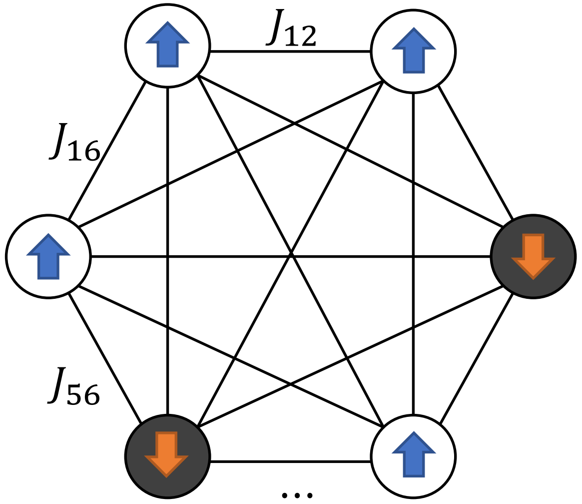
Daniels et al, Physical Review Applied ***13****, 034016 (2020)*

Ising models are physics models that map to combinatorial optimization problems.



$$H = \sum J_{ij} m_i m_j$$

(X, Y) such that $XY = F$



In fact, Ising models are equivalent to many problems of interest.

frontiers in
PHYSICS

REVIEW ARTICLE
published: 12 February 2014
doi: 10.3389/fphy.2014.00005

Ising formulations of many NP problems

Andrew Lucas*

Lyman Laboratory of Physics, Department of Physics, Harvard University, Cambridge, MA, USA

Edited by:
Jacob Biamonte, ISI Foundation, Italy

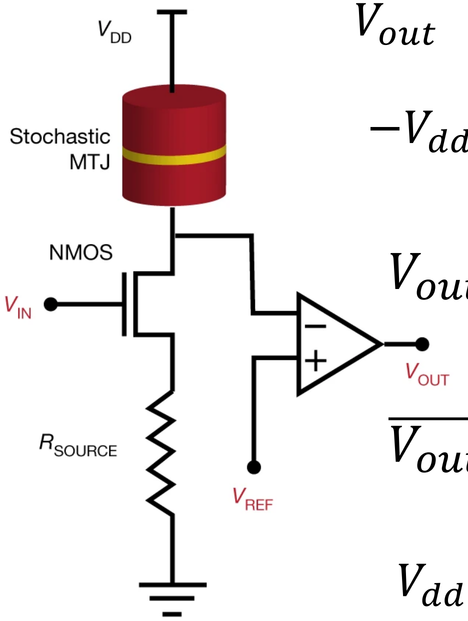
Reviewed by:
Mauro Faccin, ISI Foundation, Italy
Ryan Babbush, Harvard University, USA
Bryan A. O'Gorman, NASA, USA

We provide Ising formulations for many NP-complete and NP-hard problems, including all of Karp's 21 NP-complete problems. This collects and extends mappings to the Ising model from partitioning, covering, and satisfiability. In each case, the required number of spins is at most cubic in the size of the problem. This work may be useful in designing adiabatic quantum optimization algorithms.

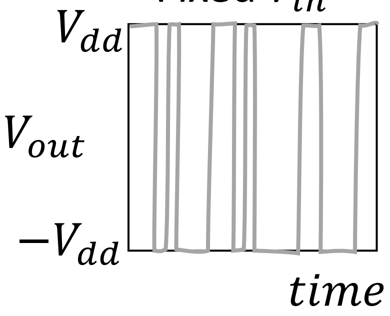
Keywords: spin glasses, complexity theory, adiabatic quantum computation, NP, algorithms

You can implement an Ising spin using an MTJ.

p–bit:

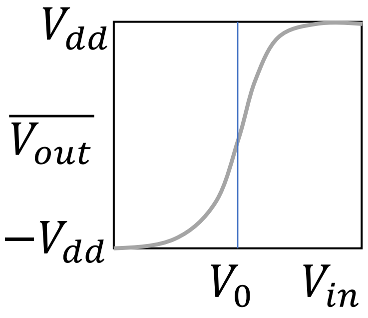


Fixed V_{in}

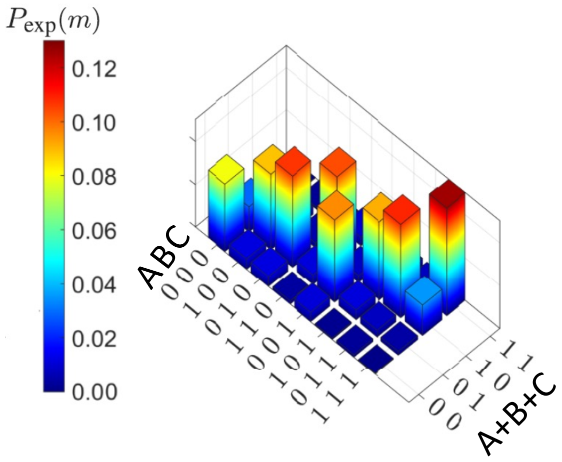


$$V_{out} = \pm V_{dd}$$

$$\overline{V_{out}} = \tanh(V_{in})$$

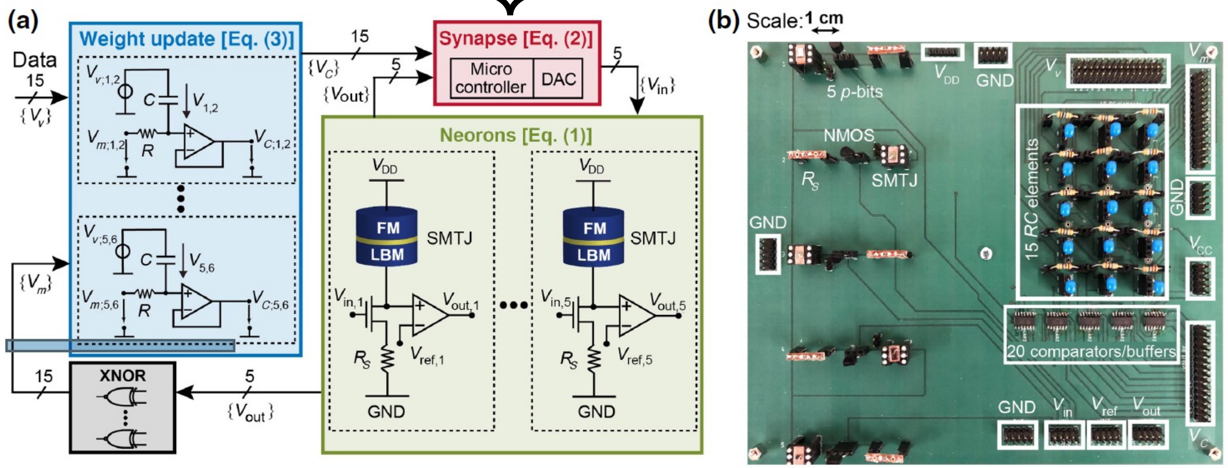


Five coupled p-bits: A, B, C, and the two bits of their sum



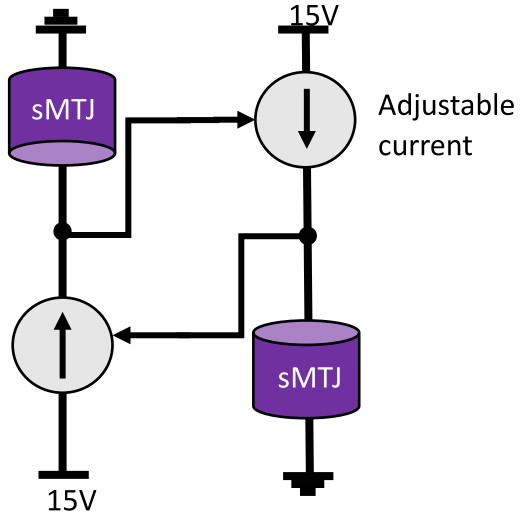
Equal probability of all "correct" configurations

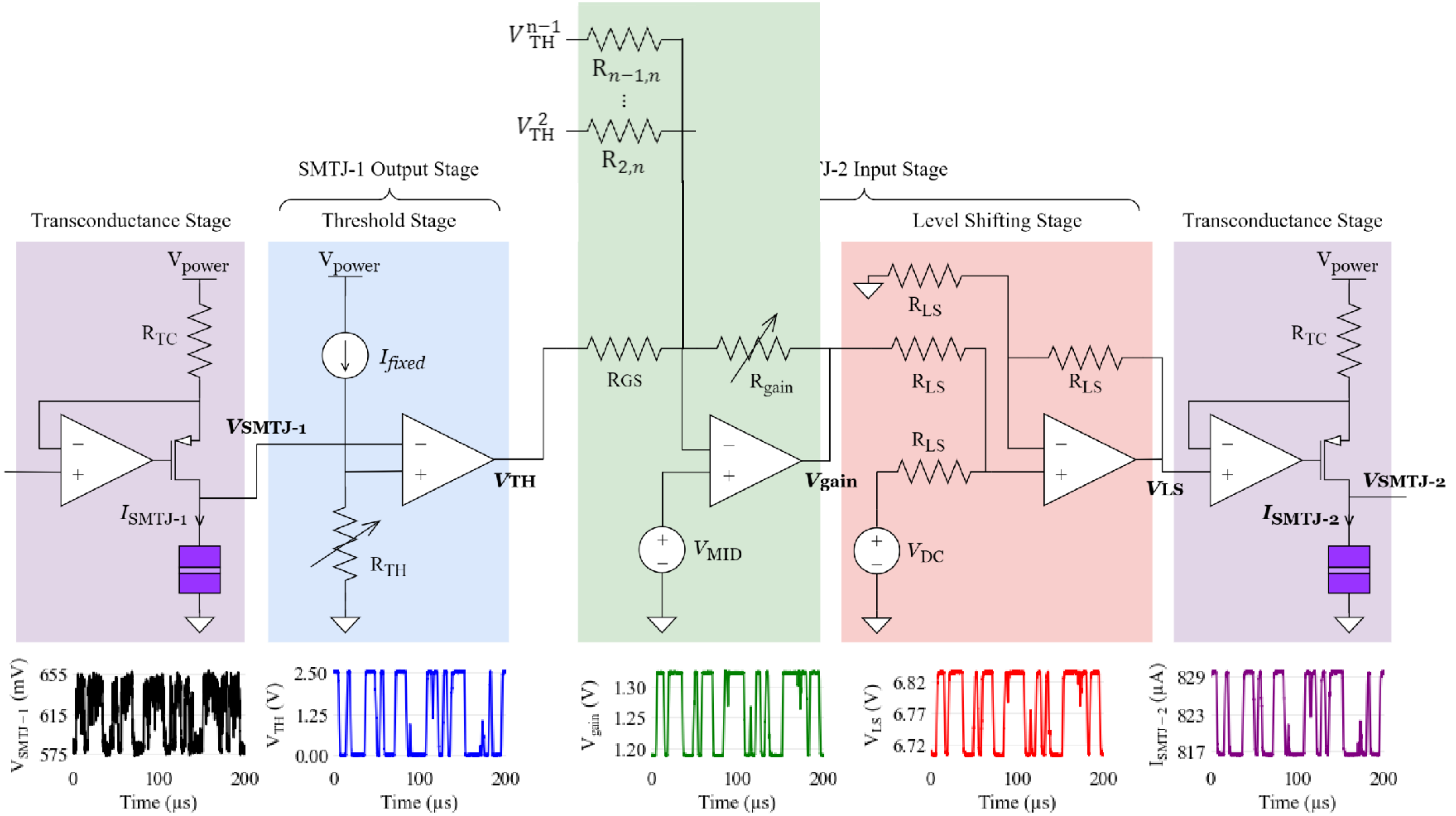
Makes scaling difficult



Kaiser, Borders, Camsari, Fukami, Ohno, and Datta, Phys. Rev. Applied **17**, 014016 (2022).

Our experiment shows direct analog coupling of these Ising spins, and can scale via crossbars.

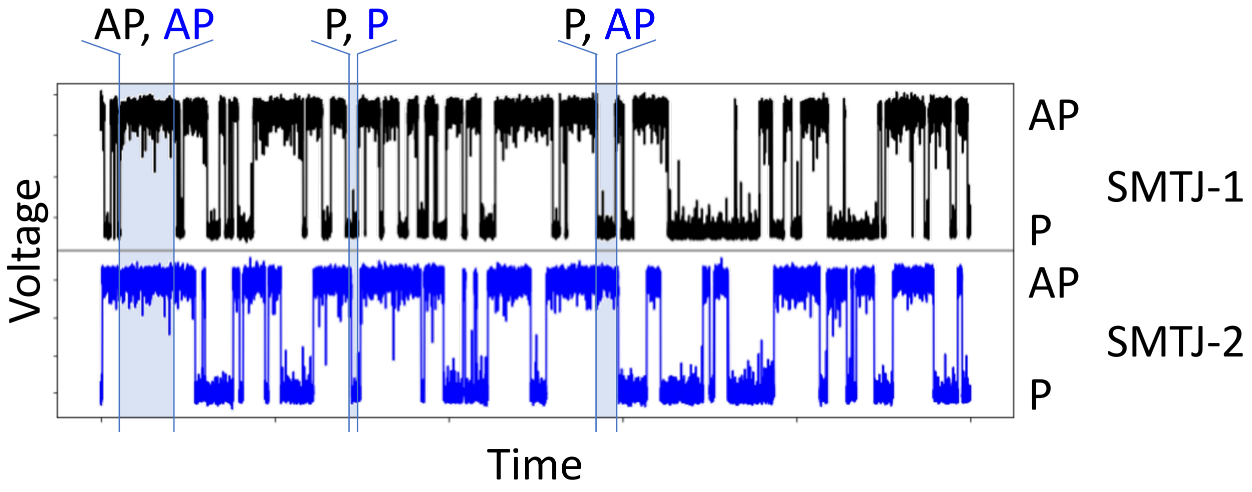




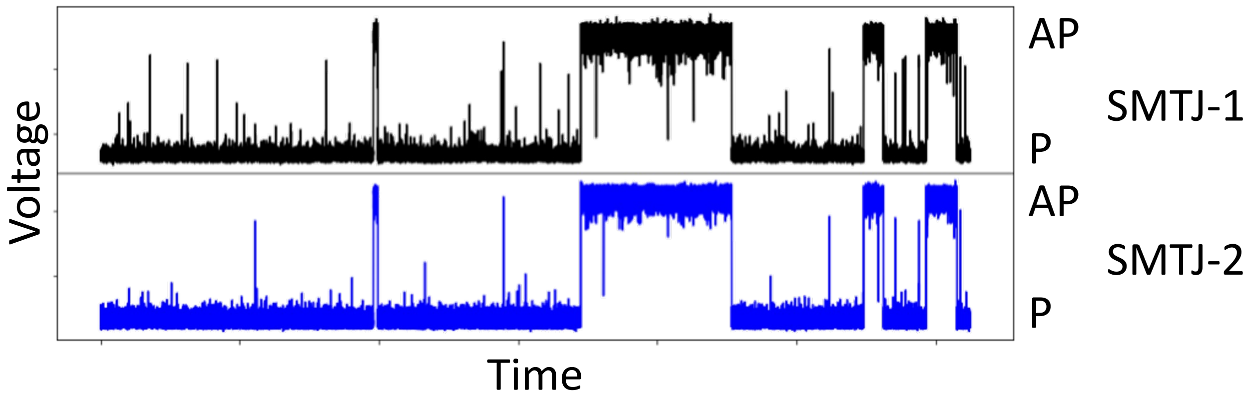
Gibeault, Adeyeye, Pocher, Lathrop, Daniels, Stiles, McClelland, Borders, Ryan, Talatchian, Ebels, Madhavan, Phys. Rev. Applied 21, 034064 (2024).

Increasing the gain corresponds to lowering the model's effective temperature.

Gain =0.0



Gain =0.06



P - parallel $s = 1$

AP – antiparallel $s = -1$

$$\text{Energy} = -GJ\, s_1 s_2$$

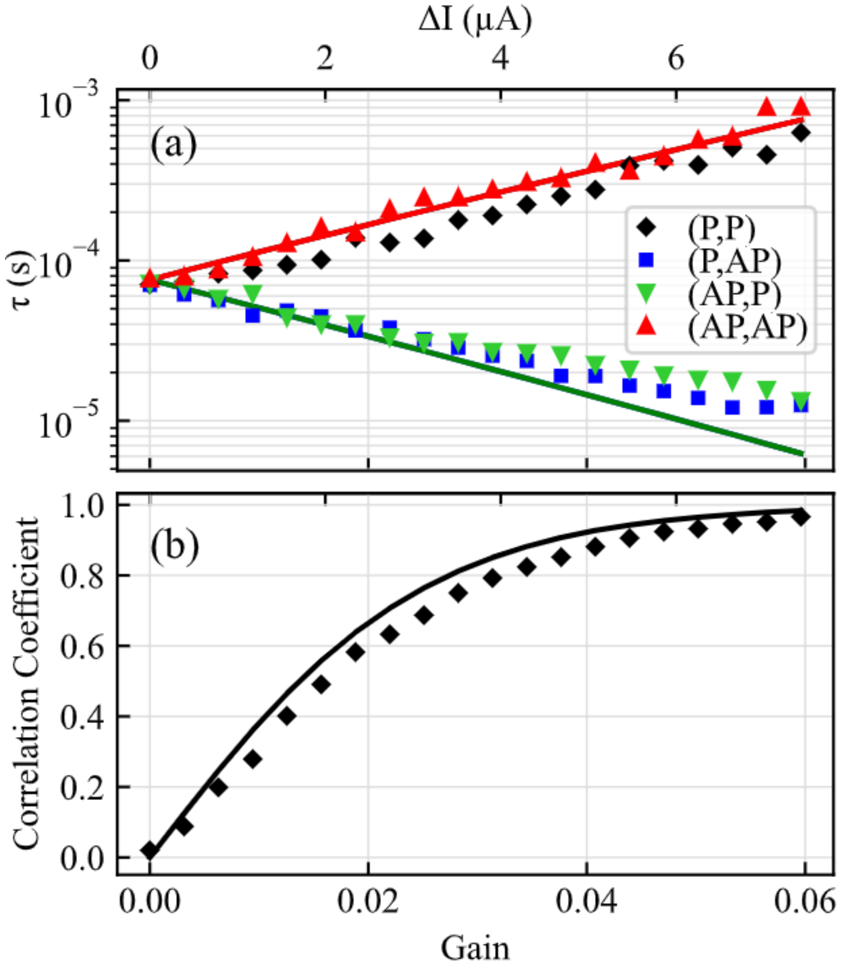
$$\text{Population} \sim \text{Exp}\left[G\frac{Js_1s_2}{kT}\right]$$

$$T_{\text{eff}} = T/G$$

Gibeault, Adeyeye, Pocher, Lathrop, Daniels, Stiles, McClelland, Borders, Ryan, Talatchian, Ebels, Madhavan, Phys. Rev. Applied 21, 034064 (2024).

Increasing the gain corresponds to lowering the model's effective temperature.

Mean time in each configuration



$\sim 1/T_{\mathrm{eff}}$

Lines and curve –
four-state Markov model

P - parallel $s = 1$

AP – antiparallel $s = -1$

$$\mathrm{Energy} = -\,GJ\,s_1 s_2$$

$$\mathrm{Population} \sim \mathrm{Exp}\left[G\,\frac{J s_1 s_2}{kT}\right]$$

$$T_{\mathrm{eff}} = T/G$$

$$\tau_{ij} = \tau_0\,\mathrm{Exp}\left[G\,\frac{\Delta E_{ij}}{kT}\right]$$

Next steps:
- Connect a larger set of "spins"
- Design an ASIC with a crossbar

Gibeault, Adeyeye, Pocher, Lathrop, Daniels, Stiles, McClelland, Borders, Ryan, Talatchian, Ebels, Madhavan, Phys. Rev. Applied 21, 034064 (2024).

CMOS+X:

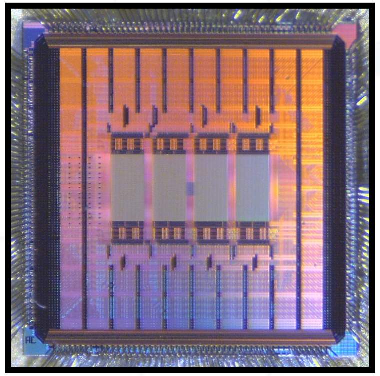
- Adding new devices with novel functionality can influence how you think about the whole stack.

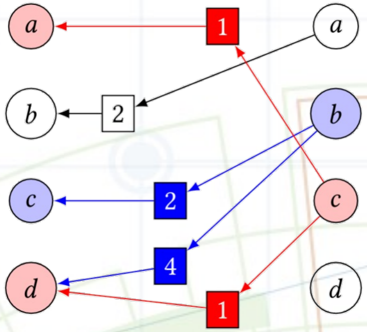
Temporal computing:

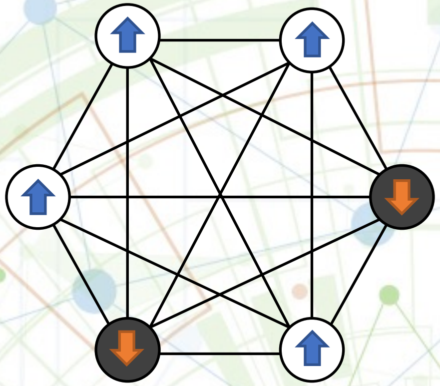
- Natural, efficient graph computation
- *Future:* log-domain analog neural networks

Stochastic computing:

- Fast, high-quality random bits ⟹ new architectures
- Ising machines for combinatorial optimization
- *Future:* stochastic training of low-precision neural networks







NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Matthew W. Daniels
matthew.daniels@nist.gov