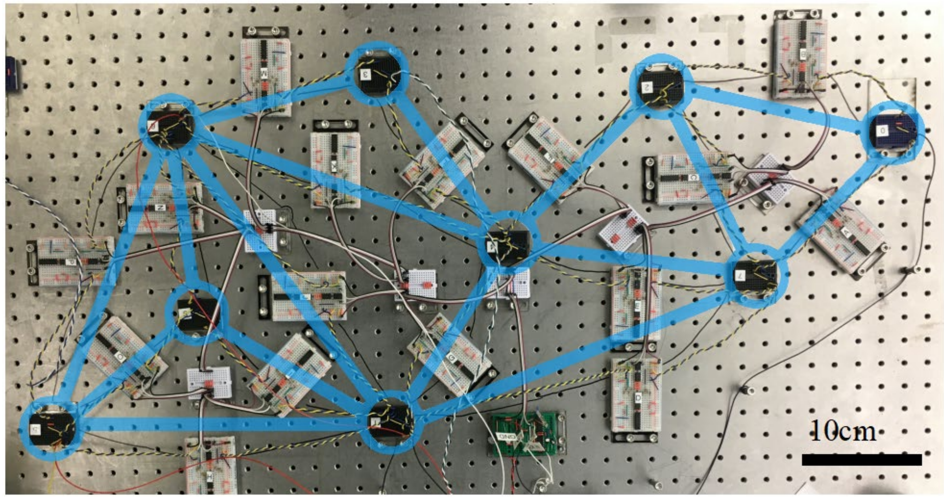
Physical Learning in Resistor Networks

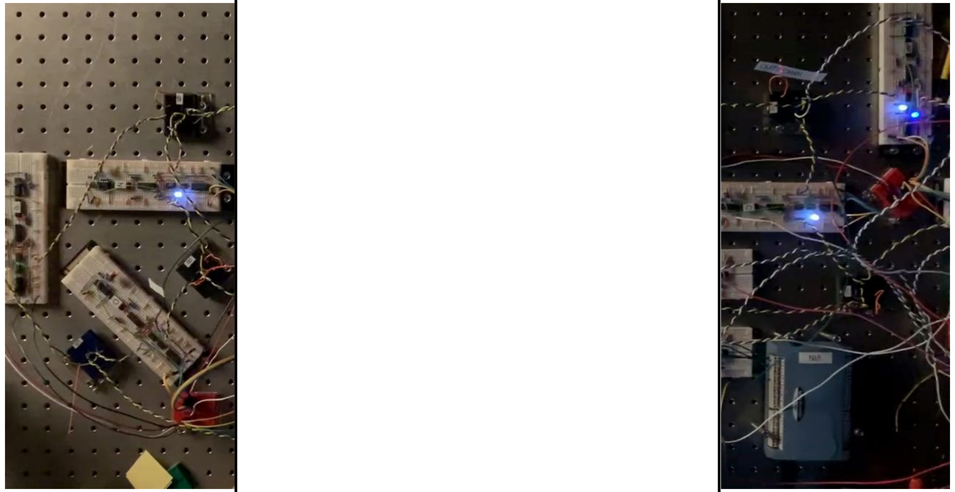
Sam Dillavou, Menachem Stern, Jacob Wycoff, Ben Beyer, Marcelo Guzman, Lauren Altman, Marc Z. Miskin, Andrea J. Liu

Douglas J. Durian

University of Pennsylvania, Phil *SA*

Electronic networks of identically-constructed edges with uctances:

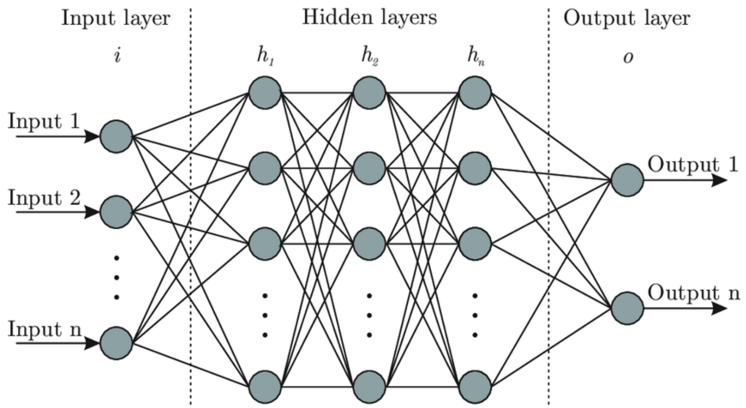




Doing AI without a computer – fast and ene

Artificial vs Real Neural Networks

- Brain circuitry is slow & noisy, yet brains are vastly more capable & energy efficient than ANNs. Why, and can we use their trick?



- Relatively narrow range of tasks
- Fragile wrt damage
- Tremendous hidden costs
  - ➢ 500 kJ and 500 ml water per 100-word ChatGPT response

- Controls thoughts, memory, senses, motor skills, regulation… & adaptable
- Robust to damage
- Energy efficient
  - ➢ 20 W = 1700 kJ/day

Contrastive Local Learning Networks

- Goal: adjust edge conductances so that output voltages are a desired function of input voltages ("computed" physically, for free).

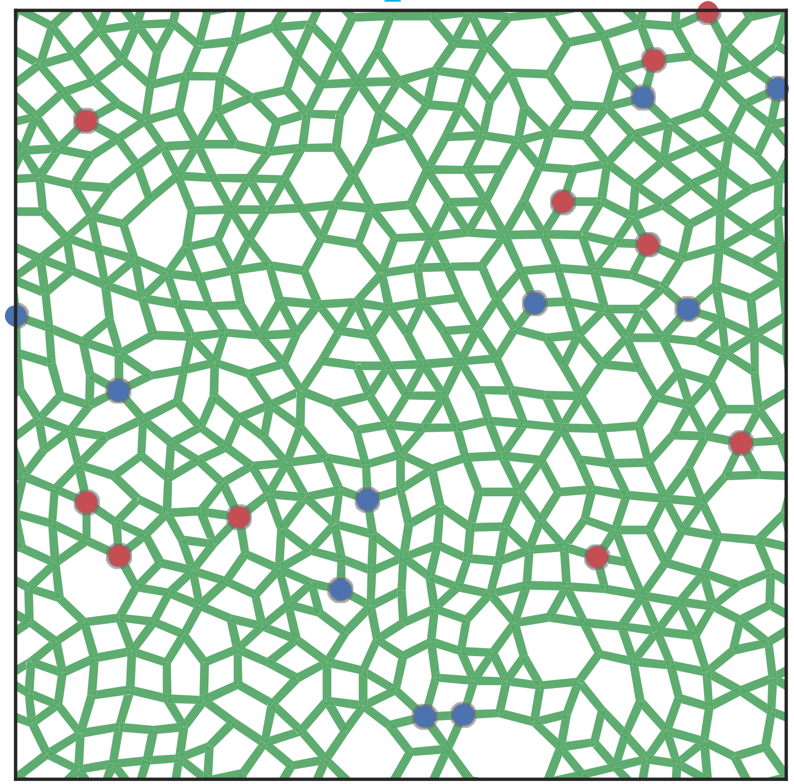
- Approach: ***local*** learning rules based on ***contrast*** of behavior for two different boundary conditions:
  1. "free" where output node values are measured
  2. "clamped" where training data are imposed on output nodes

Contrastive Hebbian Learning (Movellan, 1991)

Equilibrium Propagation (Scellier & Bengio, 2017)

Coupled Learning (Stern, Hexner, Rocks, Liu, 2021)

- *inputs*
- *outputs*



resistors, pipes, springs…

bottom up & **highly recurrent**

Should get brain-like advantages over ANNs – eg energy efficiency

Coupled Learning Rule (LOCAL)

- Traditional cost function = (desired response – free response)^2 {>0}
  - squared to guarantee it's positive, then minimized by gradient descent

- New **contrast function** = dissipation rate difference, $\mathcal{P}^{clamped} - \mathcal{P}^{free}$ {>0}
  - positive due to minimization of power as currents equilibrate for the given BCs

Menachem "Nachi" Stern

Evolve the edge conductances $\{k_j\}$ to drive contrast function to zero:

$$\dot{k}_j \propto -\frac{\partial}{\partial k_j}\left[\mathcal{P}^{clamped} - \mathcal{P}^{free}\right]$$

$$= -\frac{\partial}{\partial k_j}\left[\sum_i (V_i^2 k_i)^{clamped} - \sum_i (V_i^2 k_i)^{free}\right]$$

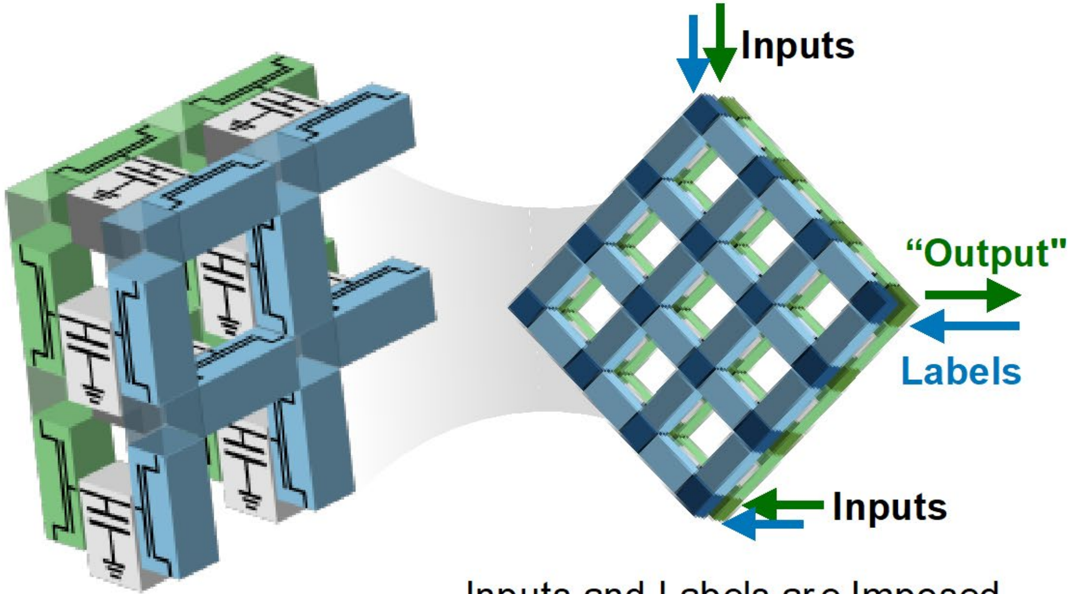
$$= -\left[(V_j^2)^{clamped} - (V_j^2)^{free}\right]$$

This rule is LOCAL
$\{\dot{k}_j$ depends only on what edge j is doing$\}$

[Stern, Hexner, Rocks, Liu, Phys Rev X 2021]

In Laboratorium? Twin Network Trick

- Build repeat units (edges) consisting of two variable resistors kept at same resistances, with circuitry to update their shared value in unison
- Connect the edges together in some chosen architecture, giving twin networks
  - One runs free BCs, the other runs clamped BCs
  - Circuitry on each edge performs the local learning rule



Twin Networks

Inputs and Labels are Imposed Voltages, Outputs are Measured.

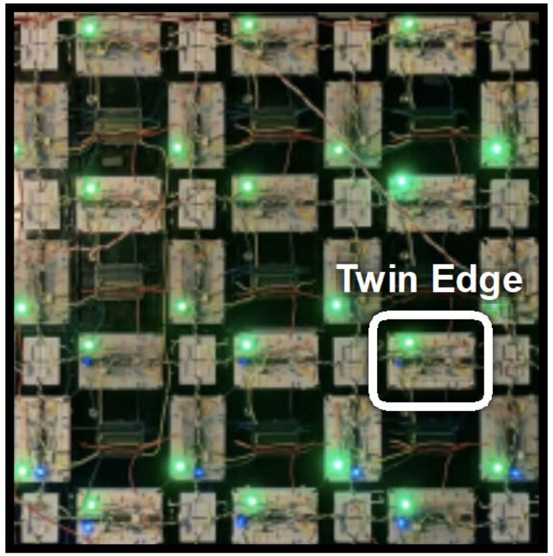


Image of Learning Material

Sam Dillavou

[Dillavou, Stern, Liu, Durian (Phys. Rev. Applied 2022); Dillavou, Beyer, Stern, Liu, Miskin, Durian (PNAS 2024)]

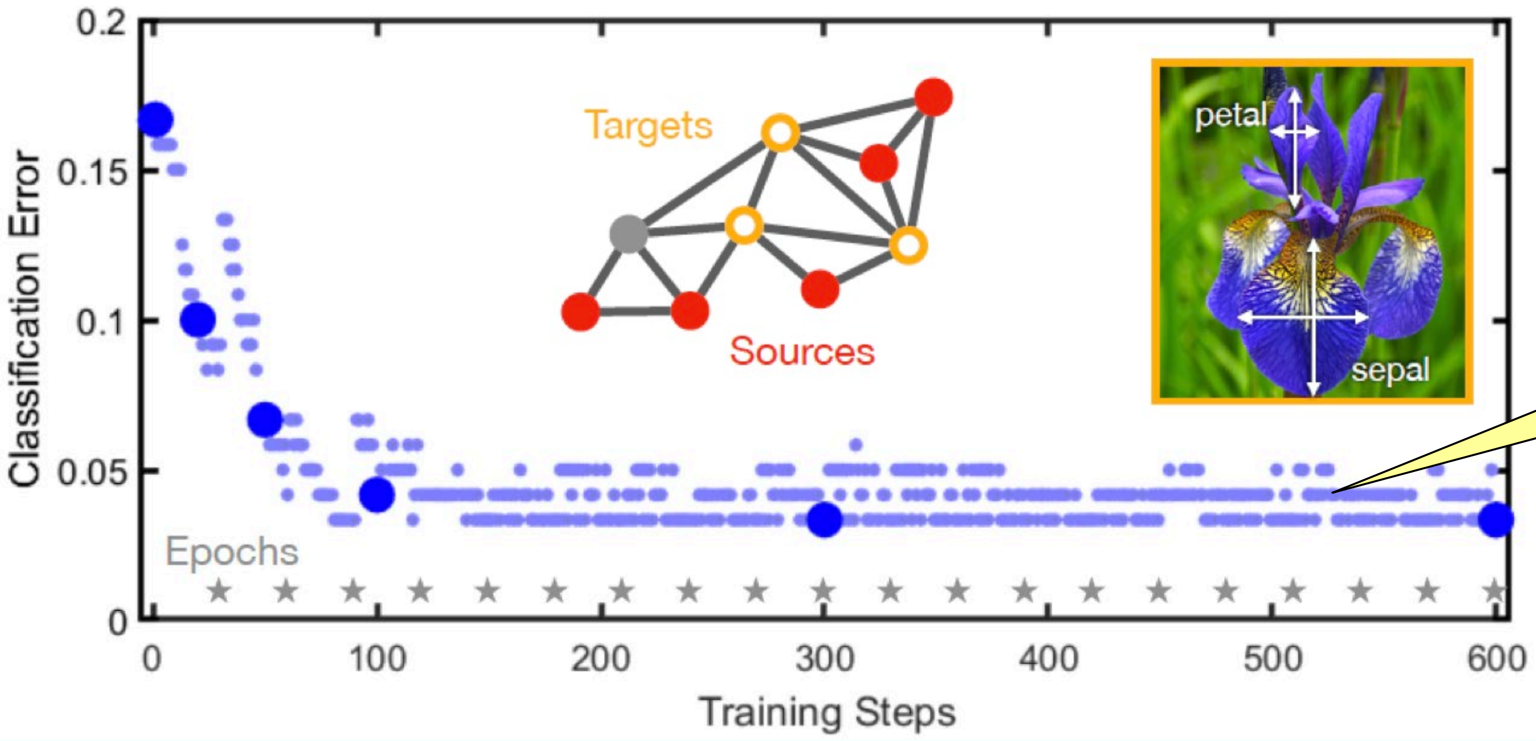
Example: Classification of Iris Species

- Dataset: 4 measurements x 50 flowers x 3 species

  {10 of each species for training + 40 of each species for testing}

- Five input nodes: these four measurement plus one ground
- Three target nodes: L_2 norm of outputs from <input> for each species

  {reset target every epoch = 30 training steps}

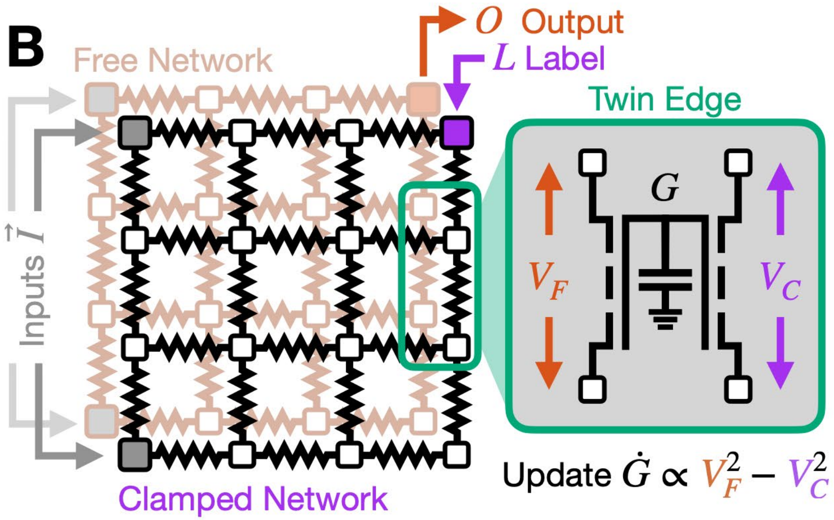


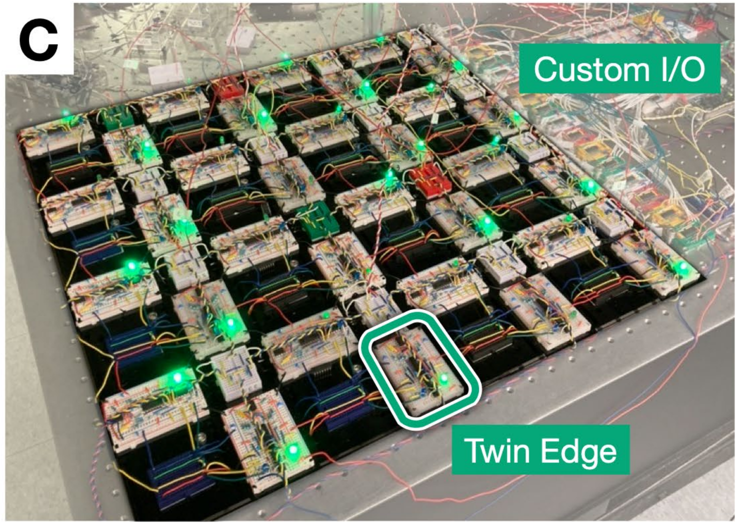
Better than 95% classification accuracy

Continuous nonlinear conductances

- 2nd gen. made with transistors (gate voltage G is learning degree-of-freedom)

  {implements actual coupled learning rule}





32-edge 16-node network
(with or without periodic BCs)

Sam Dillavou

Benjamin Beyer

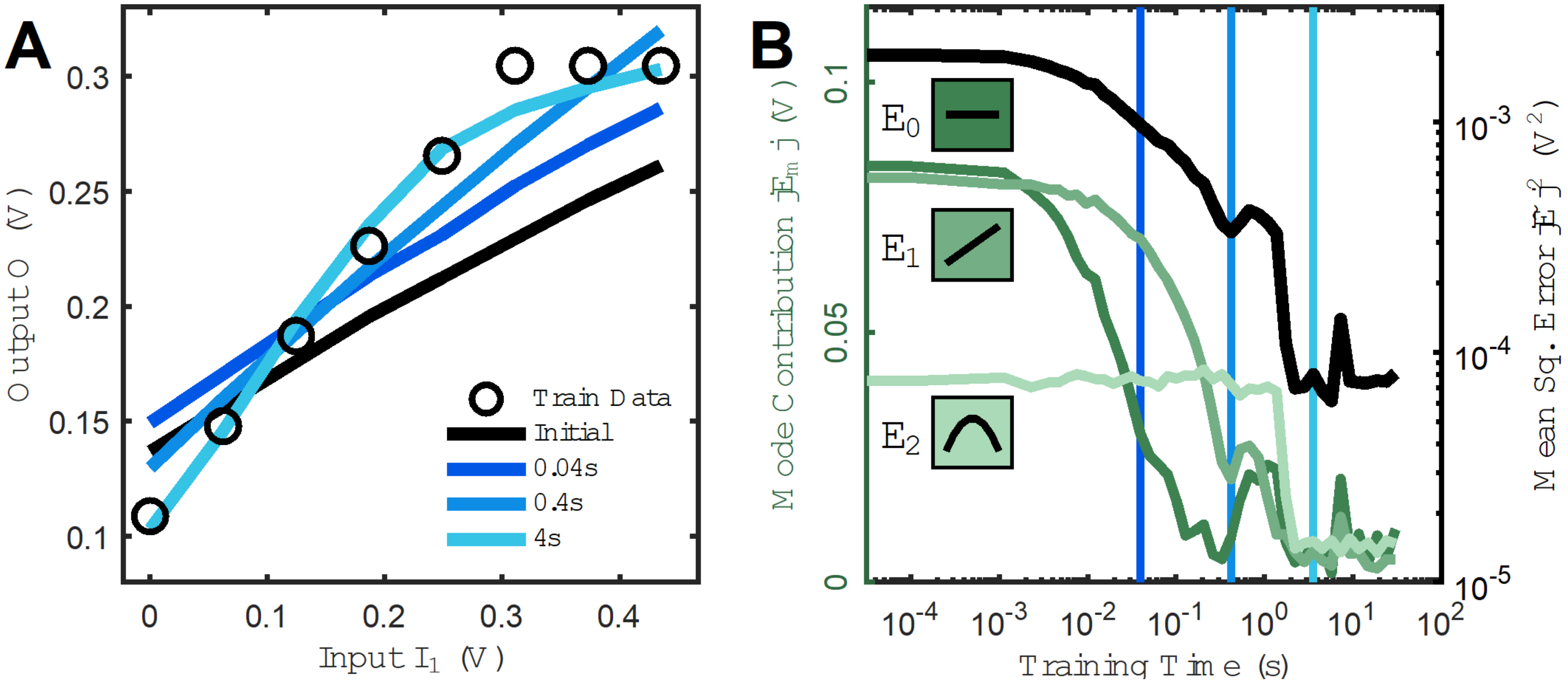
Marc Miskin

[Dillavou, Beyer, Stern, Liu, Miskin, Durian (PNAS 2024)]

Nonlinear regression

- Learns in order of complexity: first the mean, then the slope, then the curvature

c.f. spectral bias



[Dillavou, Beyer, Stern, Liu, Miskin, Durian (PNAS 2024)]

Nonlinear classification

- 2 analog inputs (I_1, I_2); 2 fixed voltages (+,-); 1 analog output ($O = O_+ - O_-$)
  - Training data: blue and orange dots
  - Testing results: background color (decision boundary in black, O = 0 volts)

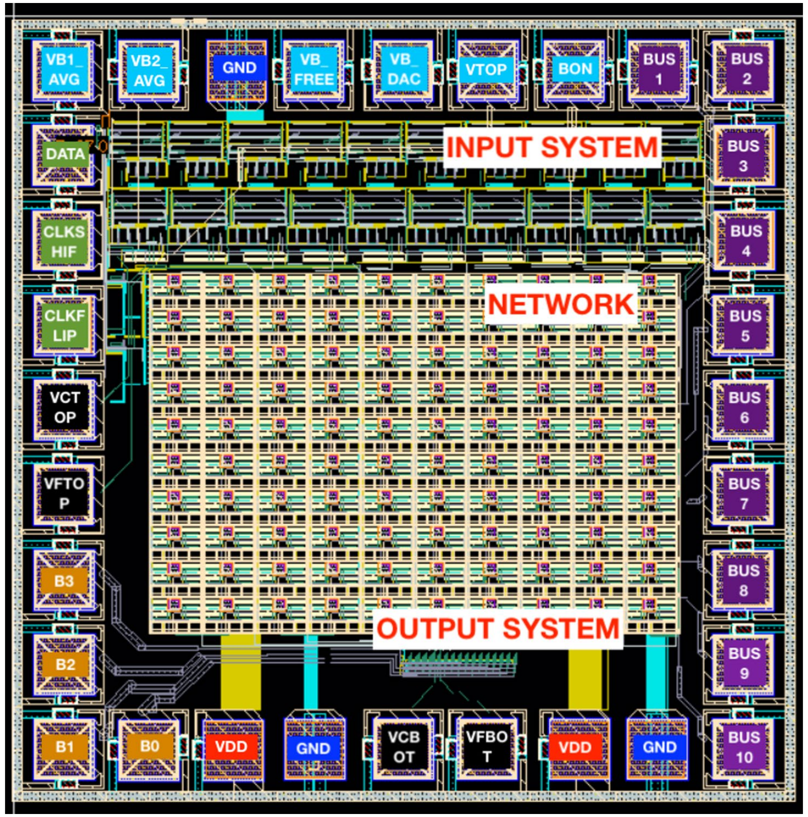
Square network with PBCs

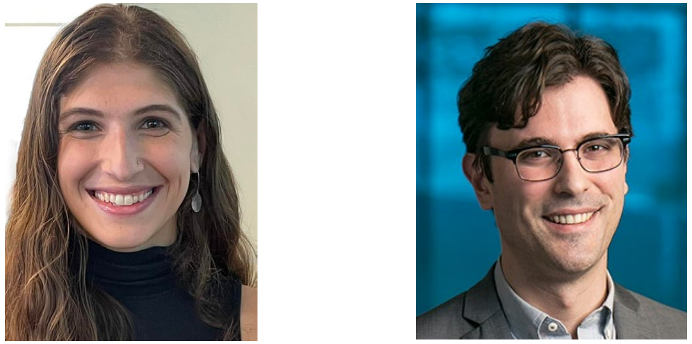
thickness = conductivity
color = Δconductivity

[Dillavou, Beyer, Stern, Liu, Miskin, Durian (NeurIPS 2023)]

How far can we scale it up?

1. Gen.2 PCBs with one edge/board: for up to O(300)-edge networks…
2. Gen.3 chips: A 10x10 network in 1x1 mm^2…





Lauren Altman

Marc Miskin & Sophia Handley & Tarunyaa Sivakumar

Edge nonlinearity & network architecture for max expressivity?
Effects of noise & physical imperfections? Learning dynamics?

CLLNs for applications

- A new hardware compute platform that is fast & energy efficient?

    Qualcomm AI chip: 10^{-12} J/parameter/inference

    LLMs: 10^{12} parameters

    - Gen.2 on breadboards: 10^{-11} J/parameter/inference
    - Gen.3 on chips: 10^{-14} J/parameter/inference
    - Potentially: 10^{-18} J/parameter/inference for 10^{9} parameters on chip

- Disrupt current artificial neural network paradigm for doing AI?
- Find niche applications where the energy cost must be minimized, or the task must be adapted to changing conditions?
    - eg edge computing in sensors (incl. medical), robotic swarms, triggers in particle detectors…

CLLNs for science

- Not just "analog in-memory learning for analog in-memory computation"

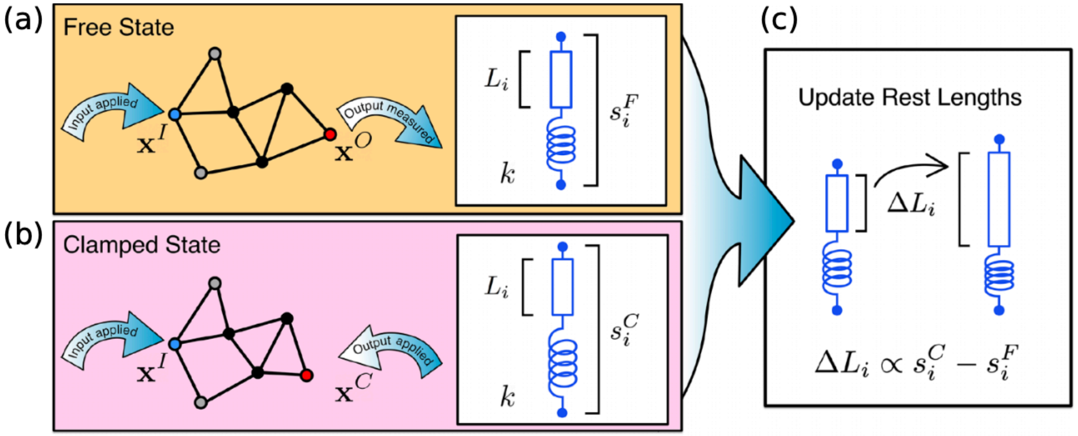
- Model systems where the emergence of learning can be isolated, hopefully understood as a nonlinear many-body physical process, and perhaps translated to neuroscience

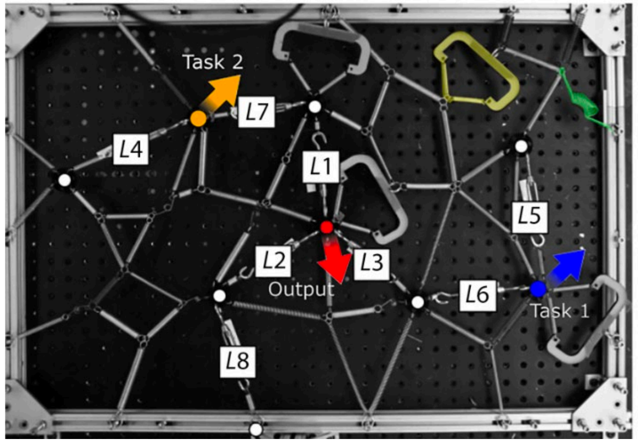
- Novel form of physical matter where ***<u>interactions</u>*** are individually tunable
    - "metamaterials" consisting of many copies of identically-constructed repeat units
    - c.f. Hopfield networks and Boltzmann Machines (2024 Nobel Prize in Physics, but *in-silico*)
  - Electrical
  - Fluidic
  - Mechanical

Mechanical CLLNs

- 1st GEN: Turnbuckle (spring rest length) is tunable learning degree of freedom
- 2nd GEN: Coil thickness (spring stiffness) is tunable learning degree of freedom, with local motor and strain gauge for doing local learning [Altman… & Sung, in progress]

[1]

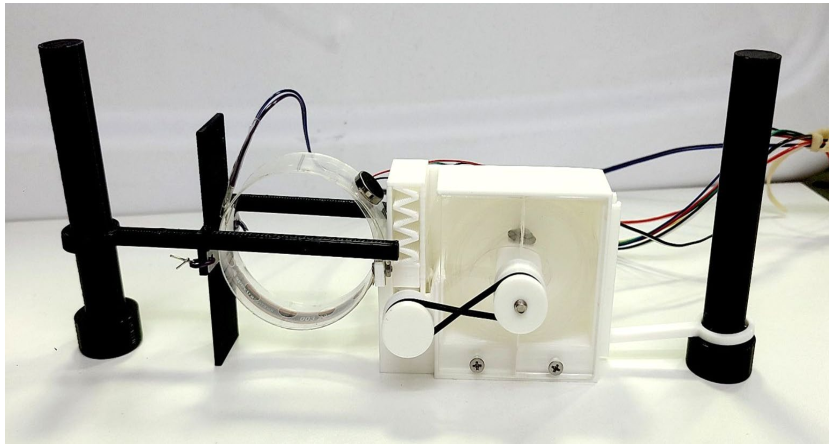




Lauren Altman

L. E. Altman, M. Stern, A. J. Liu, D. J. Durian, "Experimental Demonstration of Coupled Learning in Elastic Networks," Physical Review Applied 22, 024053 (2024)

[2]



Many open questions…

- Circuits:
  - Best network architecture and input/output node selection?
  - Best edge nonlinearity? Universal approximation theorem for edges?
  - Noise: when annoying and when helpful?
  - Energy vs accuracy tradeoff?
  - Statistical physics and nonlinear dynamics of learning?
    - *simultaneous relaxation/optimization in two high-dimensional rugged landscapes*
    - *an interesting new form of matter with variable interactions between identical elements*

    "Many more is more different"
  - Insights portable to brain function and vice-versa?

- Learning in other types of network?
  - Microfluidic and mechanical (bio- and architectural)
  - Optical?

ACKNOWLEDGEMENTS

experiment

Sam Dillavou

Jacob Wycoff

Benjamin Beyer

Lauren Altman

theory

Menachem Stern

Marcelo Guzman

Prof. Andrea Liu

2nd generation

Prof. Marc Miskin

& Sofia Handley, Prof. Dinesh Jayaraman, Maggie Miller, Shivangi Misra, Tarunyaa Sivakumar, Prof. Cynthia Sung

UPenn MRSEC

LRSM

Center for Soft & Living Matter
University of Pennsylvania

Dillavou, Stern, Liu, Durian (Phys. Rev. Applied 2022)
Dillavou, Beyer, Stern, Liu, Miskin, Durian (PNAS 2024)
Dillavou, Guzman, Liu, Durian (arXiv:2505.22887)