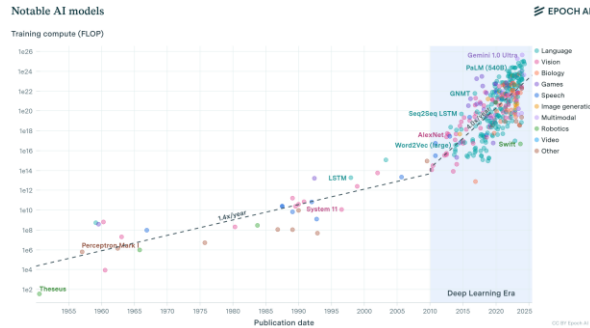


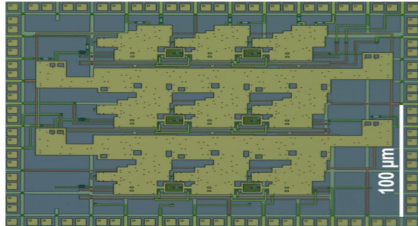
Current State of, and Potential for, Superconducting Neuromorphic Computing

Mike Schneider

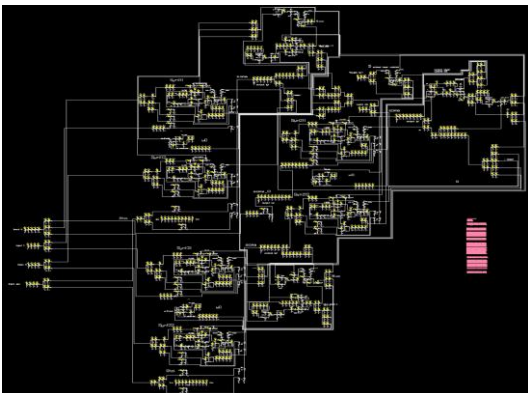
National Institute of Standards and Technology, Boulder, CO



- Introduction / why neuromorphic



- Survey current state of the field

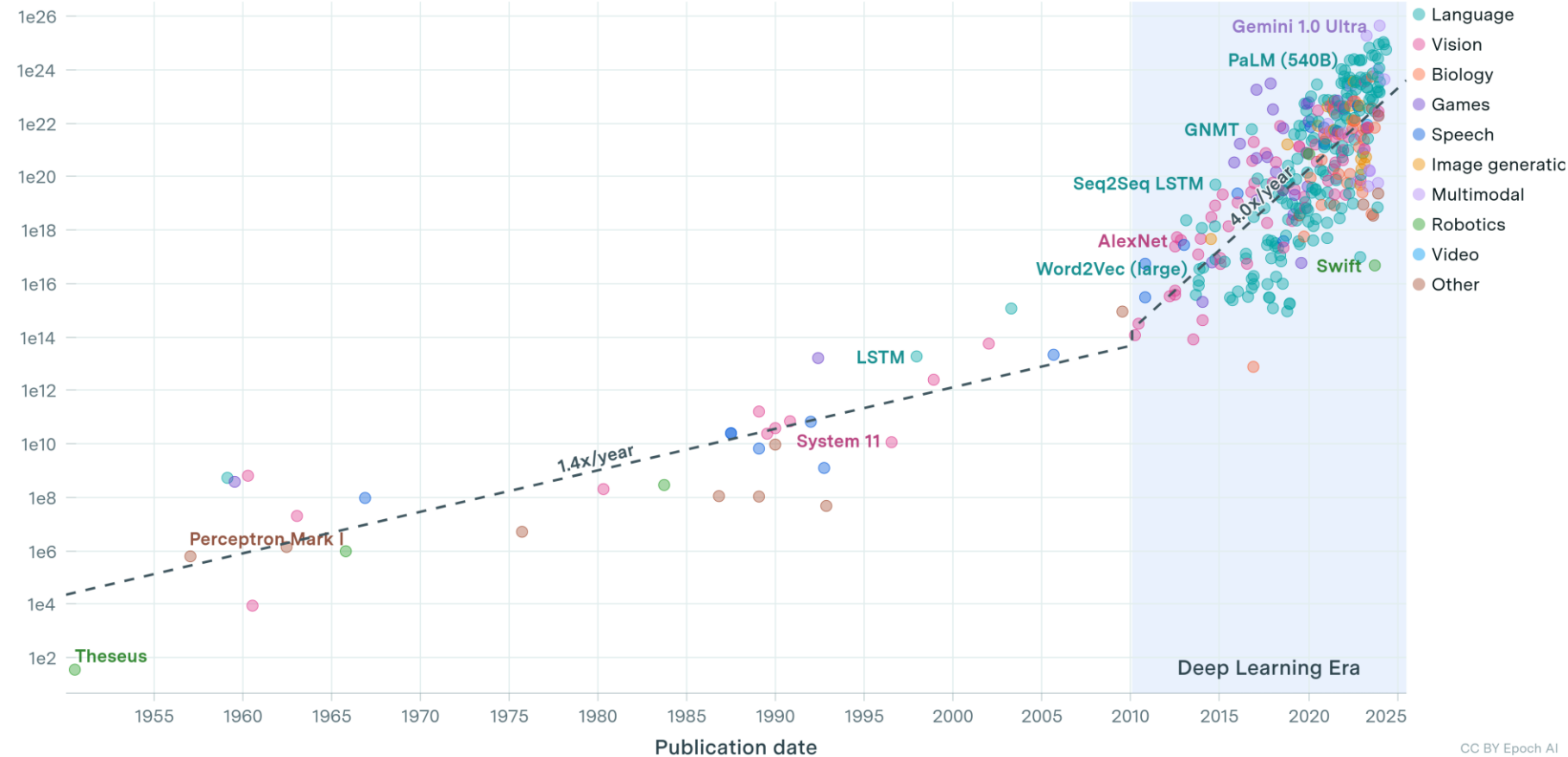


- Reinforcement learning architecture for SFQ based SNNs

AI has become economically driven in the last 10 years

Notable AI models

Training compute (FLOP)



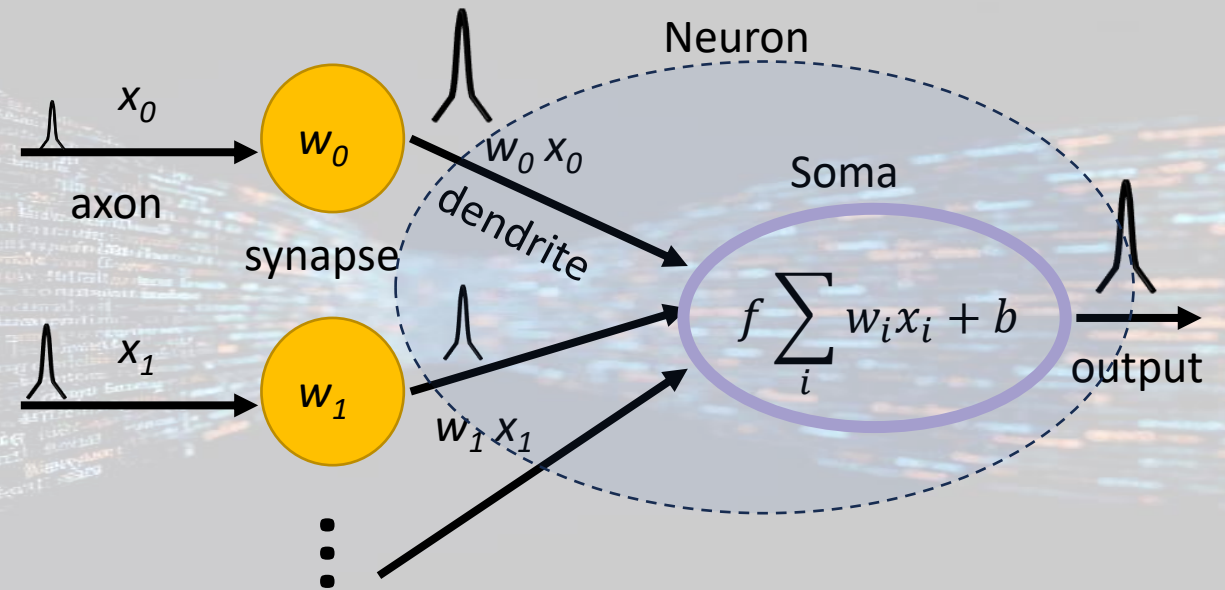
CC BY Epoch AI

- 2012 Alex Net
 - Image recognition
 - Large datasets and GPUs beat other image recognition algorithms
- 2017 “Attention is all you need”
 - Transformer Architecture
 - Large Language Models begin to demonstrate impressive results e.g. language translation
- 2025 “Deep Seek”
 - Smaller network with reinforcement learning
 - More efficient LLM

What is modern AI

- There is no single definition
- In general, the last 10 years have been dominated by:
- Neural networks
- Trained on large data sets
- Different structures depending on the application
 - Transformers for language
 - Convolutional networks for image recognition

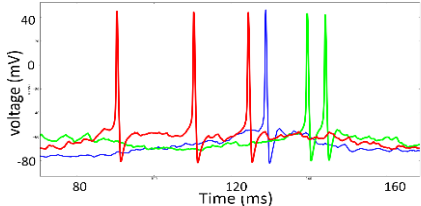
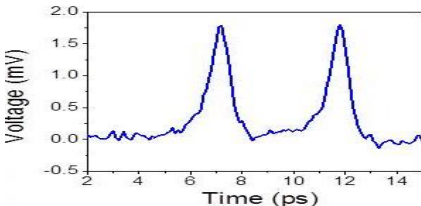
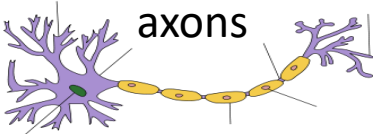
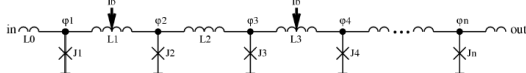
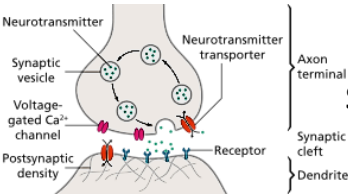
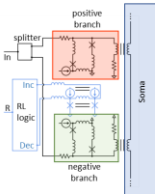
Basic Neural Network Components and their biological inspiration



Why neuromorphic?

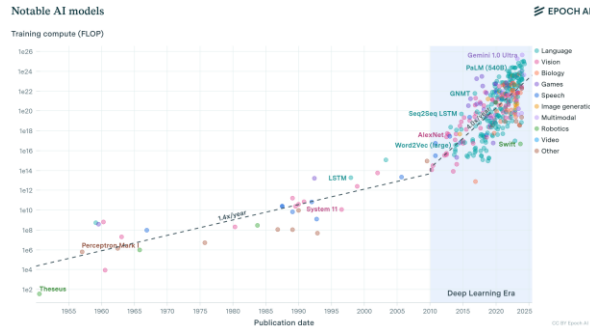
- Modern AI algorithms shows us that taking inspiration from the brain can lead to whole new computing paradigms
- The human brain is still unparalleled for many computational tasks and uses about 20 W
- Neuromorphic computing takes bio-inspired computing to the device level
- Potential advantages
 - More biological realism could advance neuroscience understanding of how the brain computes
 - Designing hardware from the device up could unlock speed and efficiency
 - More biological realism could unlock yet undiscovered computing paradigms

Why superconducting neuromorphic

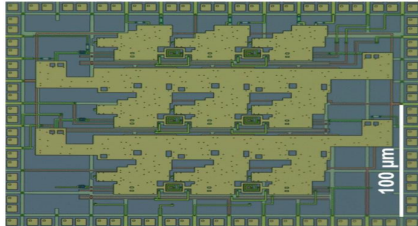
Information transfer: quantized pulse trains	<p>Human Brain</p> 	<p>Josephson junctions</p> 
Long distance “lossless” pulse transmission		<p>Active or passive transmission lines</p> 
Memory/ plasticity	 <p>synapse</p>	 <p>SQUID synapse Magnetic JJs Many more</p>
Spike energy	10 fJ	10^{-4} fJ
Spike time scale	10^{-2} s	10^{-11} s

Where does neuromorphic fit in?

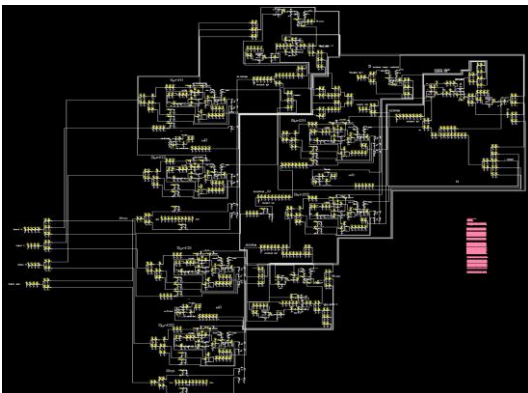
- No single answer!
- Large scale (see next talk)
 - Emergent behavior on a very large scale
 - Energy efficiency and connectivity (enabling larger systems)
 - bio-realism (enabling more advanced architectures)
- Smaller scale
 - Integrated with measurements and sensors
 - Similar to digital signal processing in the 80/90s
 - High speed communications / radar
 - Quantum readout / state preparation



- Introduction / why neuromorphic



- Survey current state of the field



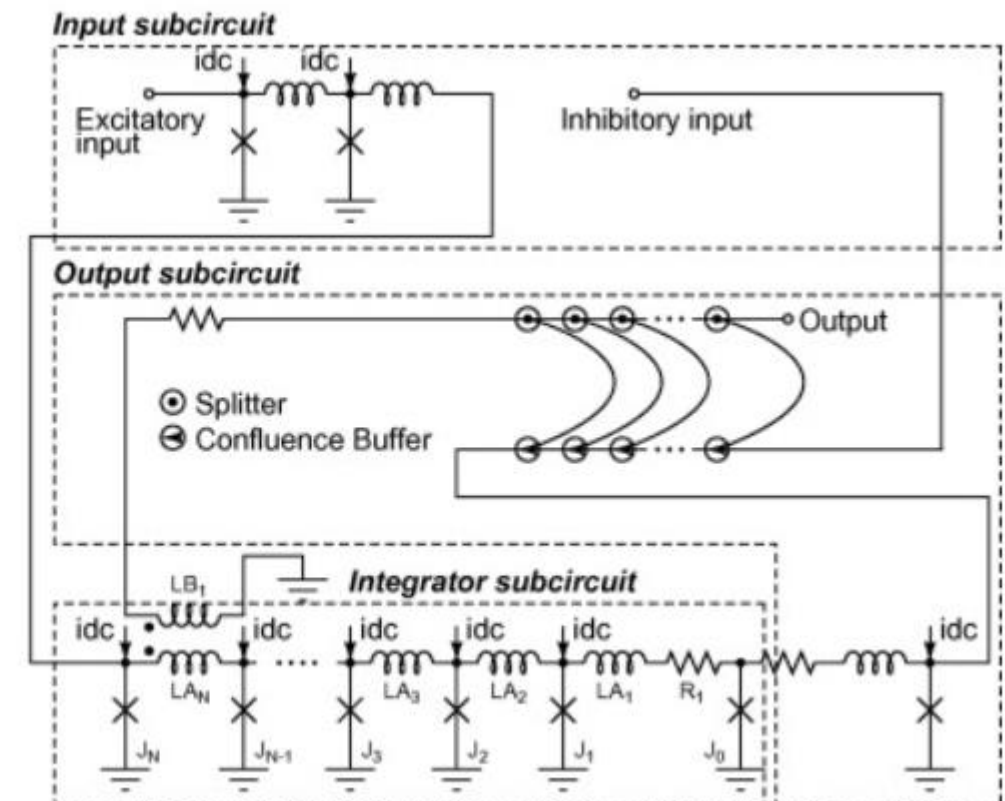
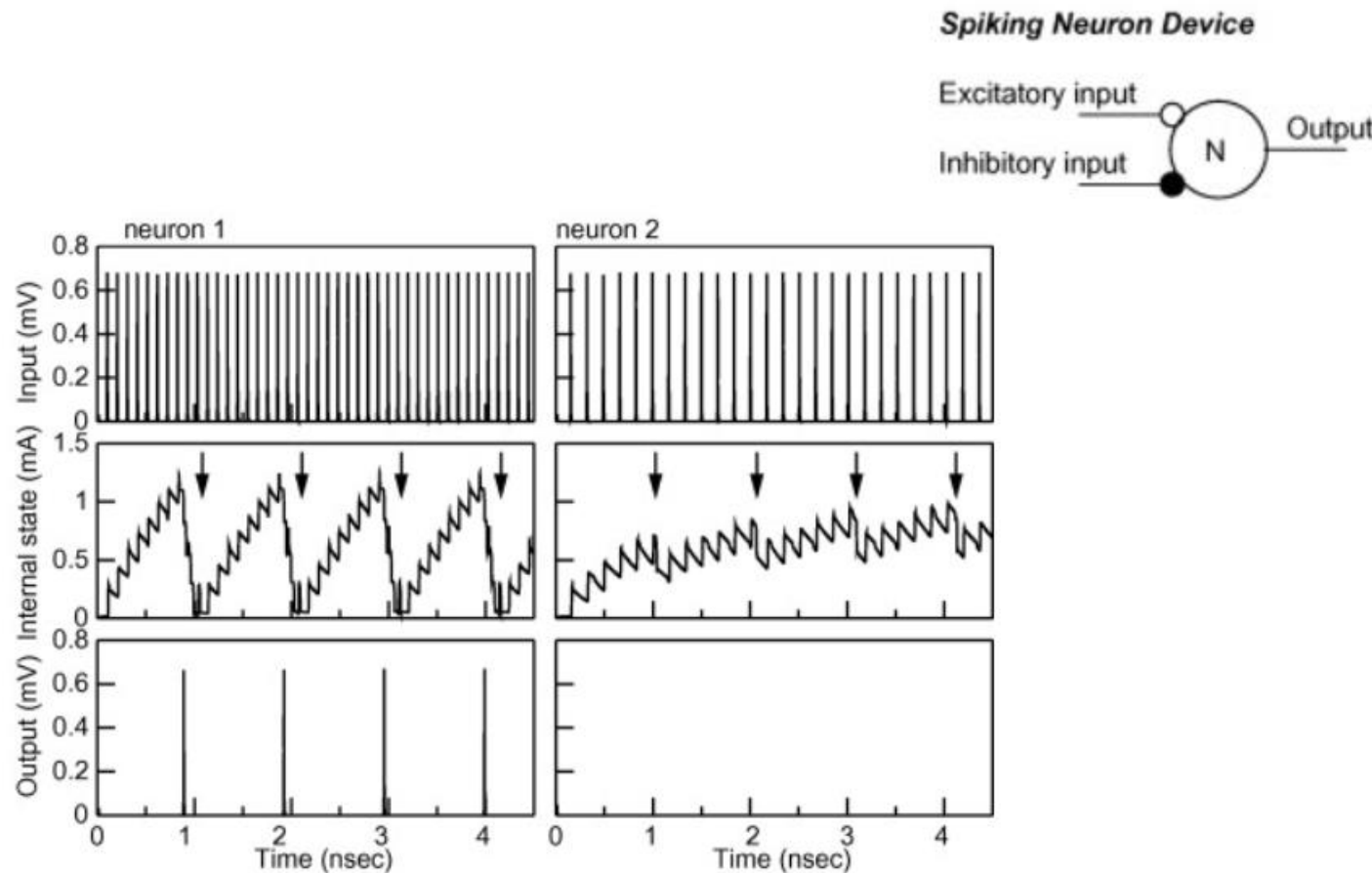
- Reinforcement learning architecture for SFQ based SNNs

Simulation

Spiking NN using SFQ neurons

- Inhibitory spiking neural network using SFQ circuits
- Spiking neurons coupled through all-to-all inhibitory connections.
- Simulated demonstration of winner take all neuron behavior

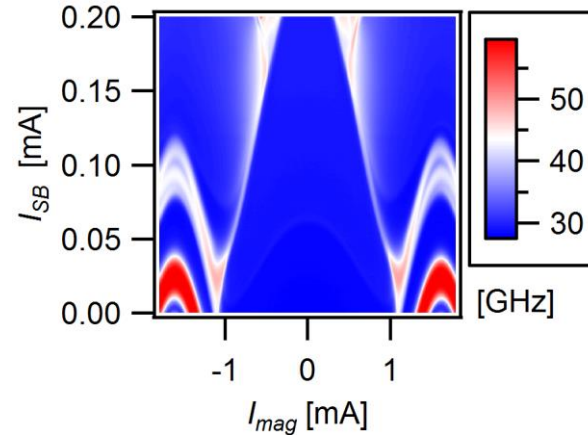
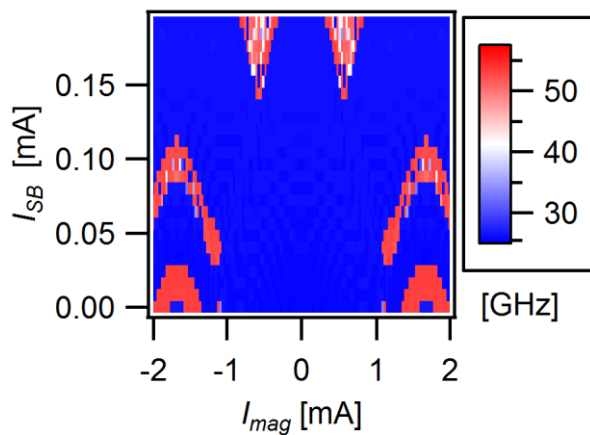
Hirose et al. Physica C 2006



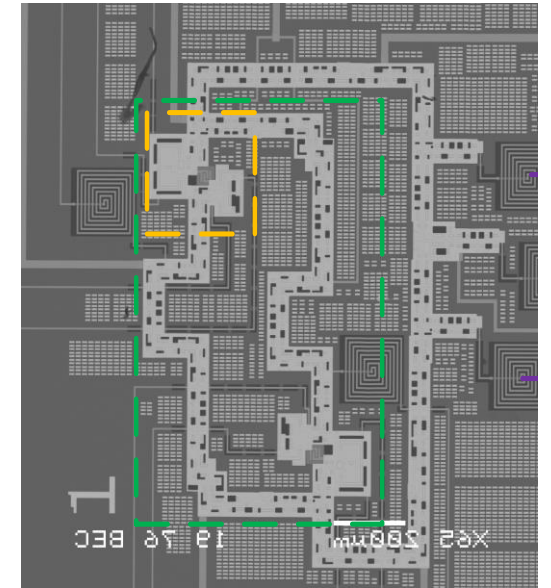
Experiment

2 coupled neurons

- JJ Neuron as a modified DC-SFQ circuit
- Shows biologically plausible behavior
- 2 – coupled JJ Neurons showing phase flip bifurcation, also seen in neuroscience



Blue = in-phase
Red = antiphase

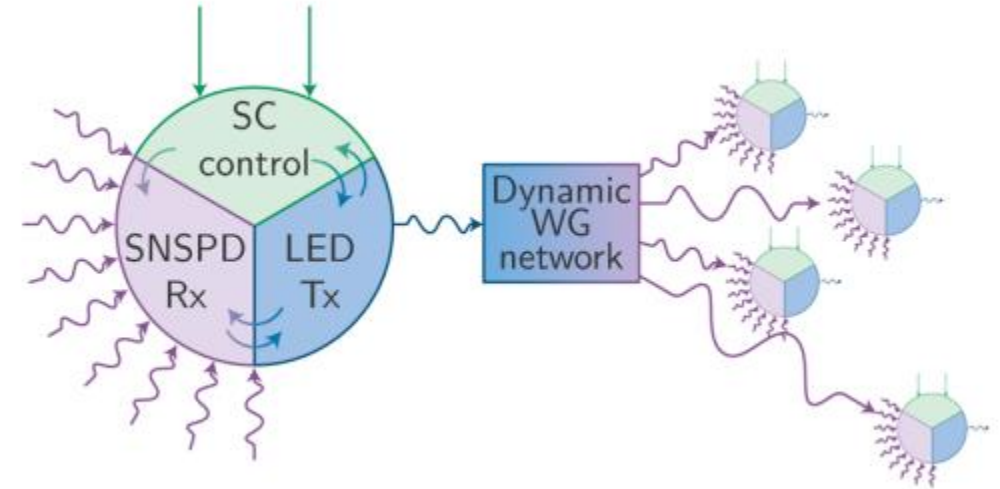


Crotty, Schult, and Segall, PRE (2010)
Segall et al. PRE (2017)

Experiment and simulation

Superconducting optoelectronic networks

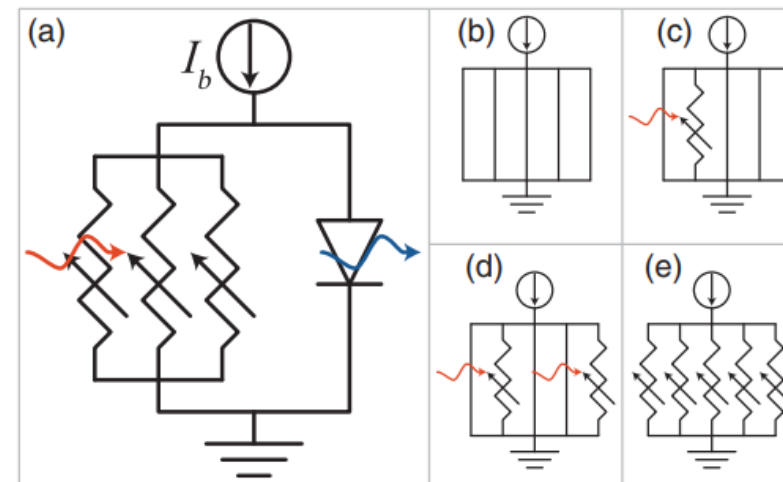
- Integration of superconducting circuits with photonics for ultimate scaling
- Large scale fanout
- Single photon sensitivity



Shainline et al. IEEE ICRC (2016)
Shainline et al. PR Applied (2017)

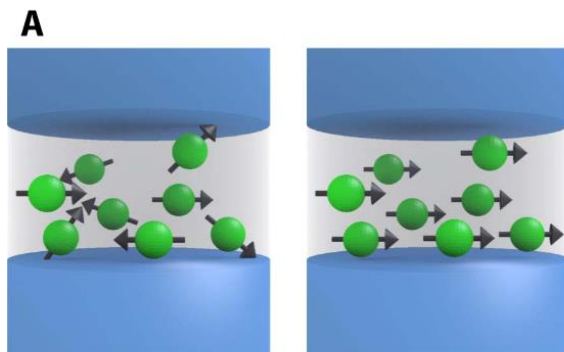
•
•
•

Shainline et al. arXiv:2409.18016 (2024)

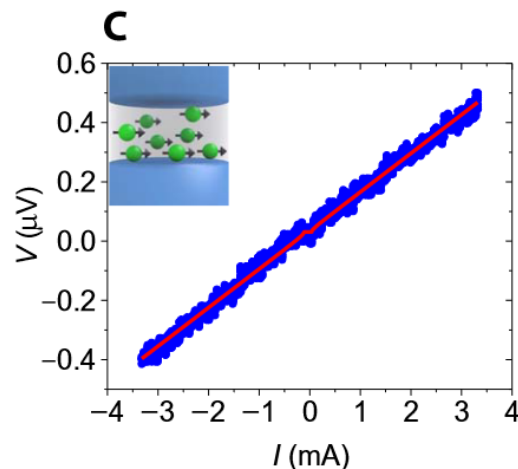
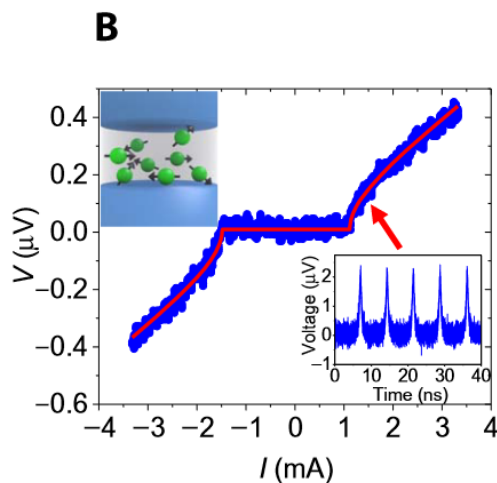
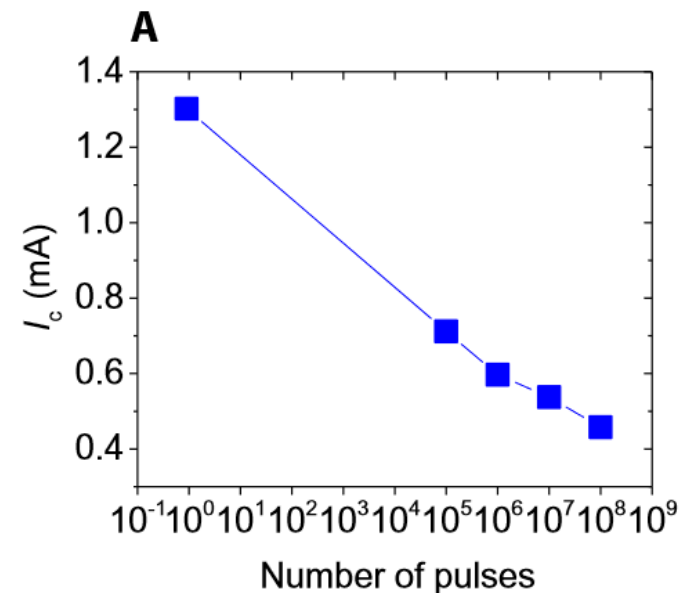


Experiment

Compact superconducting spintronic synapse

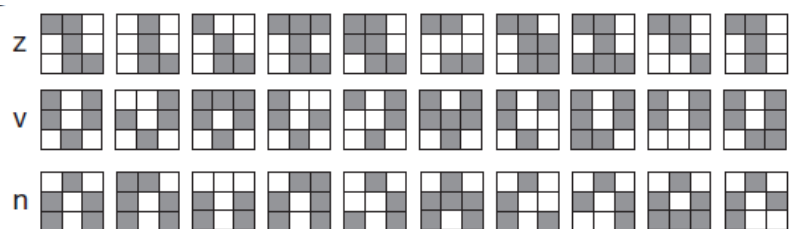
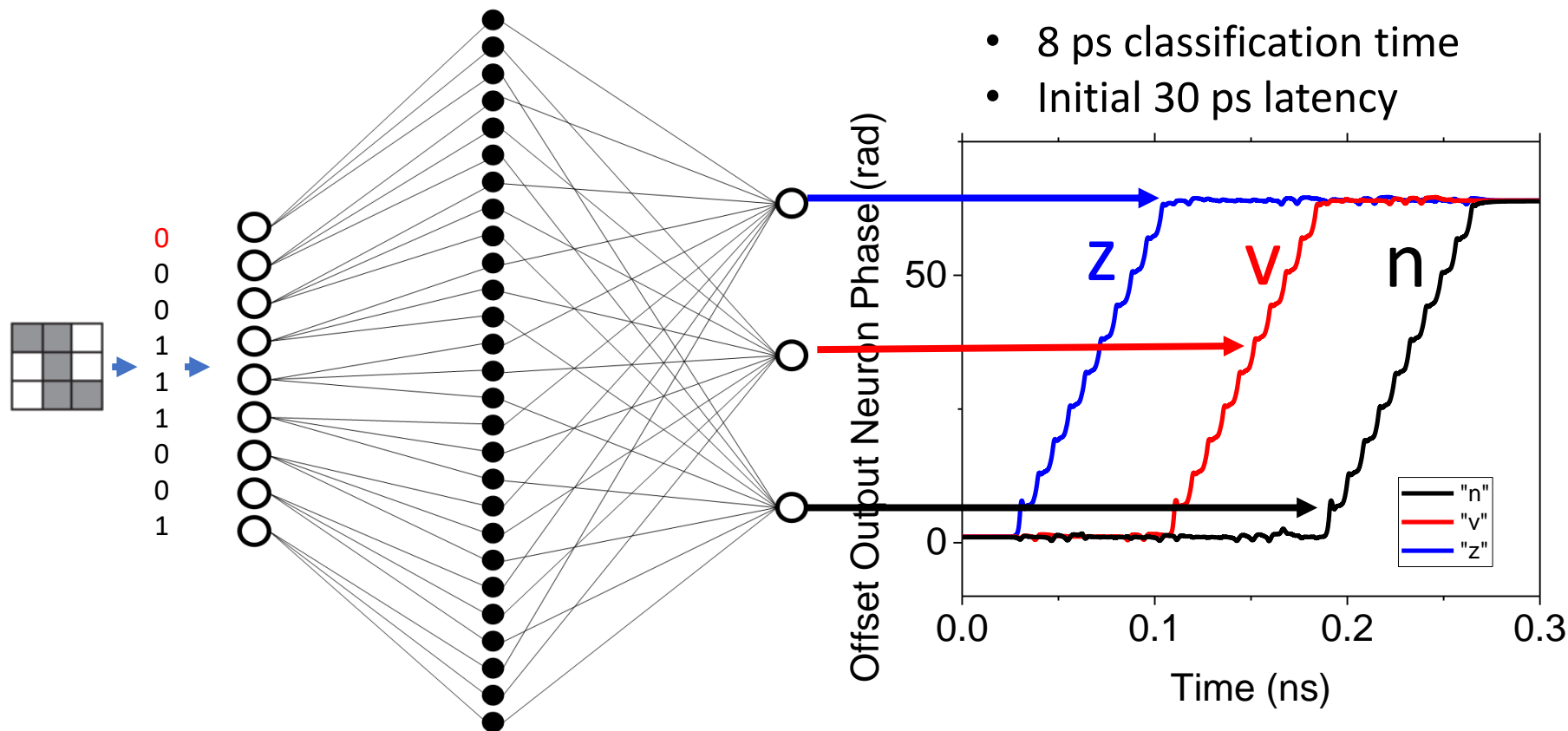


- Single device – tunable critical current
- Analogous to memristors
- 10,000 cluster/ μm^2
 - Potential for > 16 bit depth per device



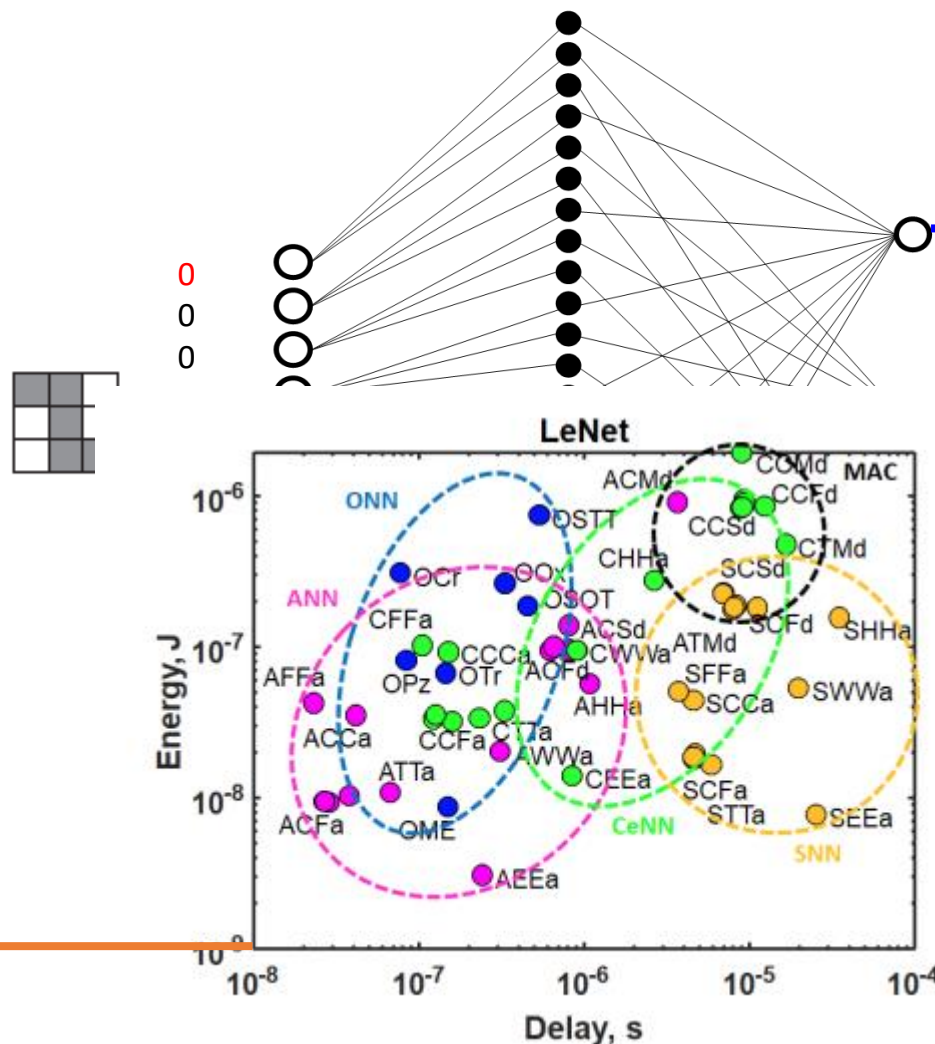
Russek et al. IEEE ICRC (2016)
Schneider et al. Sci. Adv. (2018)
Jue et al. JAP (2022)

Simulation programmable small scale 125 GHz 9-pixel JJ classifier

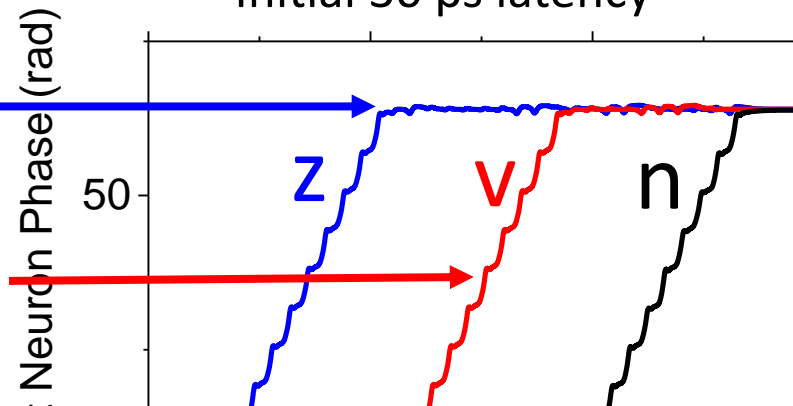


Schneider, et al. IEEE ICRC, (2017)
Schneider, J App Phys 128.21 (2020)

Simulation Comparison to other technologies



- 8 ps classification time
- Initial 30 ps latency



Calculated CNN architecture of LeNet to solve the MNIST benchmark from Nikonov and Young

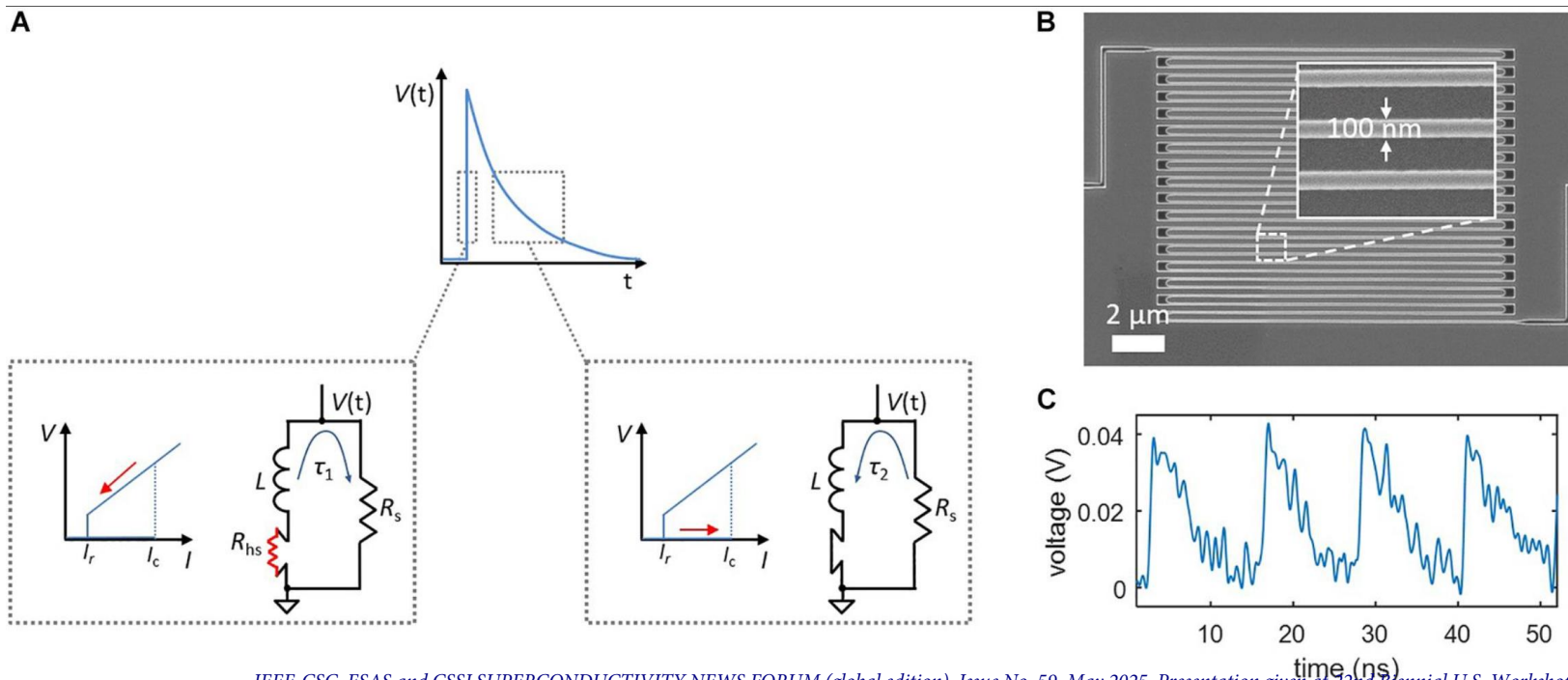
Very rough approximation for JJNN

- $1e-19$ J per element
- $1.3e7$ elements in LeNet
- $\Rightarrow \sim 1e-12$ J ($1e-9$ with cooling overhead)

Experiment

Nanowire neurons

- Superconducting nanowires demonstrated artificial neuron behavior
- The coupling of two nanowire-based oscillators acts analogously to the two ion channels in a simplified neuron model
- Neuron characteristics were simulated based on these devices and showed firing threshold behavior and bursting behavior
- Nanowire transmission line with a propagation speed of $\sim 2\% c$ may be used as an axon delay line,
- Potentially enabling easier spatio-temporal information processing



Toomey et al. *Frontiers in Neuro.* (2019)

Simulation

QPS Neuromorphic circuits

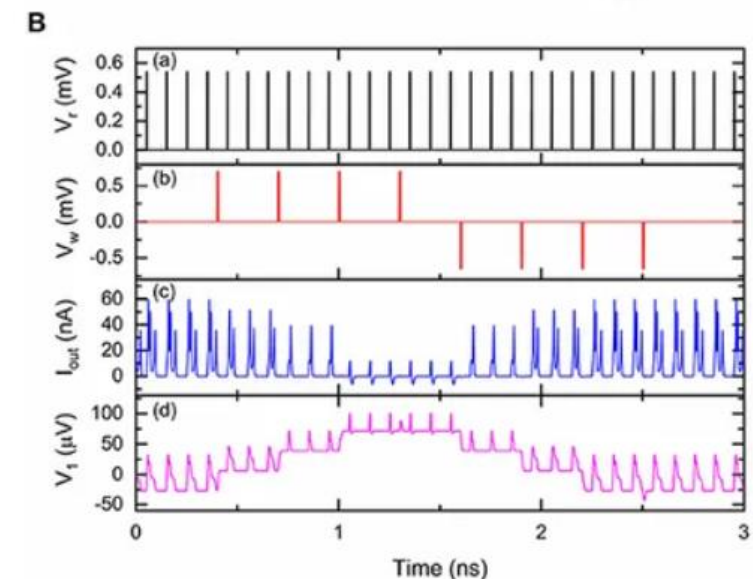
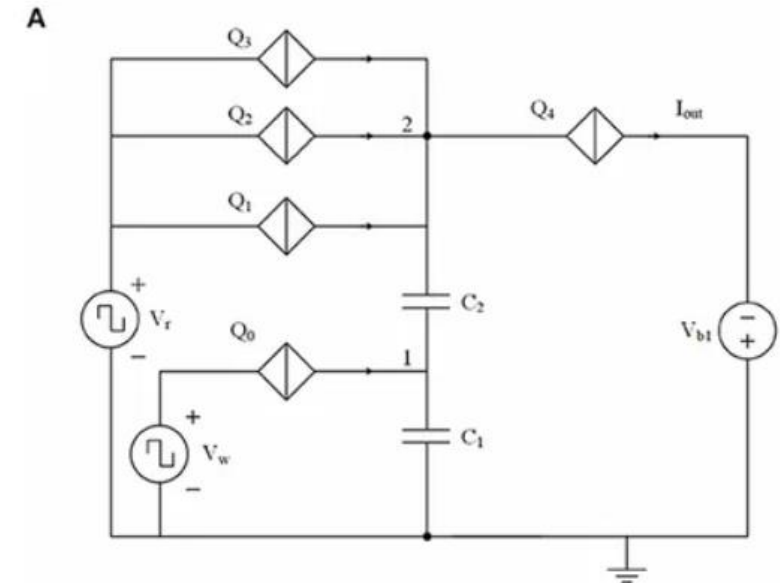
- Superconducting quantum phase slip junctions (QPSJs), an electromagnetic dual to Josephson Junctions (JJs)
- Demonstrated designs that mimic STDP learning behavior
- integrate-and-fire neurons
- multi-weight synapses
- fan-out mechanisms
- STDP can be achieved in individual synapses through the LTD and LTP circuits presented
- QPSJ-based coupled 3x3 synapse network
- Supervised and unsupervised learning mechanisms

Cheng et al. *JAP* (2018)

Cheng et al. *IEEE TAS* (2019)

Cheng et al. *IEEE TAS* (2021)

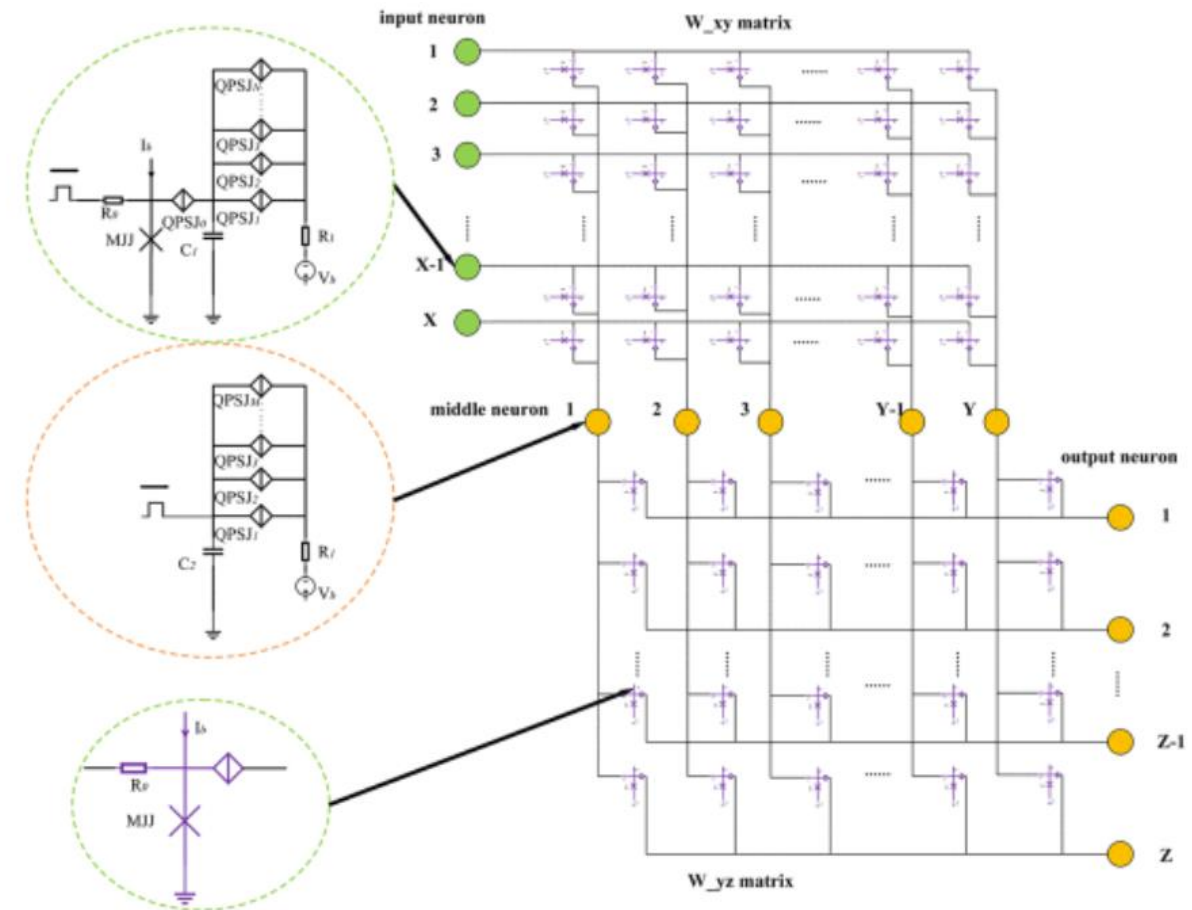
Cheng et al. *Frontiers in Neuros.* (2021)



Simulation

Rate encoded SNN with unsupervised learning

- Combined MJJ and QPS synapses and QPS neurons
- Frequency coded network
- Unsupervised learning based on STDP
- SFDP – spiking frequency dependent plasticity
- 3 layer fully connected SNN
- 0 – digit recognition tests

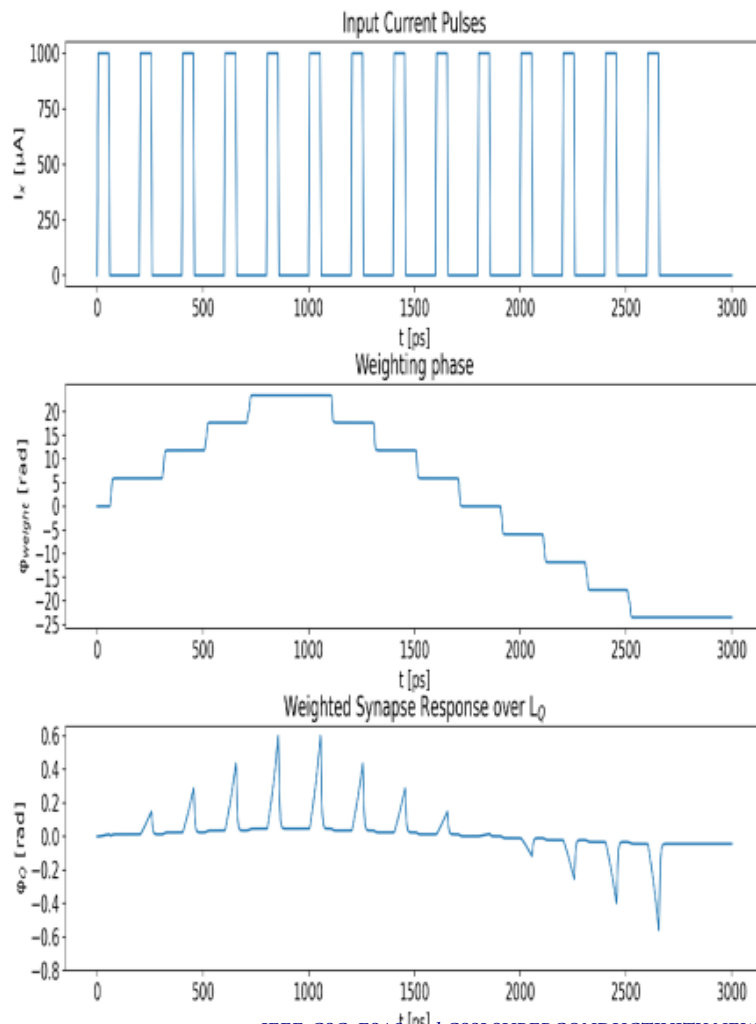


Zhang et al. IEEE Tans Emerging Topics in Comp Intelligence 2021

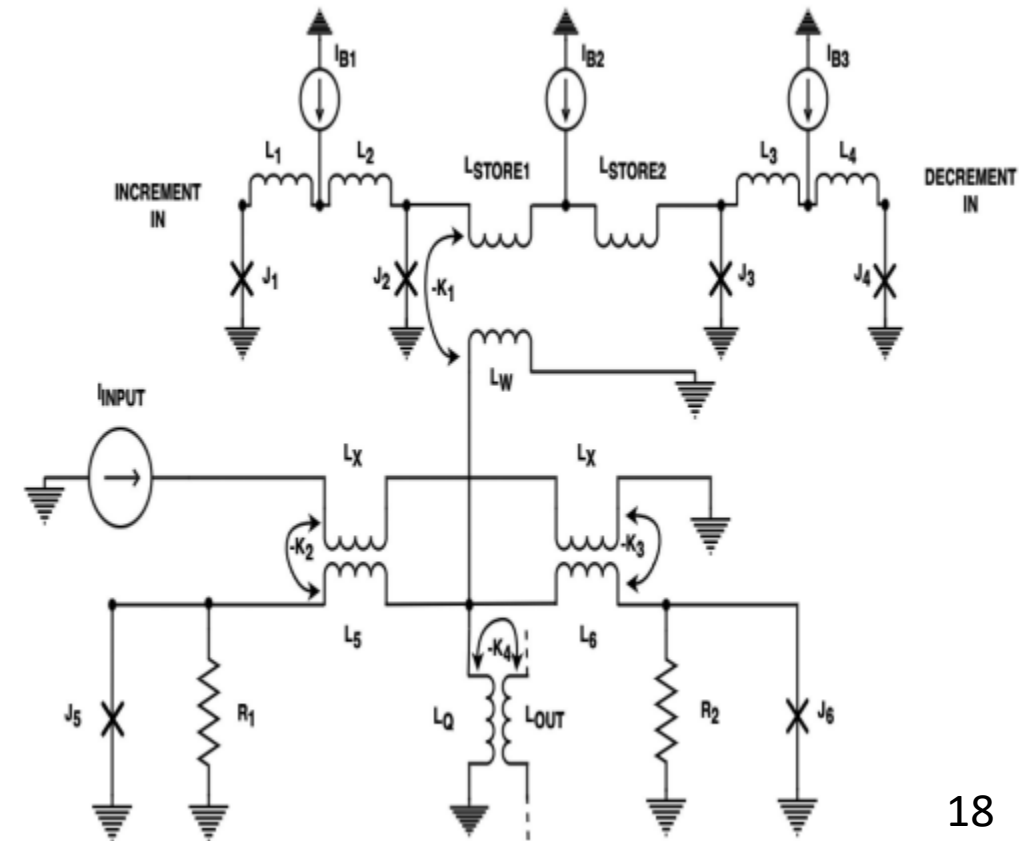
Simulation

Combine QFP SFQ cells and architecture

- Combine QFP and rSFQ architecture
- QFP to provide bi-directional weighting
- Perceptron learning algorithm
- Single layer Boolean logic demonstrated



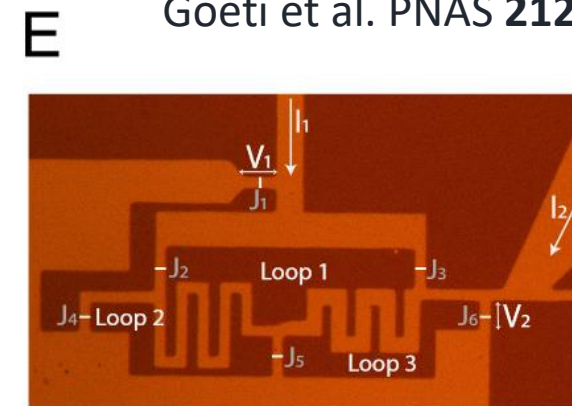
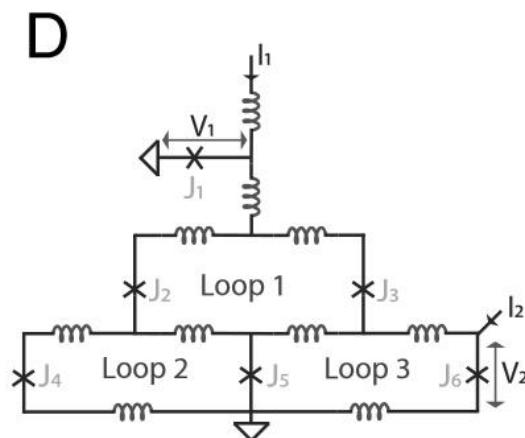
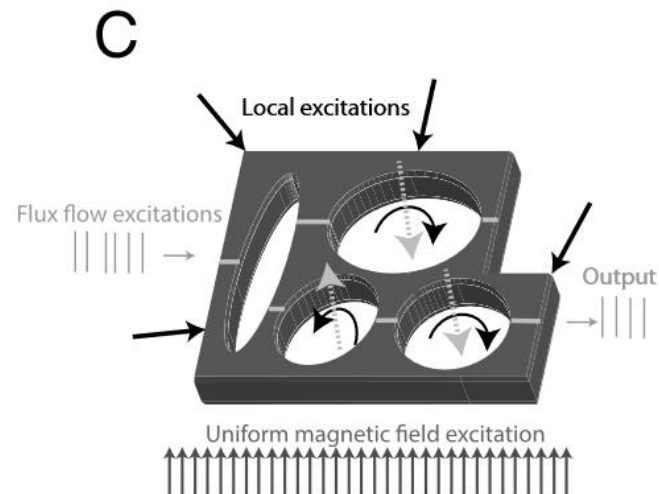
Jardine and Fourie
IEEE TAS 2023



Experiment

Associative memories

- Disordered network of superconducting loops
- Trapped flux memory states guide the flowing fluxons along different pathways
- Associative memory states are observed and can be programmed
- The flow of fluxons in these loop networks resemble the information flow in human brains with each having a different duration of dynamic stability.



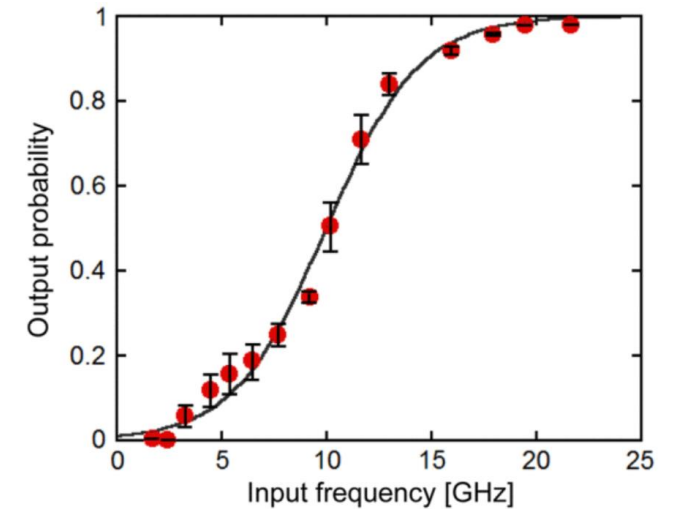
Goeti et al. Science Adv. (2022)

Goeti et al. PNAS **212** 2314995121 (2024)

Experiment and design

Neuromorphic units, Binary NNs

- Experimental demonstration Pseudo sigmoid function generator using an SFQ comparator
- Experimental demonstration Cryo-CMOS integrated with SFQ for BNNs
- Design of time domain convolution operation in SFQ BNN
- Design of Max pooling circuit for SFQ BNN
- Design of discrete Hopfield NN



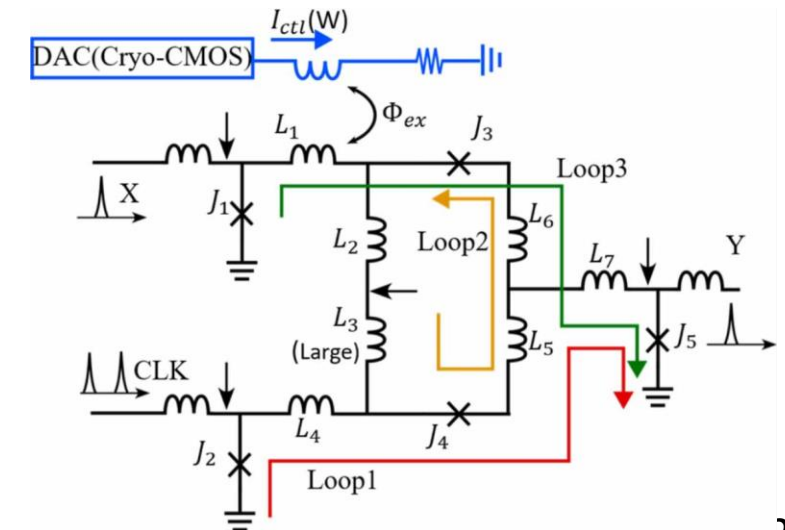
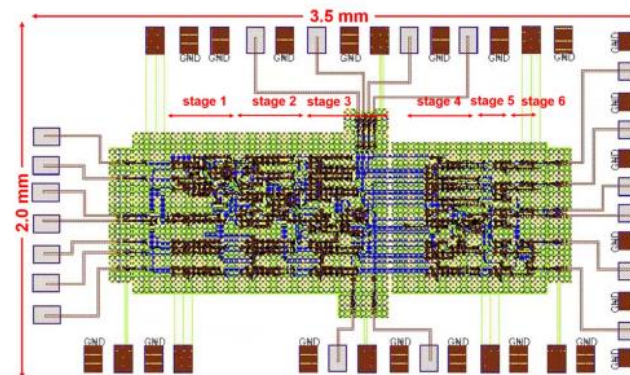
Zeyu et al. SS and T 2024

Li et al. SS and T 2024

Han et al. IEEE TAS 2023

Yamanashi and Yoshikawa IEEE TAS 2022

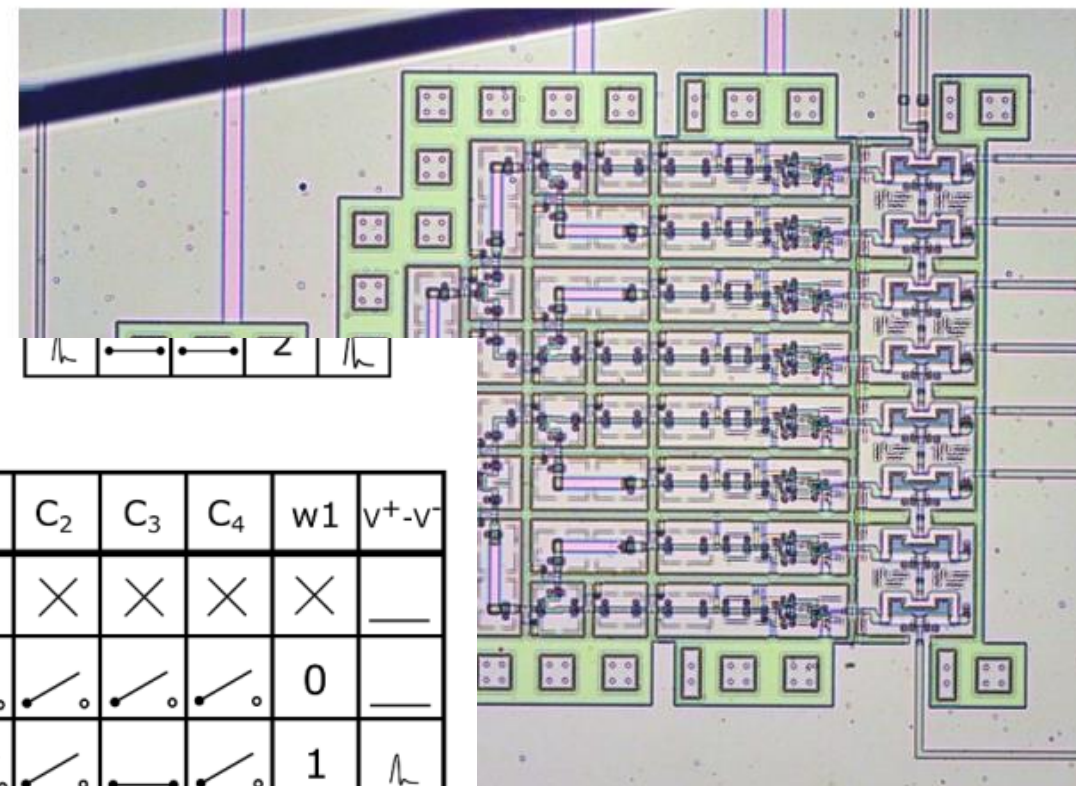
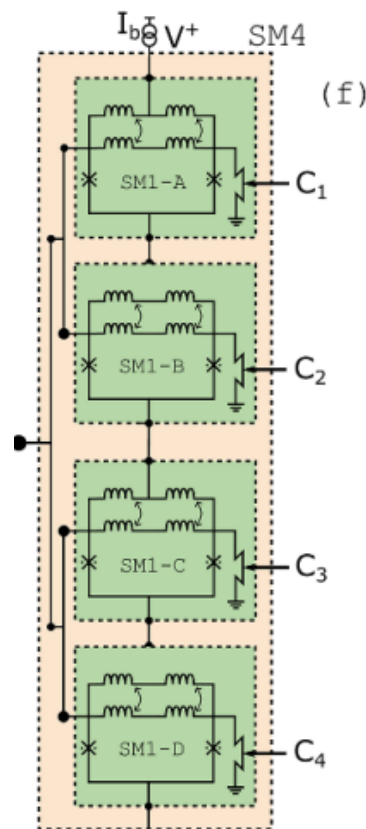
Yamanashi et al. IEEE TAS 2013



Experiment

Hybrid JJ-synapse programmable with cryo-CMOS

- JJ-Synapse uses a series SQUID circuit with no static power consumption
- Synapse weights are adjustable via an in-situ Si-Ge CMOS circuit
- The JJ-Synapse can function up to 20 GHz and have power consumption in the order of attojoule



P1	C ₁	C ₂	C ₃	C ₄	w1	v ⁺ -v ⁻
—	×	×	×	×	×	—
h	↗	↗	↗	↗	0	—
h	↗	↗	—	↗	1	h
h	—	—	↗	↗	2	h
h	↗	—	—	—	3	h
h	—	—	—	—	4	h

Karamuftuoglu, Bozbey, and Ozbayoglu.
IEEE TAS (2023)

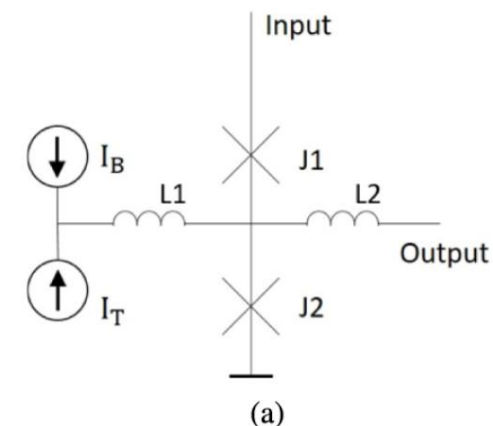
Karamuftuoglu et al. *IEEE TAS* (2023)

Razmkhah et al. *SS and T* (2024)

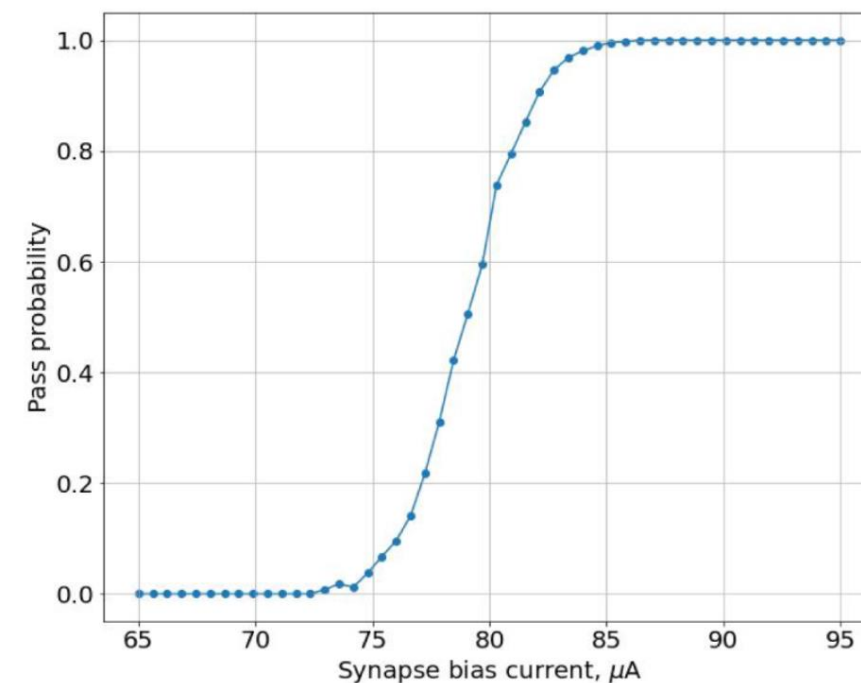
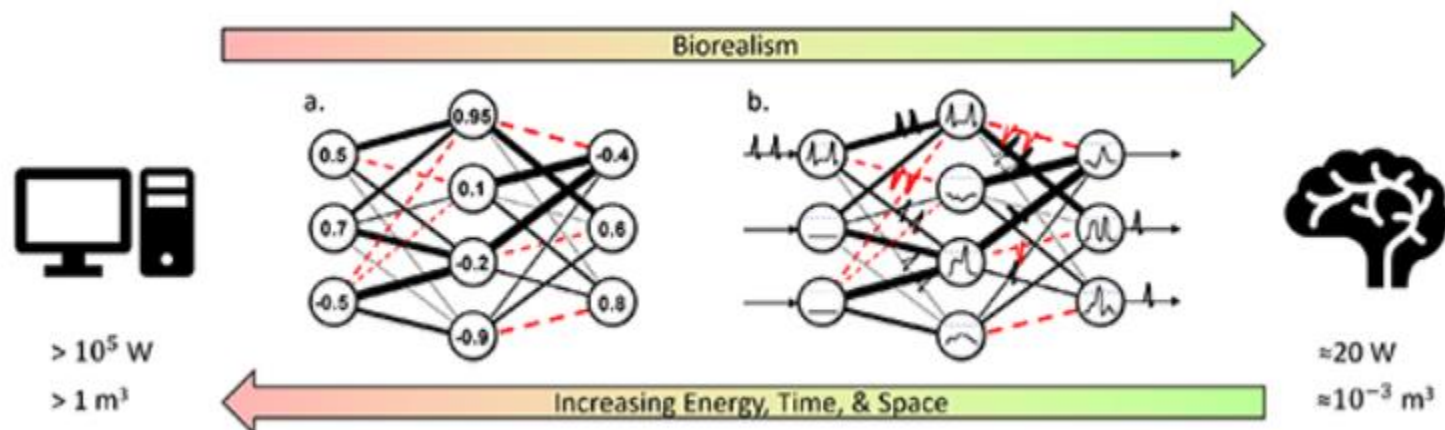
Simulations

Stochastic synapses with rate coded architecture

- Biorealism to improve efficiency
- Stochastic synapse
- Rate coded architecture
- Programmable or hard coded weight values
- XOR level network simulated



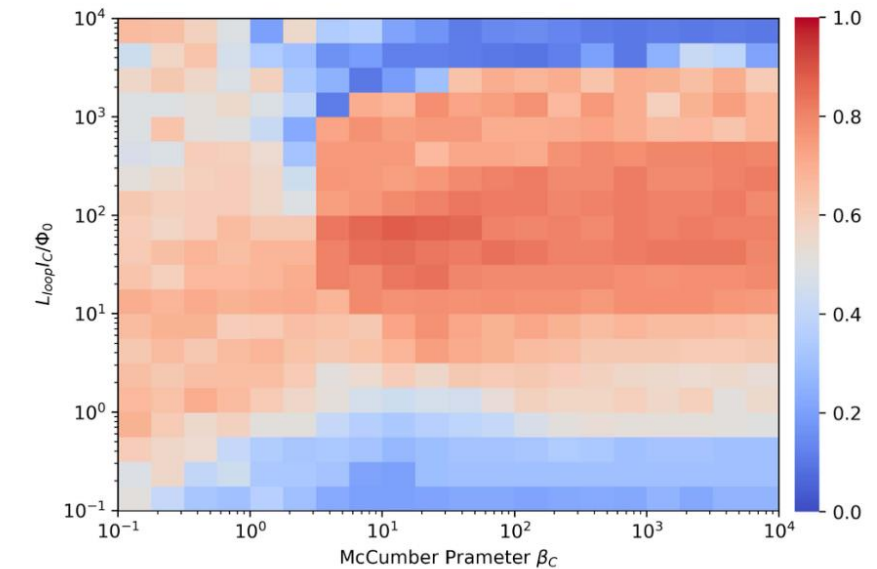
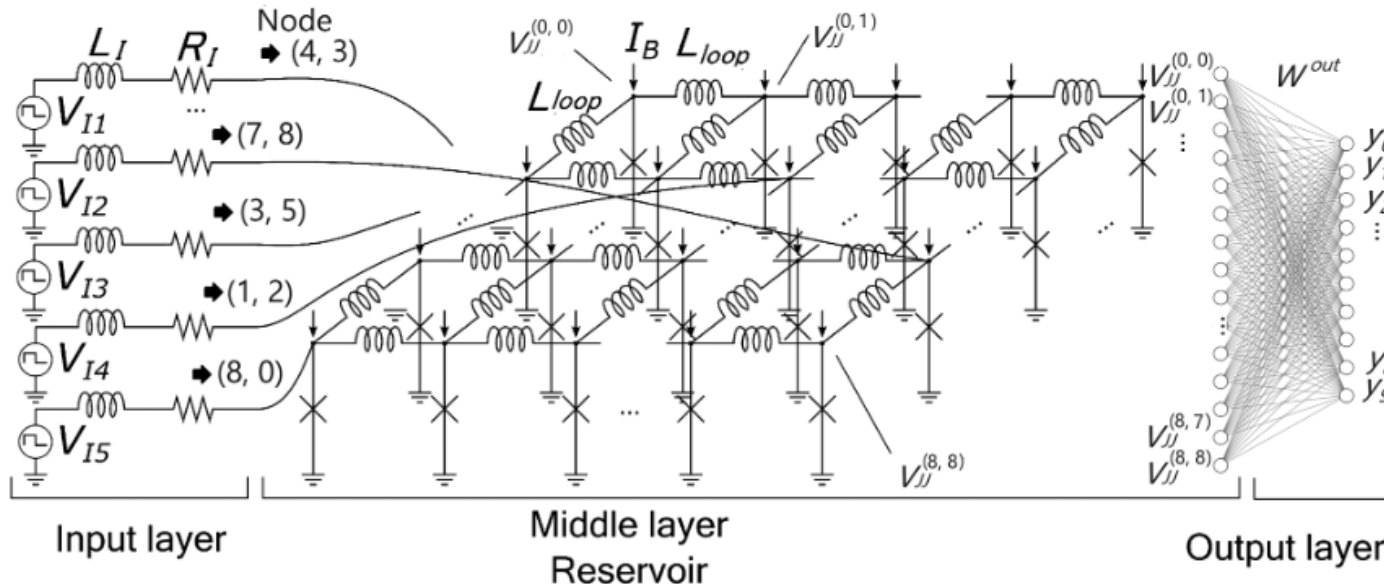
Edwards et al. SS and T (2024)



Simulation

Reservoir computing

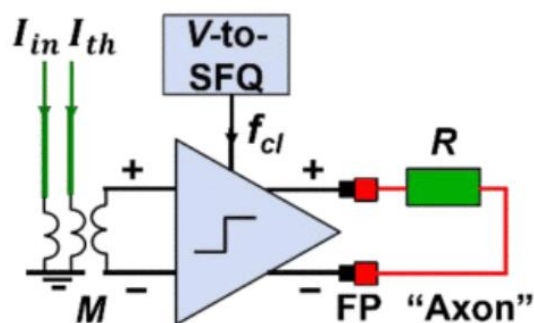
- Simulation of 9x9 pixel classifier
- Recognizing 10 digits
- Reservoir computing does not need explicit weight programming



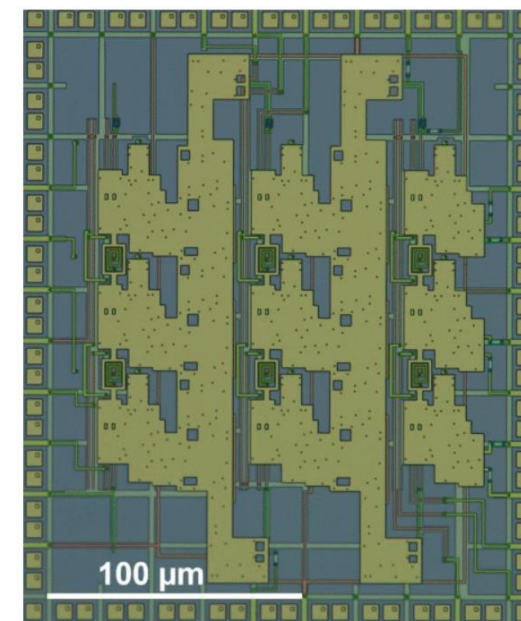
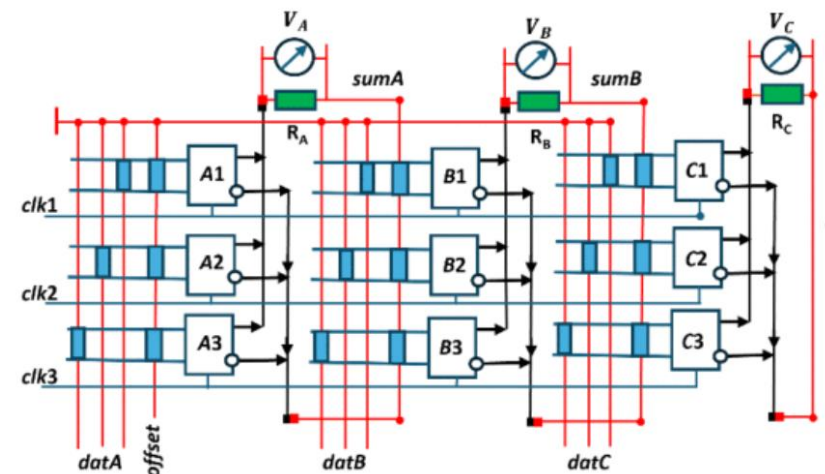
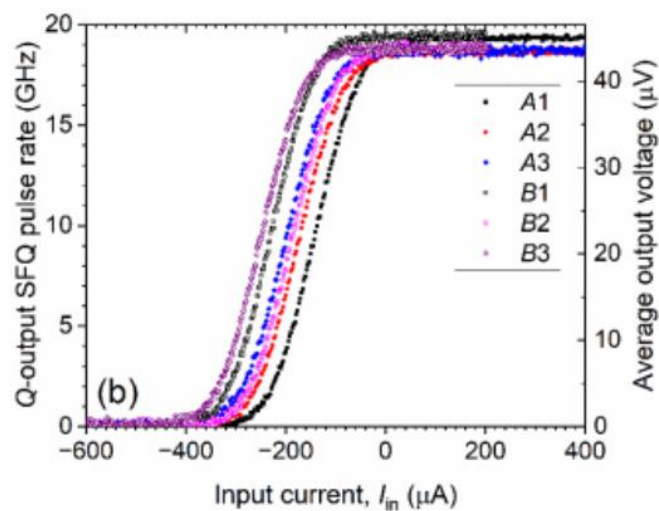
Watanabe et al. *IEEE TAS* (2024)

Experiment

3 x 3 analog ANN



- Continuous spiking regime
- Analog representation of multiply accumulate function that underpins algorithmic NN
- Proposed memory / fixed weight circuit has been fabricated

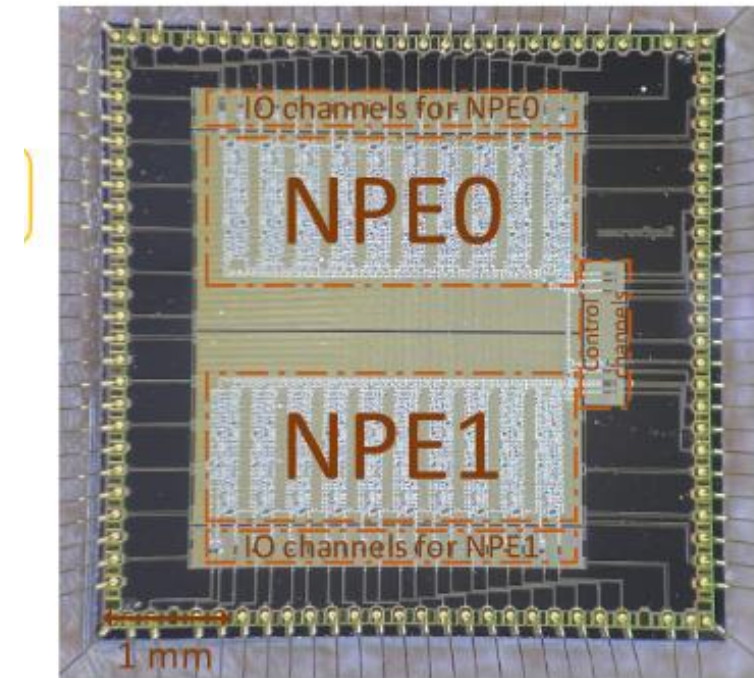


E. B. Golden, V. K. Semenov, S. K. Tolpygo, IEEE TAS (2025)
Semenov et al. IEEE TAS (2023)
V. K. Semenov, et al. IEEE TAS (2022)

Experiment

Digital SFQ SNN architecture

- Asynchronous spiking neural network
- Weight encoded with number of pulses
- Use of splitters and confluence buffers for changing the number of pulses
- NDRO used for programing weight changes and inference vs. training operation modes
- Two physical neurons demonstrated
- Simulations extrapolated to MNIST levels
- 45,000 JJ's for a 4x4 mesh network



Z. Liu, et al., *IEEE TAS* (2025)

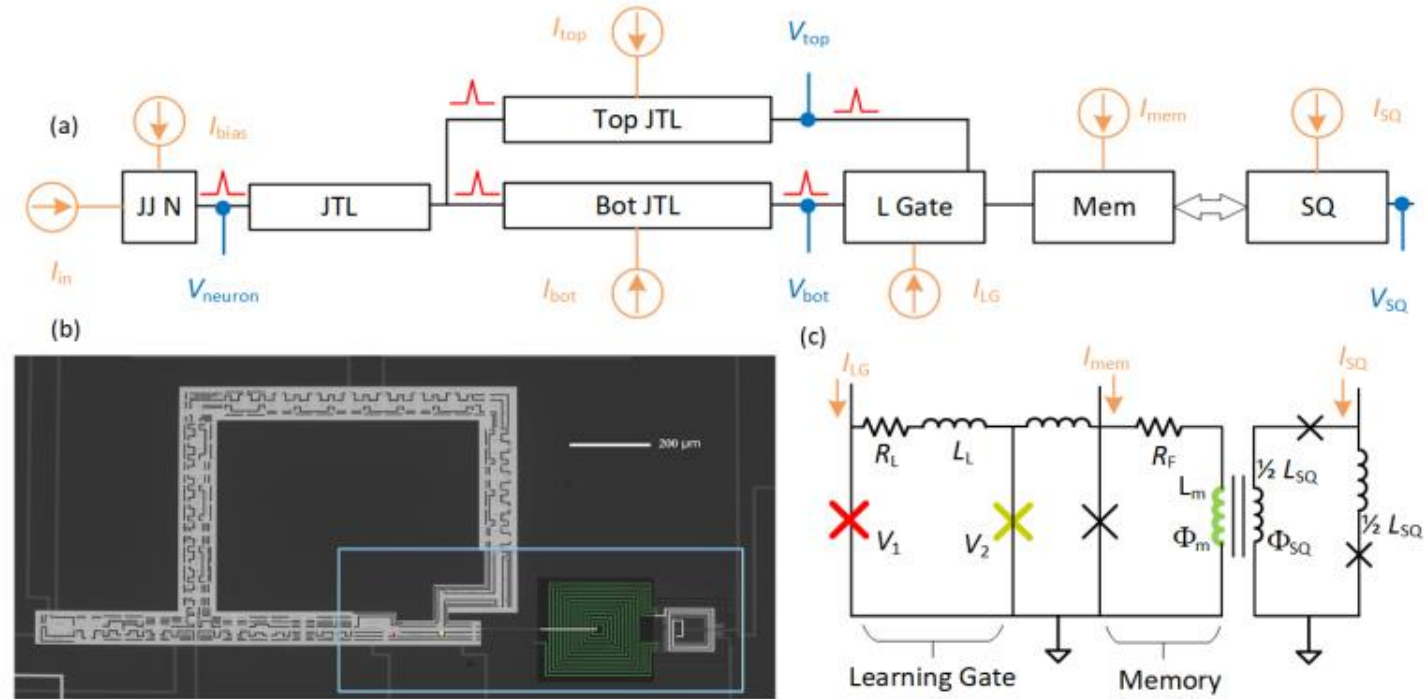
Liu et al. *MICRO '23 proceedings* (2023)

CB		SPL		NDRO	
dinA/B-dinA/B	19.9	din-din	19.9	din/rst-rst/din	39.9
dinA/B-dinB/A	5.7	TFF		clk-clk	39.9
DFF		din-din	19.9	din-clk	14.81
din-clk	8.53	JTL		rst-clk	16.61
clk-clk	19.9	din-din	19.9		

Experiment

Spike timing dependent plasticity

- Nearly all of the Izhikevich Behaviors were demonstrated in simulation
- Spike timing dependent plasticity
- Often called Hebbian learning
- Shows biologically plausible behavior
- Potential for spiking neural networks that support unsupervised learning



Crotty et al. *IEEE TAS* (2023)

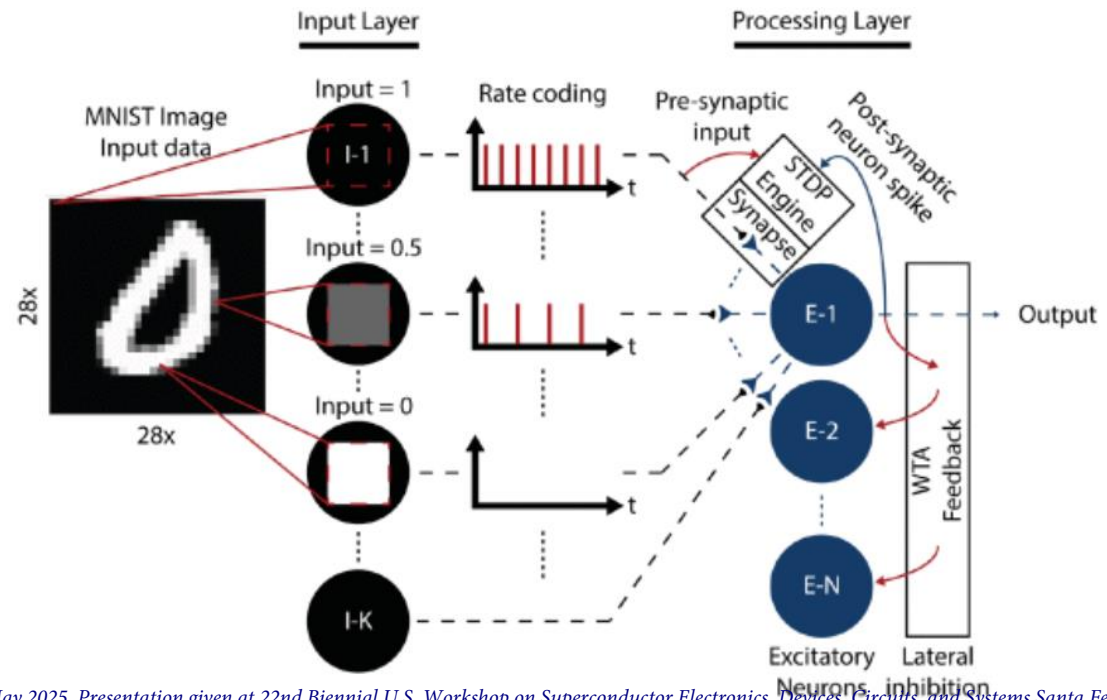
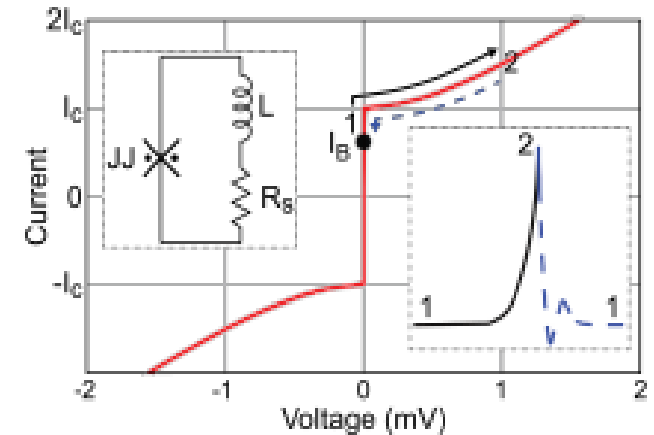
Segall et al. *Appl. Phys. Lett.* (2023)

Segall et al. arXiv:2504.02754 (2025)

Simulation

Neuron, synapse designs and NN architectures

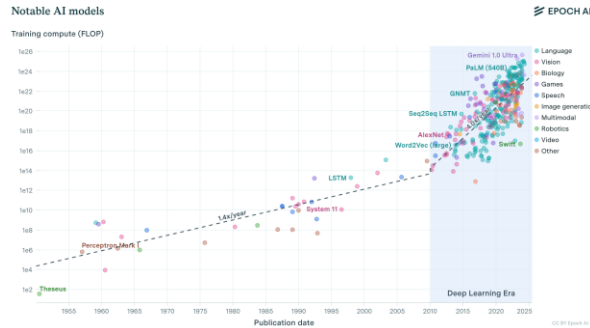
- High-fan-in superconductor neuron
- Neuron with ternary synaptic connections
- STDP with “leaky NDRO”
- Winner take all neuron connections
- Unsupervised SFQ NN
- Extended python simulations to MNIST with 96% accuracy



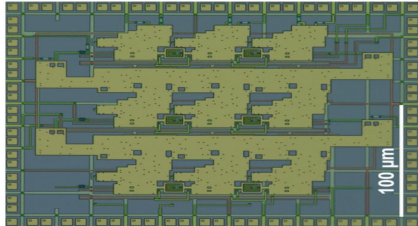
Karamuftuoglu and Pedram. *IEEE TAS* (2023)

Karamuftuoglu et al. *SS and T* (2025)

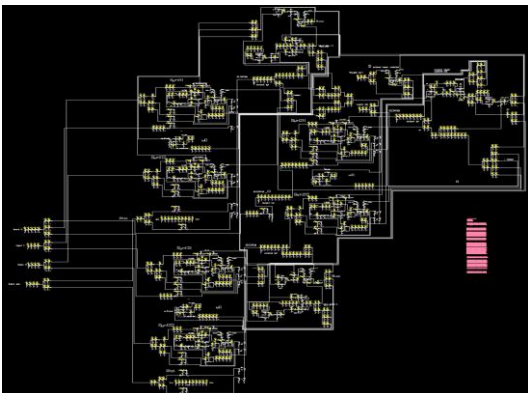
Karamuftuoglu et al. *IEEE TAS* (2024)



- Introduction / why neuromorphic

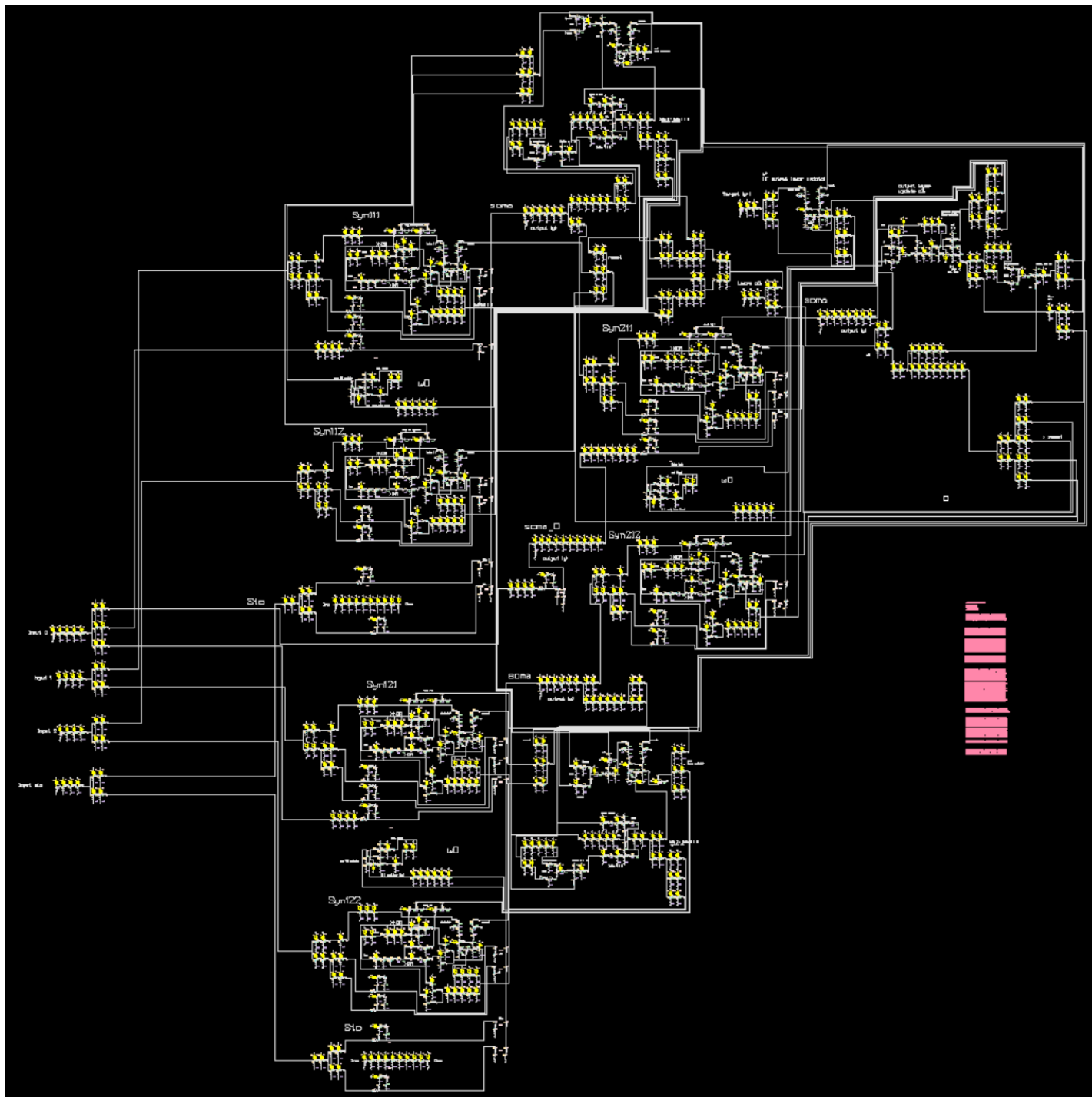


- Survey current state of the field



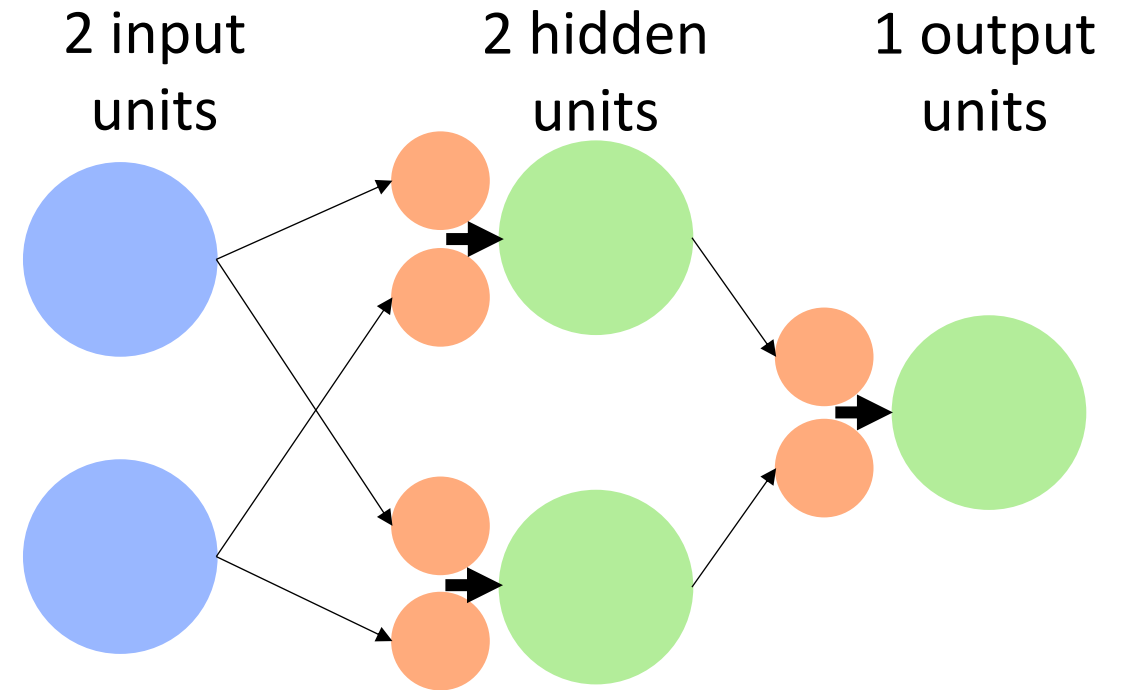
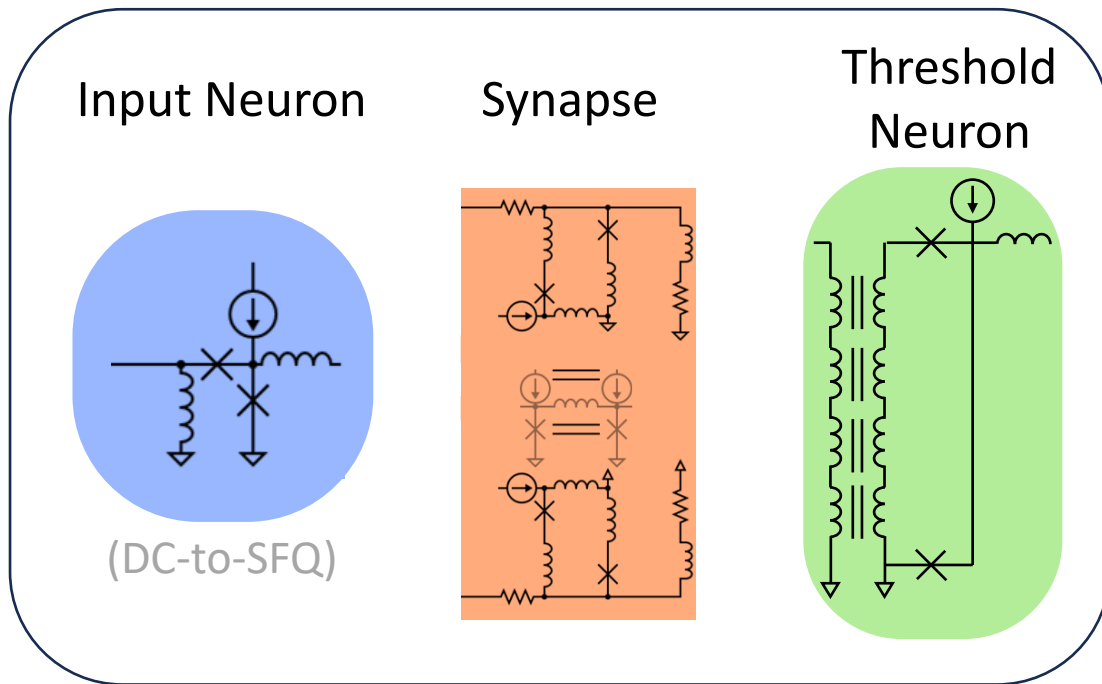
- Reinforcement learning architecture for SFQ based SNNs

WRSPICE simulation of a fully self-training 5 node network



5 node network

Building blocks



Schneider et al. NPJ Unconventional Computing (2025)

RL learning rules:

After each iteration:

1) Check:

- Did the reward get better or worse:

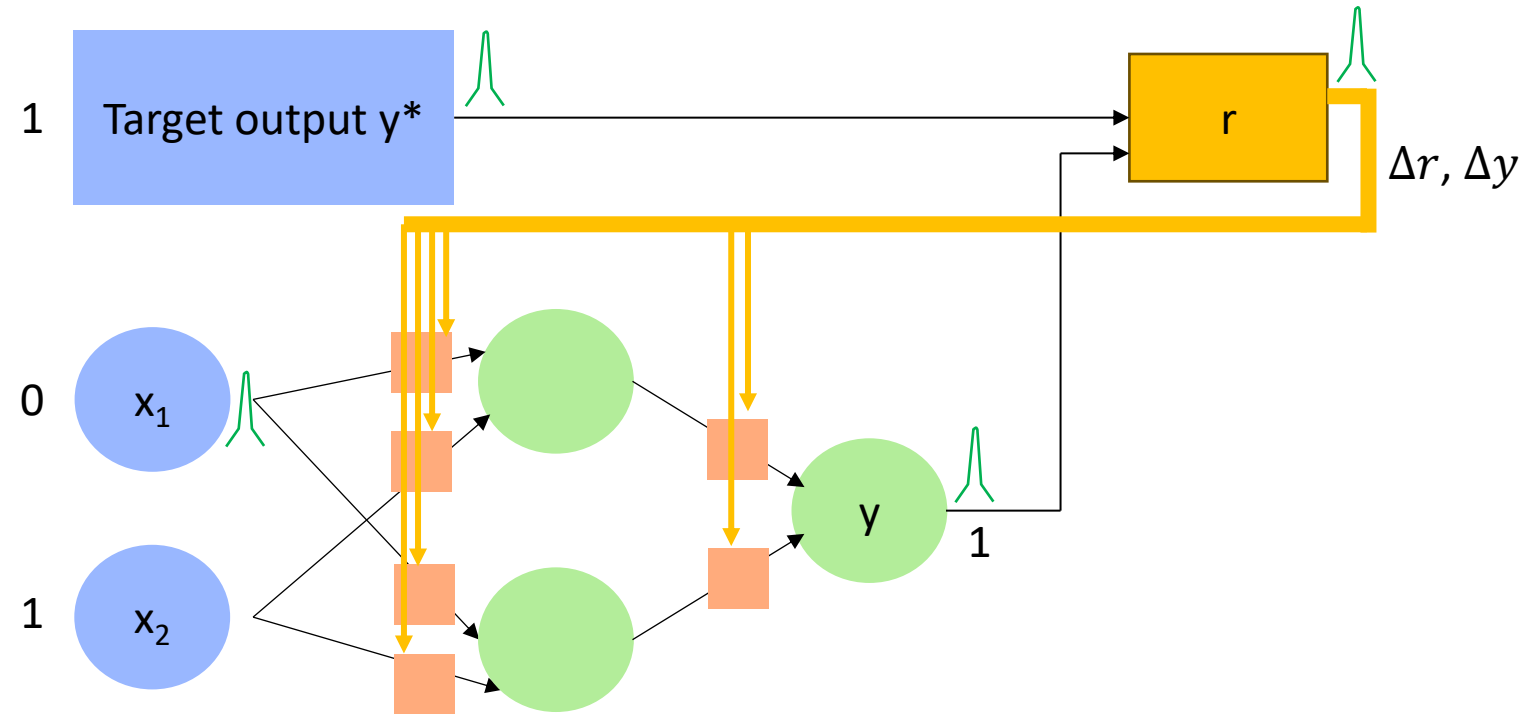
$$\Delta r = r^{now} - r^{previous}$$

- Did the output change:

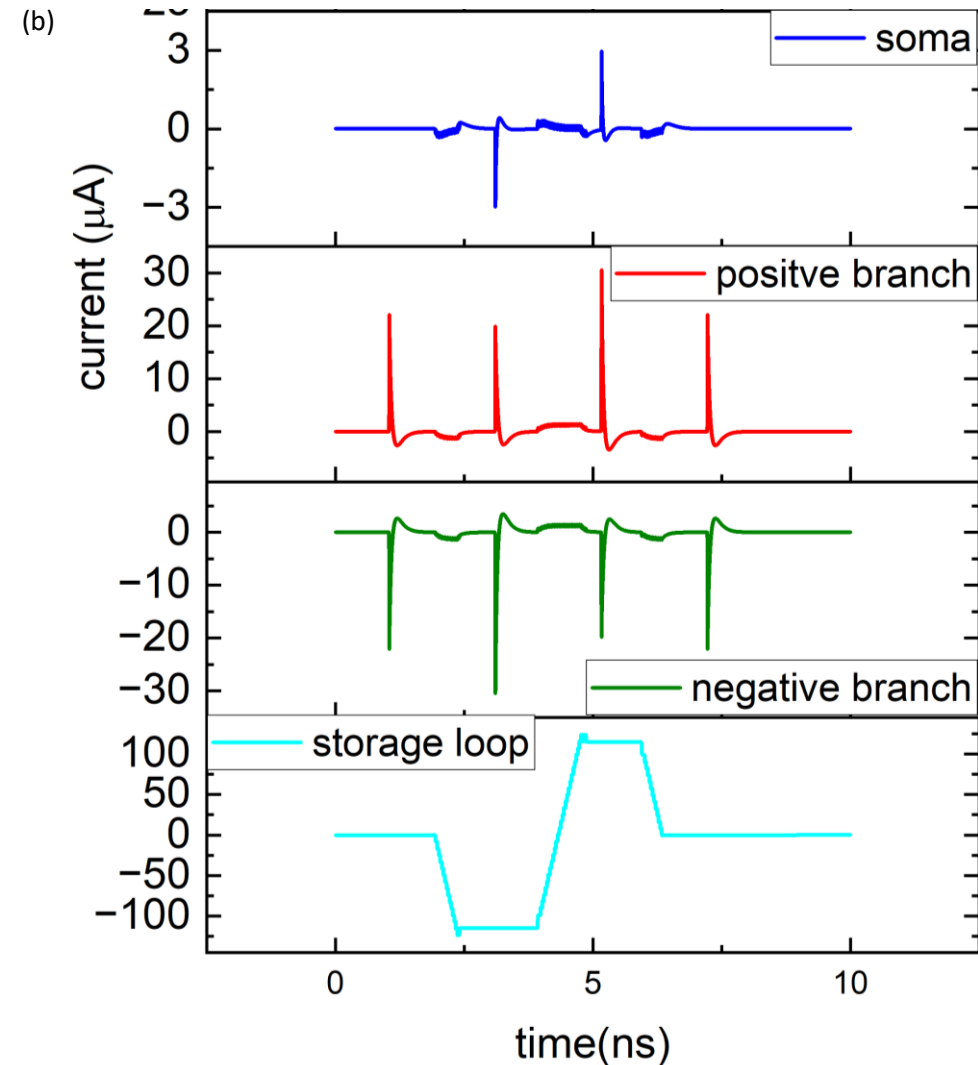
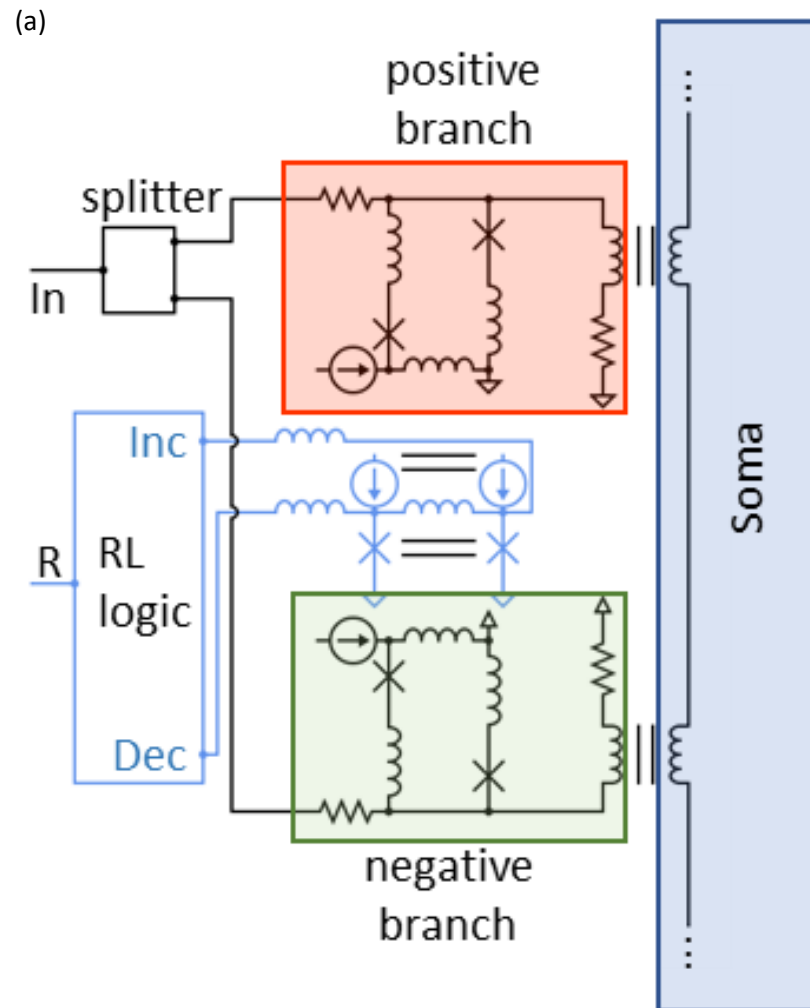
$$\Delta y = y^{now} - y^{previous}$$

2) Update synaptic weights:

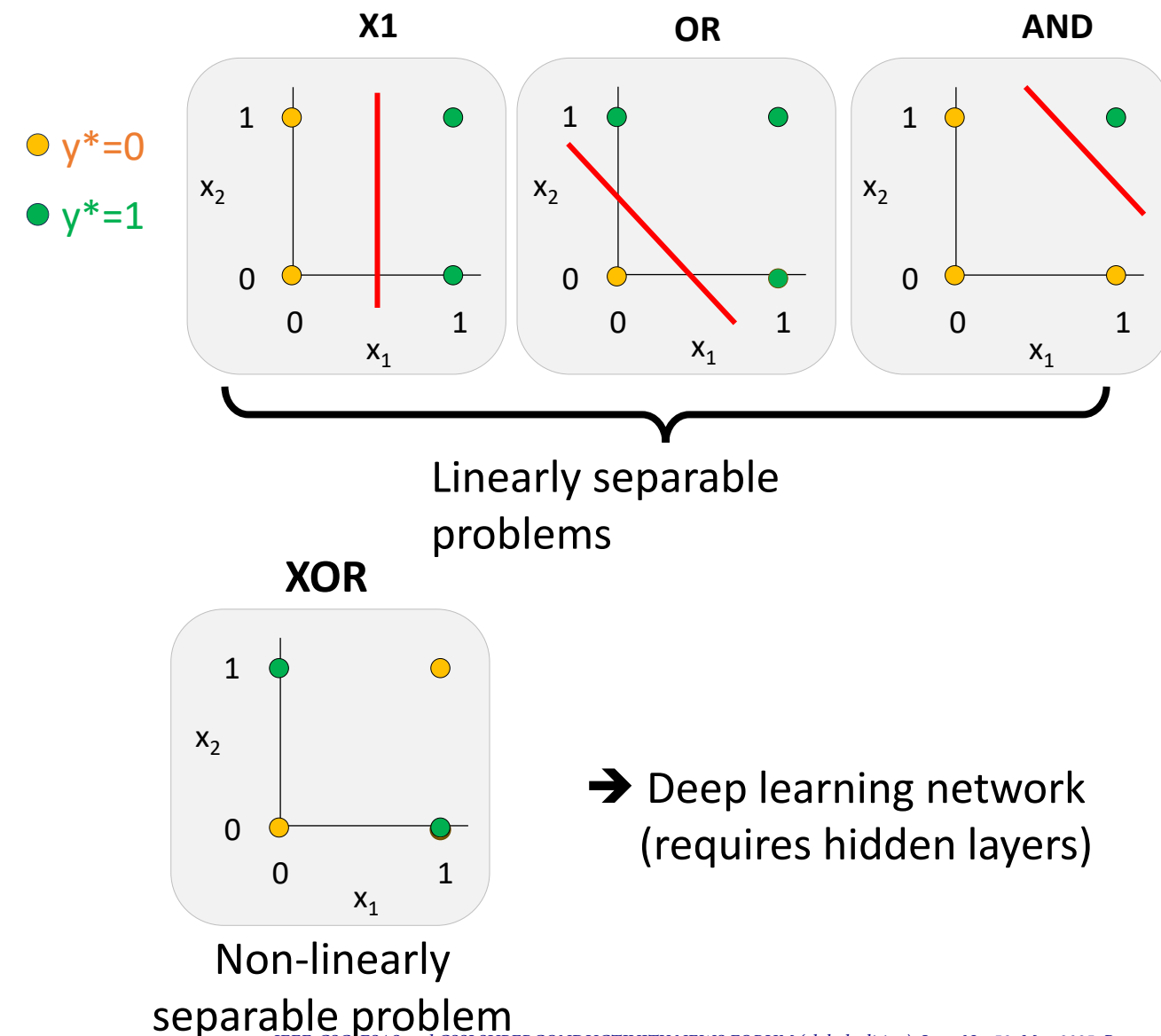
$$w_i^+ = \rho \Delta r \Delta y (x_i - 0.5)$$



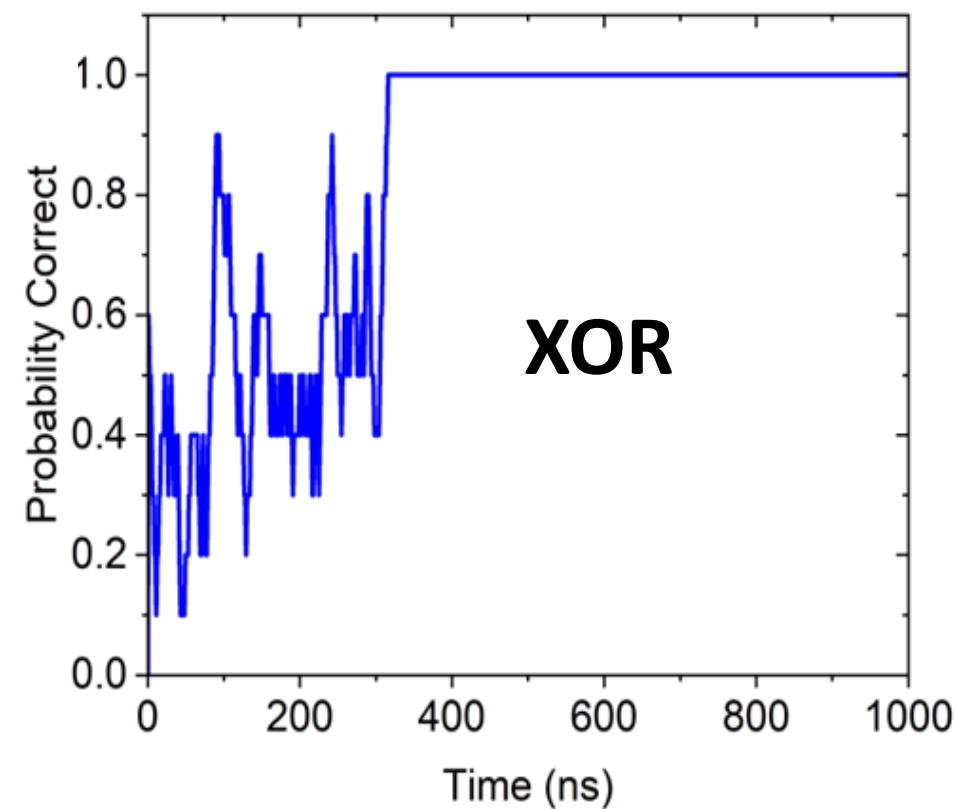
Bipolar JJ synapse (SPICE simulations)



Deep network learning



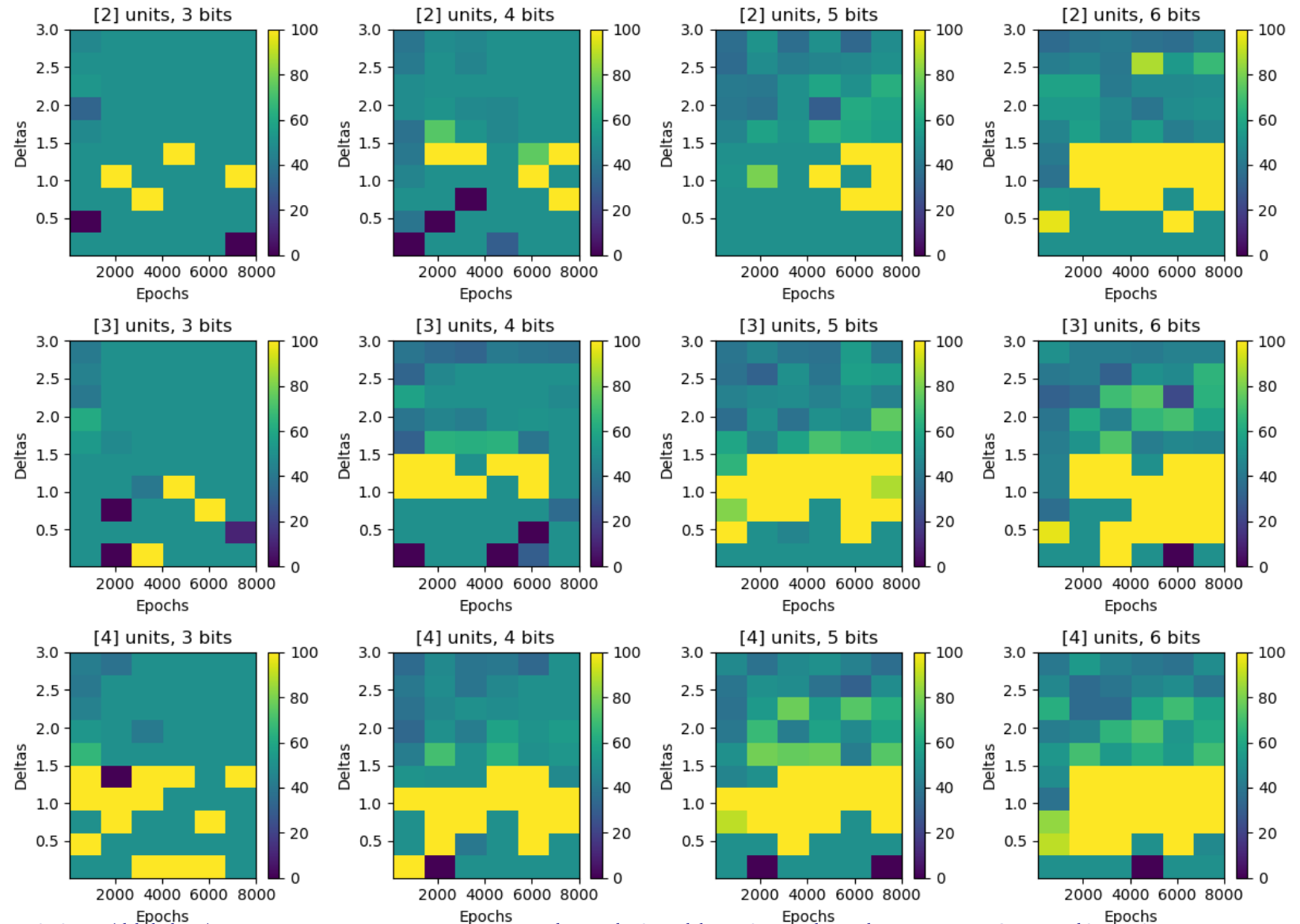
~ 20% stochastic weight exploration added



Demonstration of general deep network learning

Synaptic bit depth

- XOR training python simulations
- Weight depth matters for consistent training
- Stochastic weight exploration has a relatively consistent ideal range
- Extra hidden units also help / can be traded for bit depth

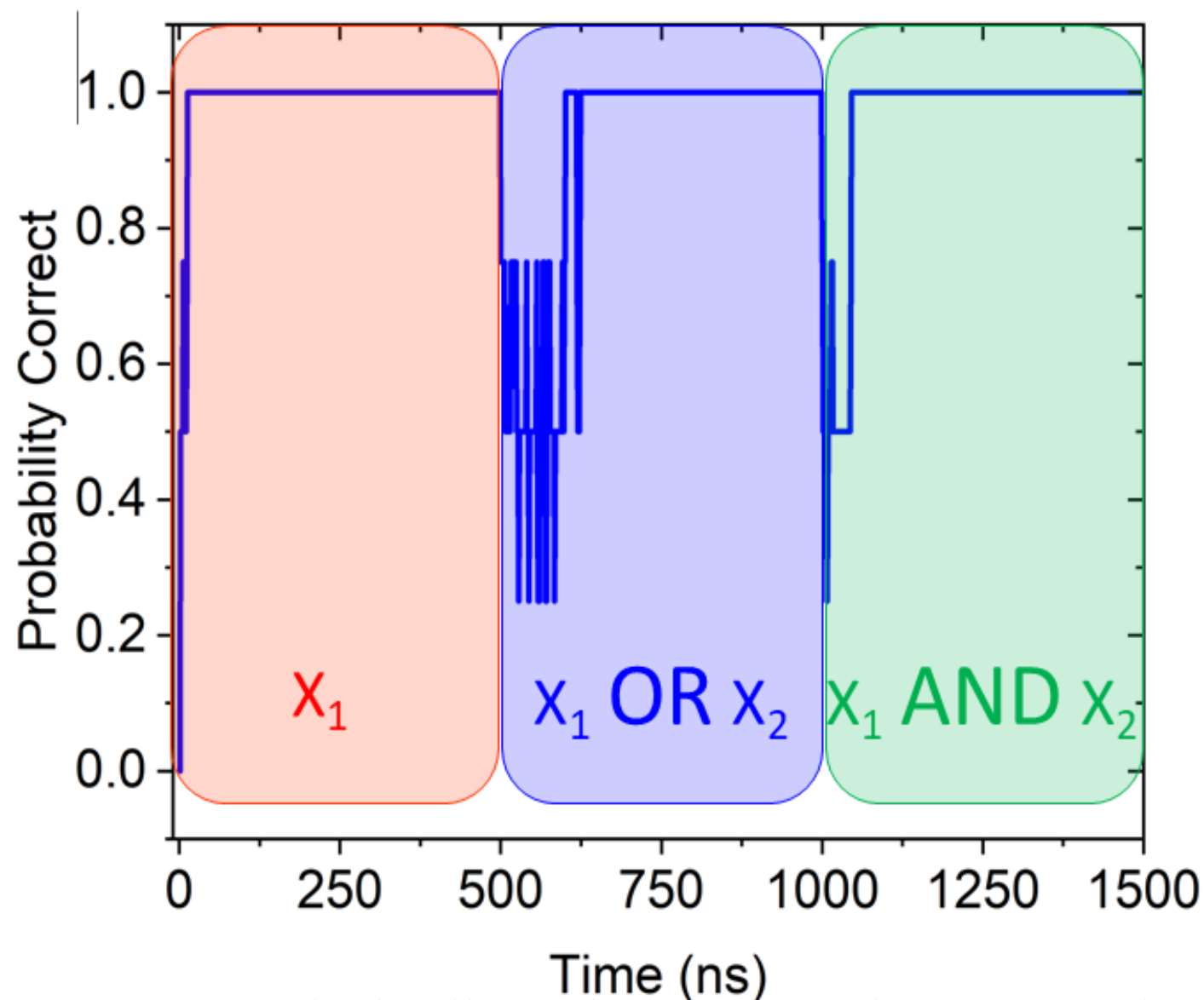


Continual Learning

Continuous random input stream
for entire 1.5 ns simulation

Only change the desired output
 y^*

Weight updates stabilize at 100%
correctly



Python extension to MNIST



- Python extension using learning rules that match SPICE simulations
- 28x28 input pixels binarized
- Limited fan-in to 50

- Nominal Network: 3 hidden layers with 200 units each
- ~ 120 ms to train 1,000 times through the full 60,000 examples
- 90.6% test accuracy (on 10,000 new examples)

- Modern AI is an example of how bio-inspired computing can outperform traditional computing paradigms
- Neuromorphic computing takes bio-inspired computing to the device level
- Superconductive technology has many advantageous in neuromorphic applications
 - Particularly spiking
 - Particularly information transfer
- Neuromorphic is at its early stages but has potential from large scale systems down to sensor integrated processing

Reinforcement learning collaborators

Emilie Jué¹, Matt Pufall¹, Ken Segall², Chuck Anderson³

1. National Institute of Standards and Technology, Boulder, CO

2. Colgate University, Hamilton, NY

3. Colorado State University, Fort Collins, CO

