SAND2022-3825C

COINFLIPS

COINFLIPS: CO-designed Improved Neural Foundations Leveraging Inherent Physics Stochasticity

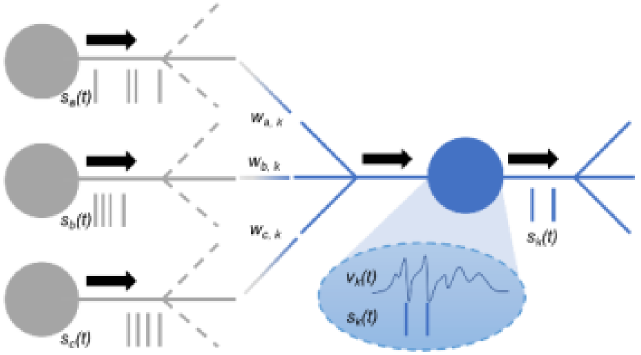
Brad Aimone

jbaimon@sandia.gov

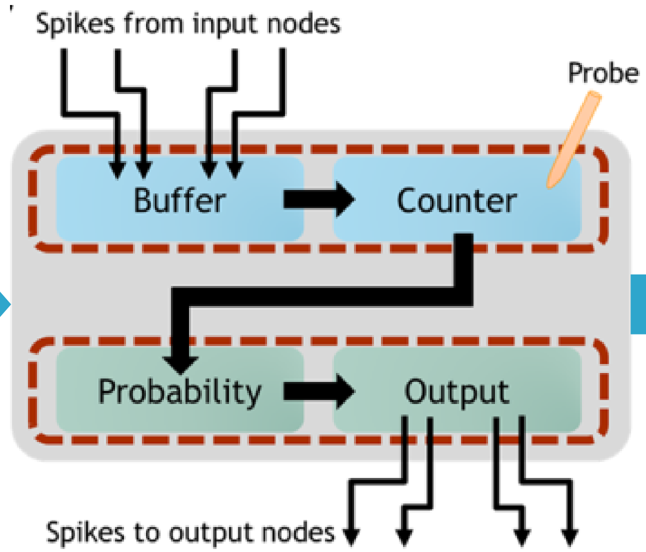
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

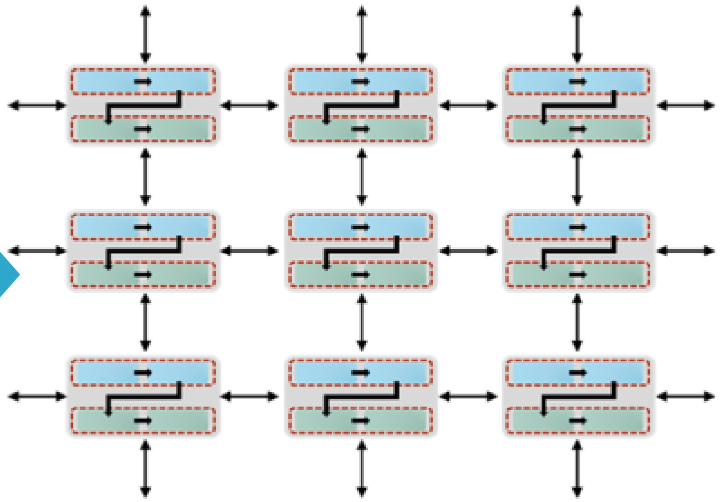
Neuromorphic hardware is advantageous on probabilistic algorithms

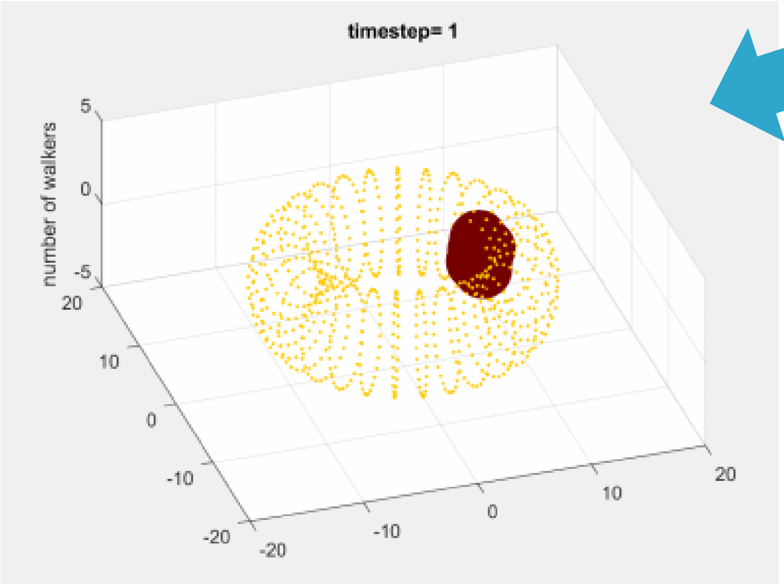
Neuromorphic algorithm can simulate random walks



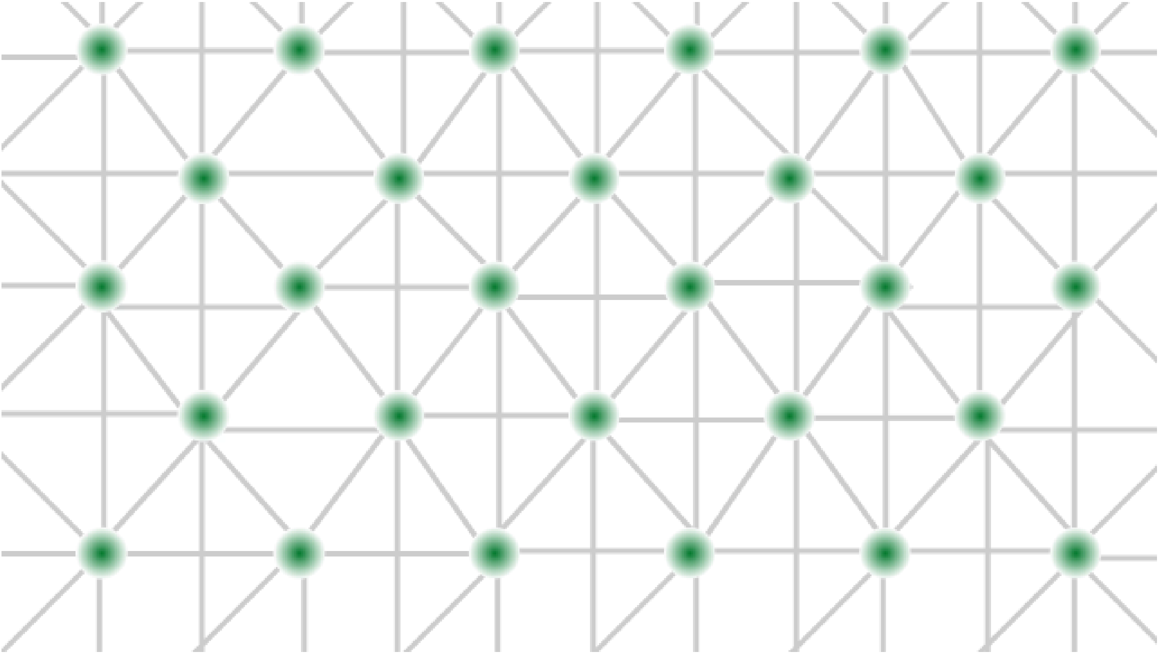
Leaky Integrate and Fire Neuron





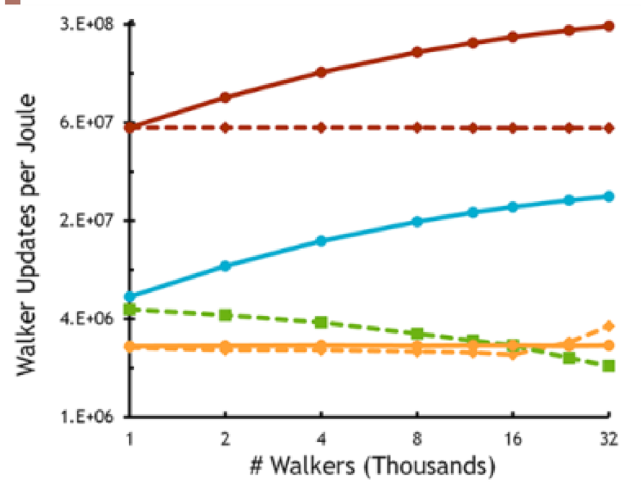


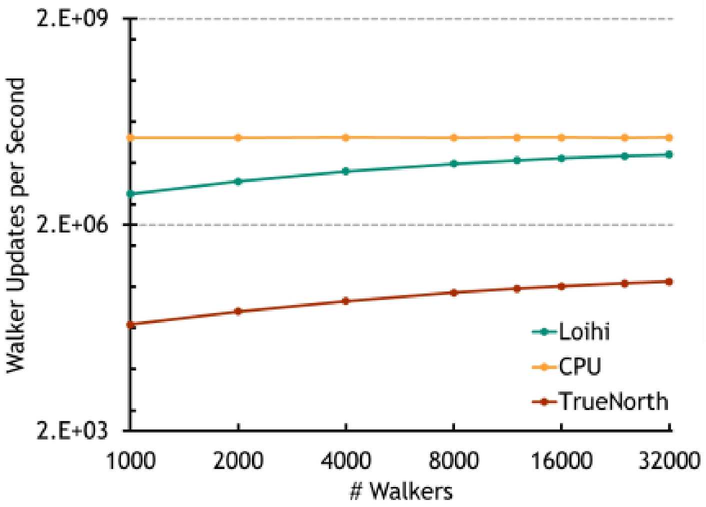
Smith et al., Nature Electronics 2022

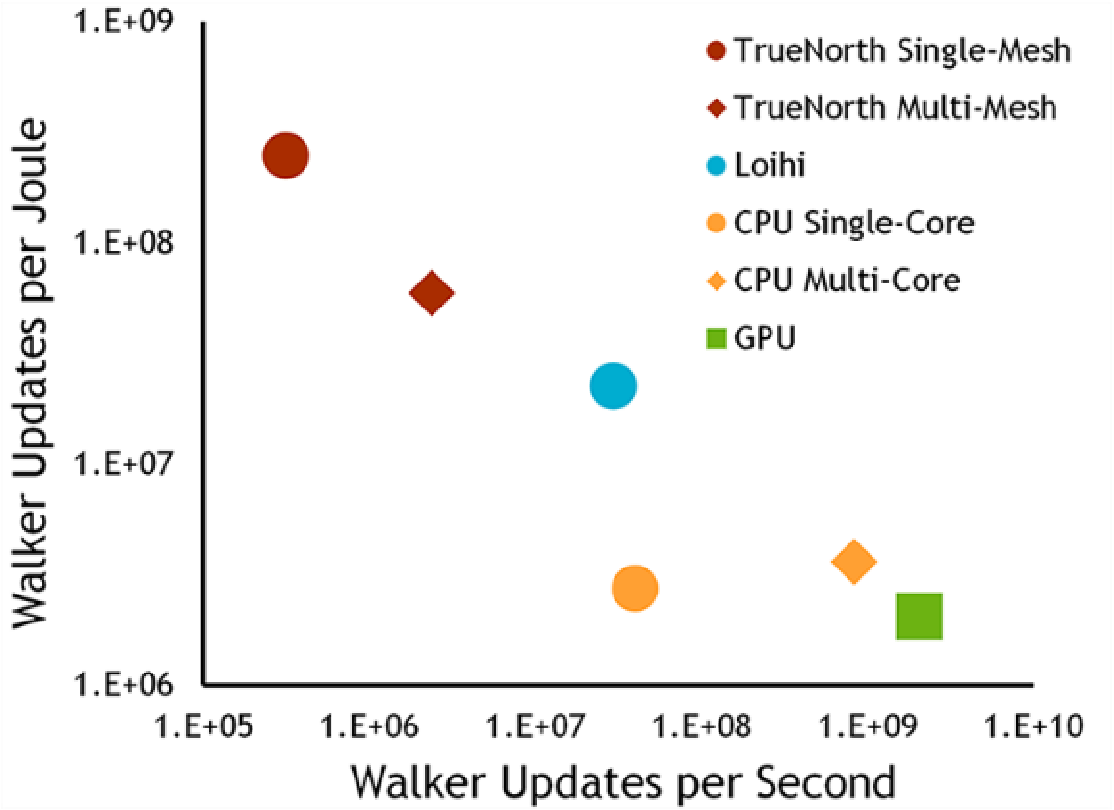


We can identify a neuromorphic advantage for simulating random walks

We define a *neuromorphic advantage* as an algorithm that shows a demonstrable **advantage** in terms of one resource (e.g., energy) while exhibiting comparable **scaling** in other resources (e.g., time).







Smith et al., Nature Electronics 2022

Where does this advantage come from?

- Extreme parallelism of neuromorphic hardware
  plus
  Embarrassingly parallel nature of Monte Carlo random walks

- Many simple calculations in parallel
  vs
  Single complex calculation

- Limiting factor going forward will likely be probabilistic component
  - Quality and form of random numbers
  - Quantity and location of random number generation

What happens if we build a neuromorphic chip centered on probabilistic sampling?

What constitutes brain inspiration?

COINFLIPS

Analog computing!

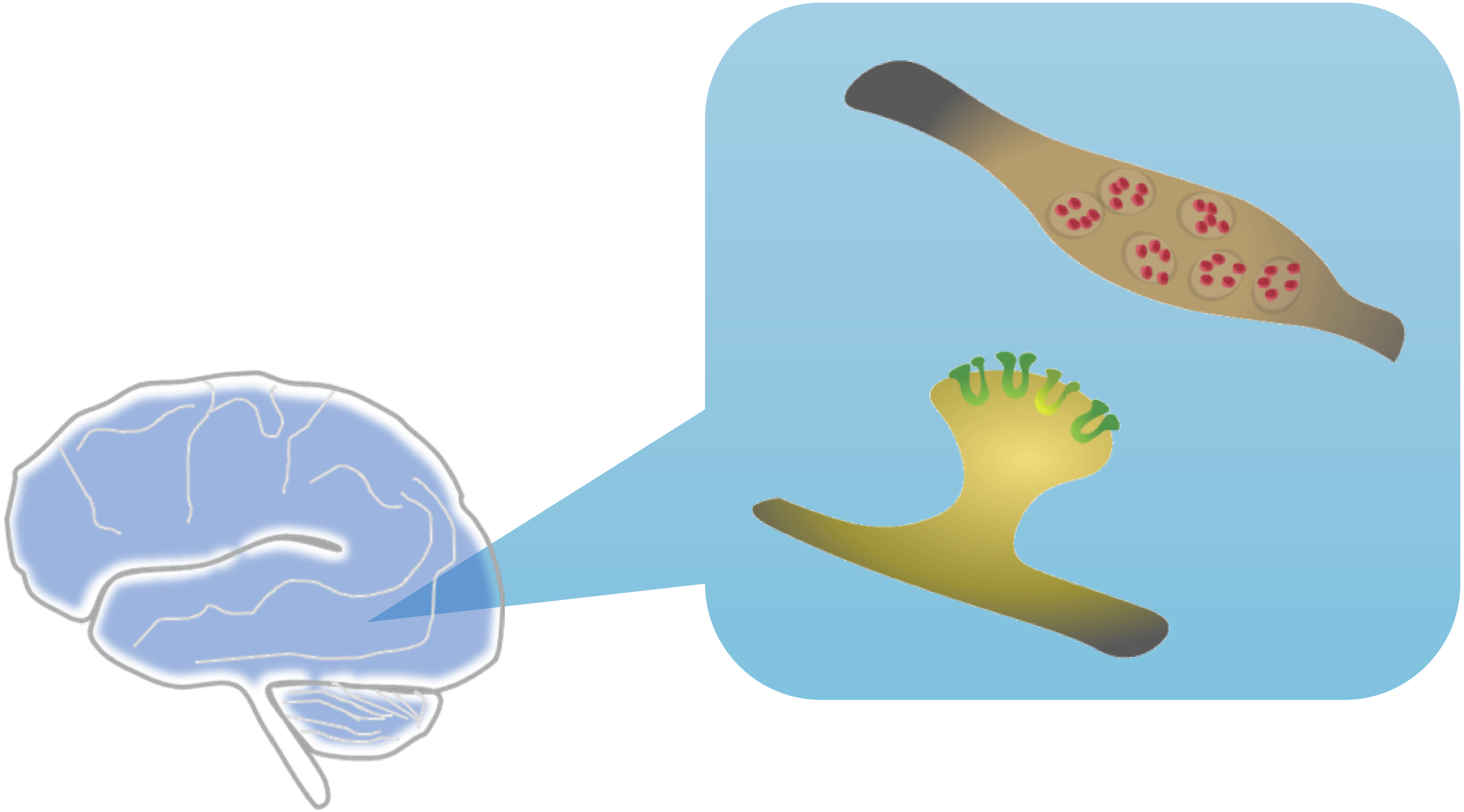
High fan-in connectivity!

Spiking!

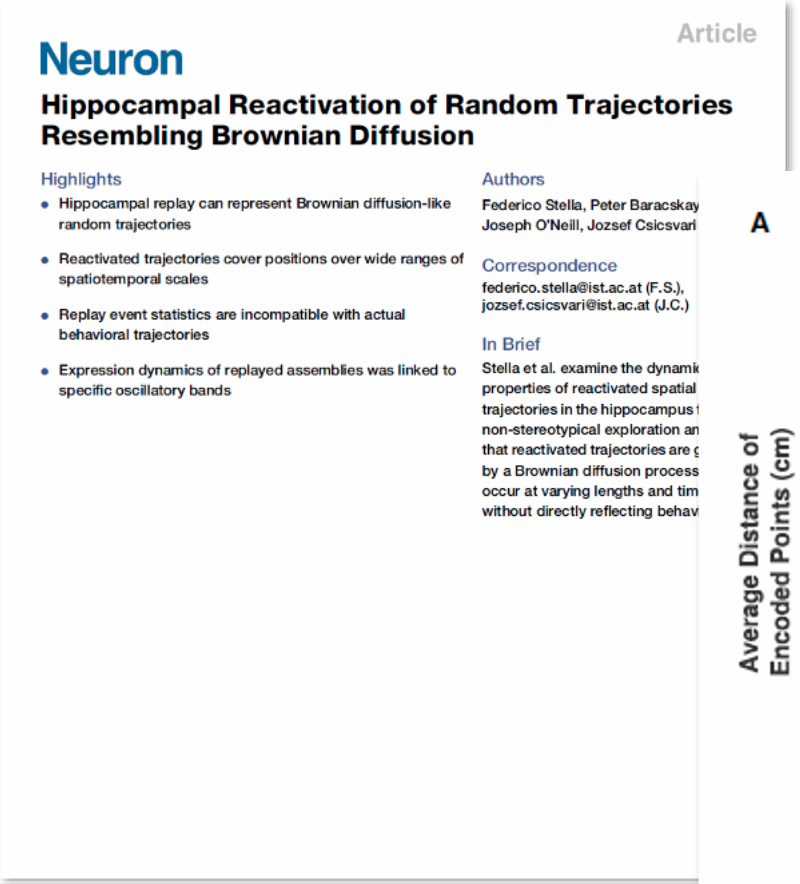
Learning!

Stochasticity!

The brain's trillions of synapses exhibit considerable stochasticity



The brain appears to use probabilistic sampling of populations



Neuron

Article

Hippocampal Reactivation of Random Trajectories Resembling Brownian Diffusion

Highlights

- Hippocampal replay can represent Brownian diffusion-like random trajectories
- Reactivated trajectories cover positions over wide ranges of spatiotemporal scales
- Replay event statistics are incompatible with actual behavioral trajectories
- Expression dynamics of replayed assemblies was linked to specific oscillatory bands

Authors

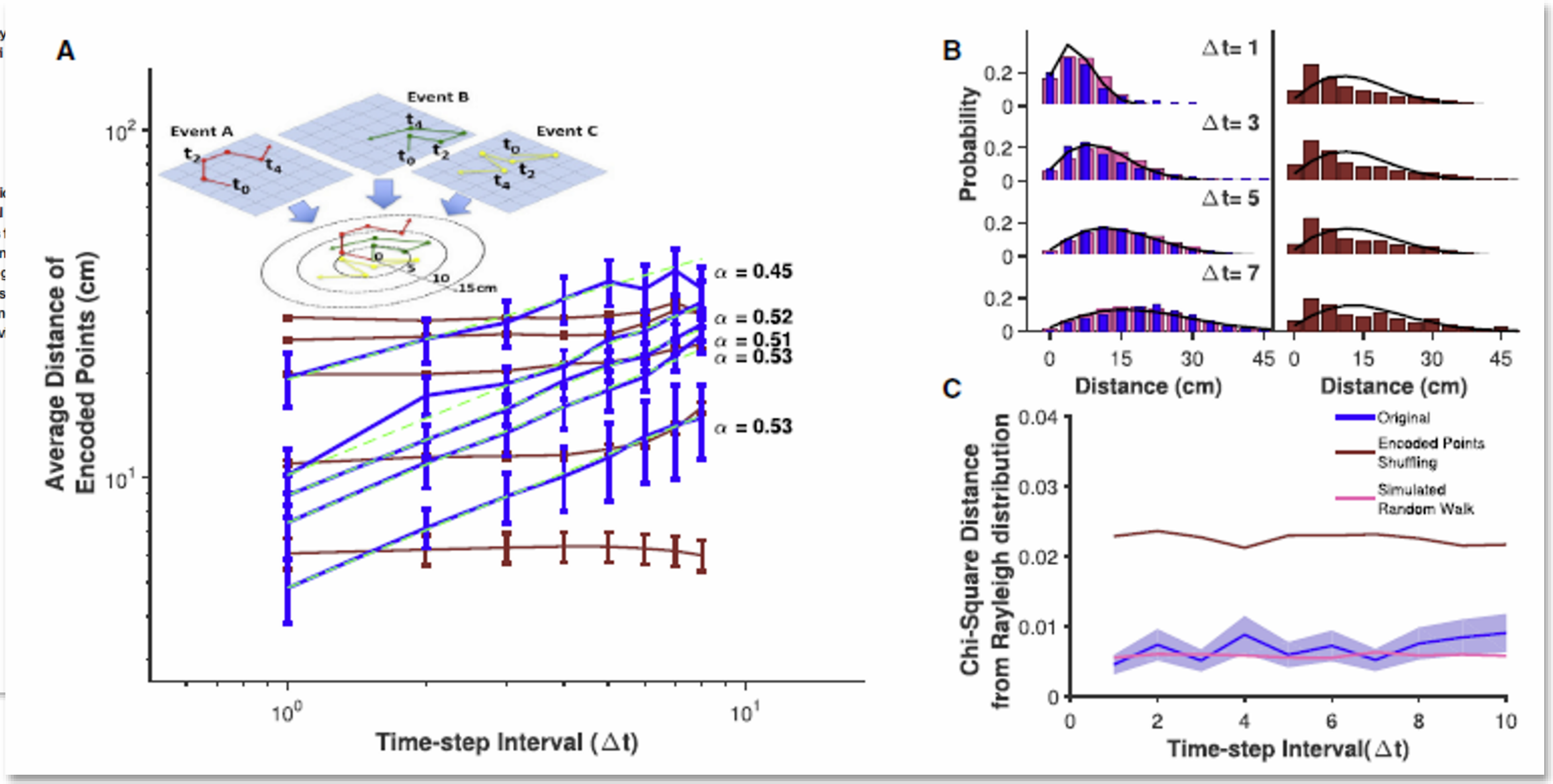
Federico Stella, Peter Baracskay
Joseph O'Neill, Jozsef Csicsvari

Correspondence

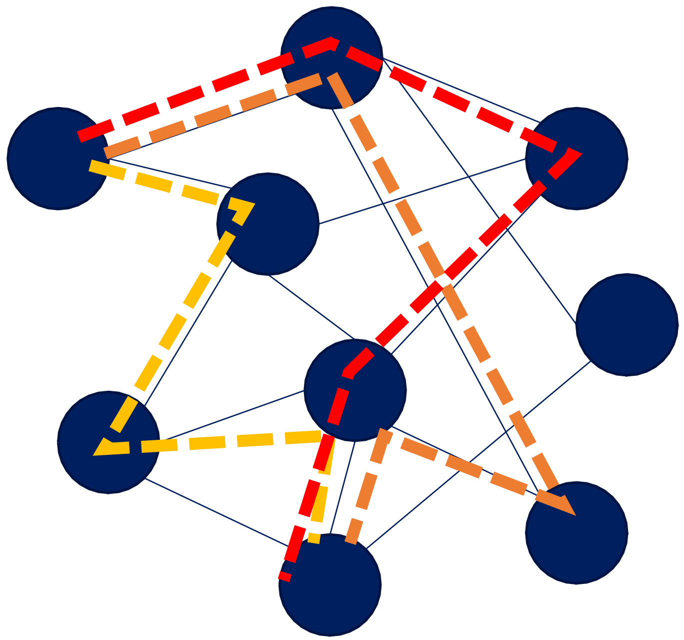
federico.stella@ist.ac.at (F.S.),
jozsef.csicsvari@ist.ac.at (J.C.)

In Brief

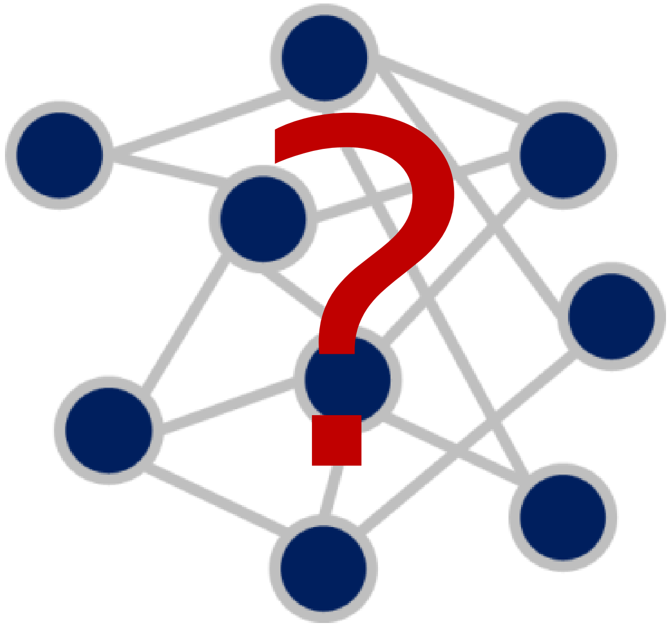
Stella et al. examine the dynami
properties of reactivated spatial
trajectories in the hippocampus
non-stereotypical exploration an
that reactivated trajectories are
by a Brownian diffusion process
occur at varying lengths and tim
without directly reflecting behav

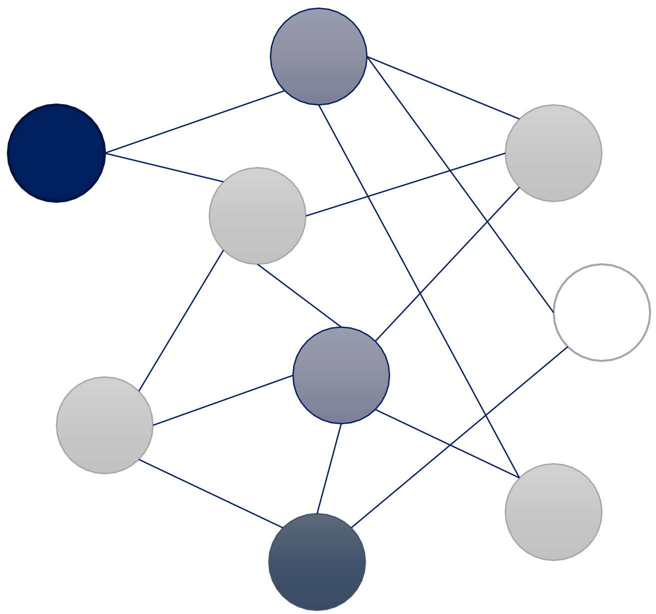


How does brain use this ubiquitous stochasticity?



DTMC random walks
(sampling network)

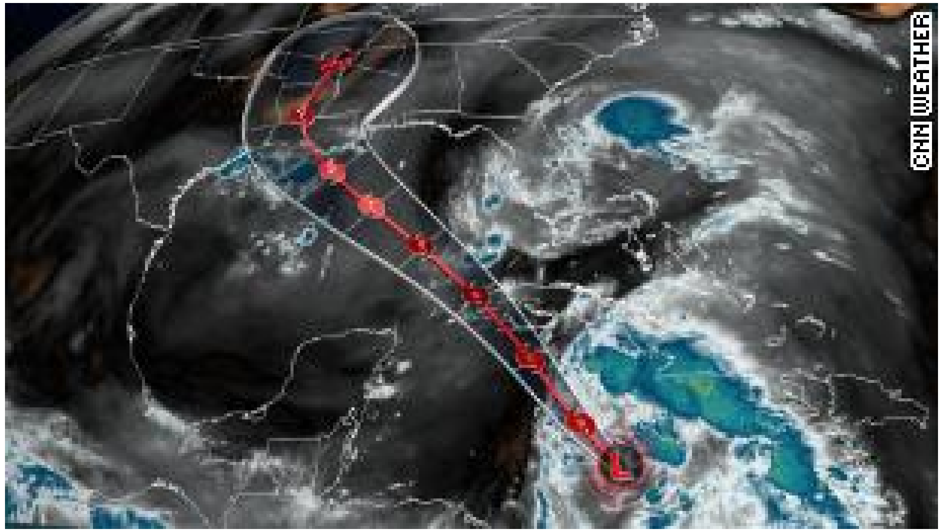


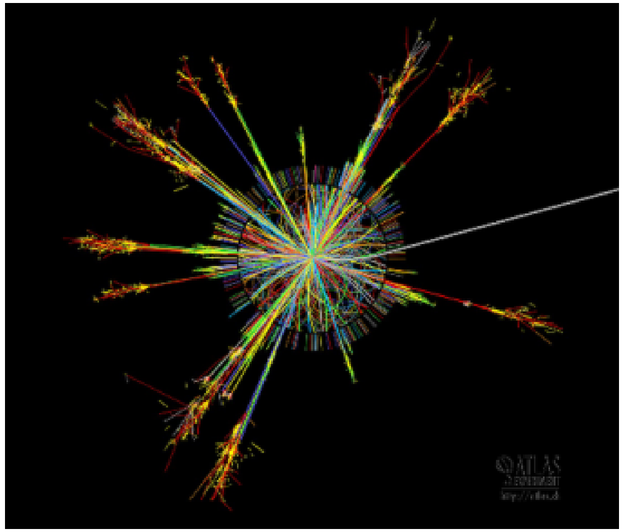


Expected value
(average over stochasticity)

Many applications of computing have inherent uncertainty

Many applications of computing have inherent uncertainty





Two main use cases:

- ❖ Mod-Sim --- Generating the random number *you need*
- ❖ Artificial Intelligence --- Effective and efficient sampling of algorithms

So what would a probabilistic neuromorphic computer look like?

COINFLIPS

Goal: *1 billion RNs per microsecond*

- ~1e11 neurons x 1e4 synapses / neuron x 1 Hz = 1e15 RNs per second in human

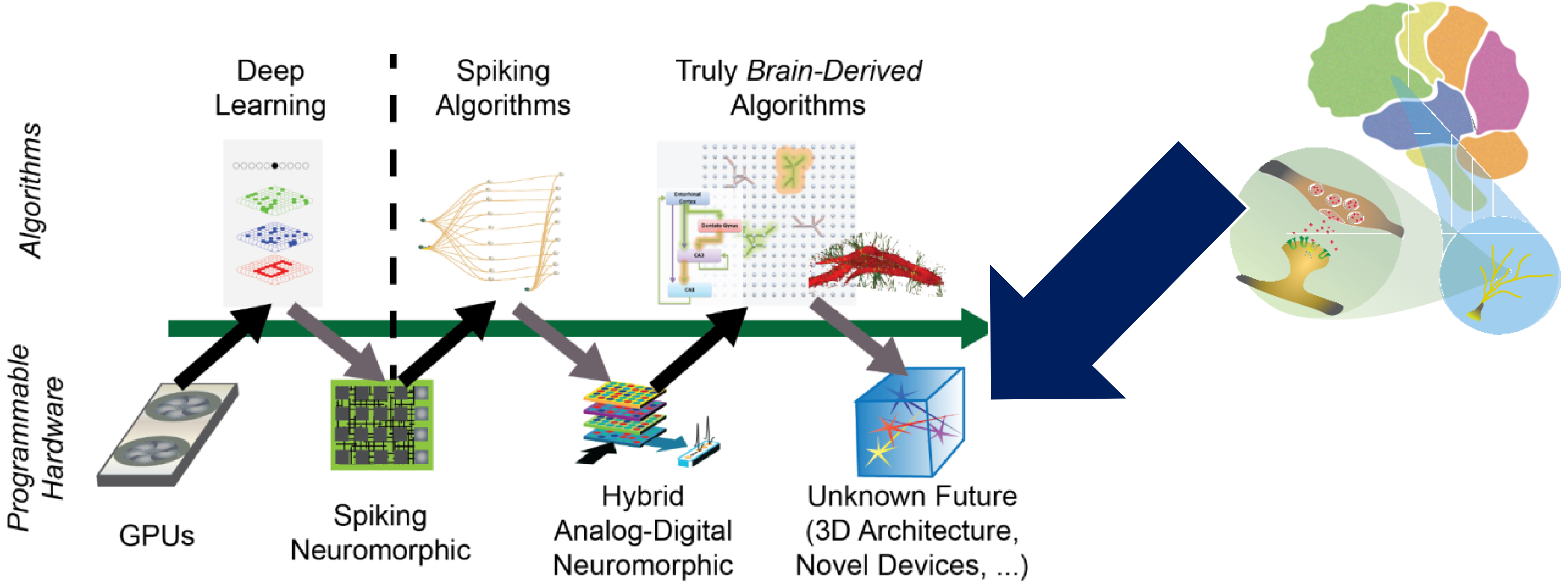
Why?

- Numerical computing
- Artificial Intelligence

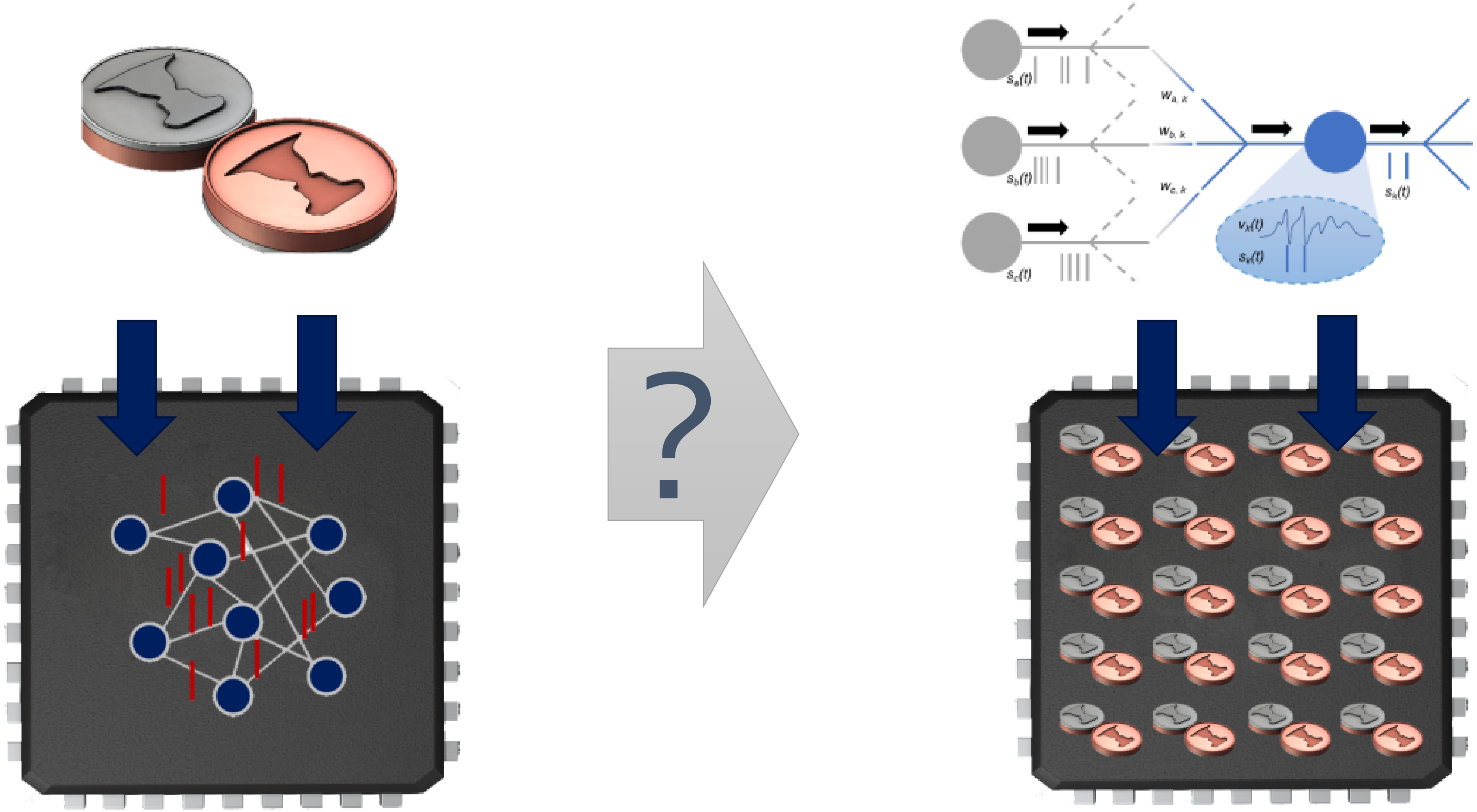
How?

- Stochastic devices
- Neuromorphic architecture

One possibility is to inject ubiquitous stochasticity into existing neuromorphic technologies



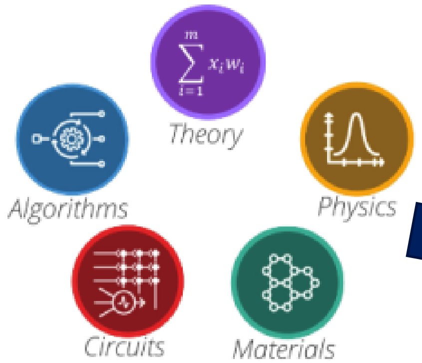
Making stochasticity ubiquitous may require that we revisit how we design neuromorphic hardware



COINFLIPS

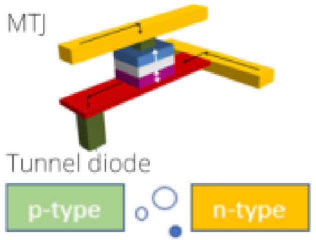
CO-designed Improved Neural Foundations Leveraging Inherent Physics Stochasticity (COINFLIPS)

COINFLIPS



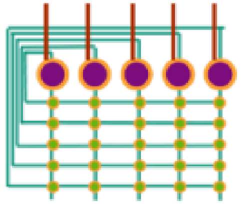
Co-design approach on new energy-efficient microelectronics

Approach



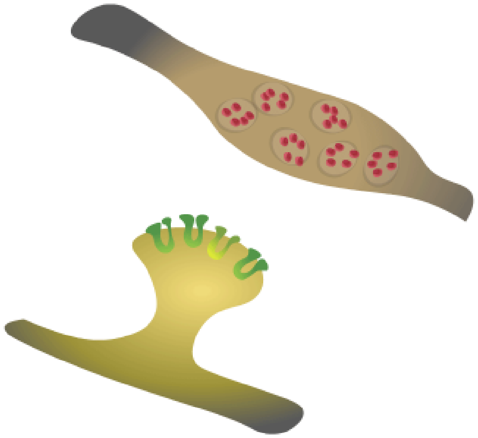
Tunable Stochastic Devices

NYU



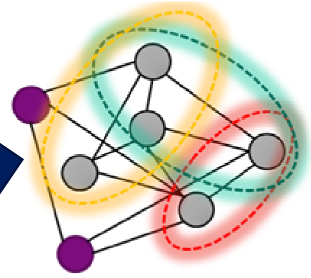
Probabilistic Circuits and Architectures

THE UNIVERSITY OF TEXAS AT AUSTIN



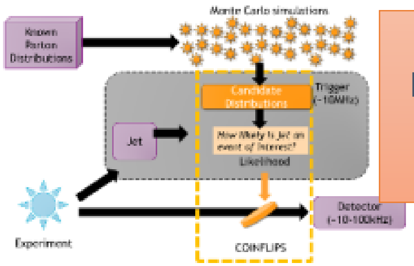
Every synapse in the brain is a stochastic "coinflip"

Inspiration



Probabilistic Neural Theory and Algorithms

OAK RIDGE National Laboratory



Particle Physics Demonstration

Show Empirical and Theoretical Application Impact

Formal Theoretical Framework

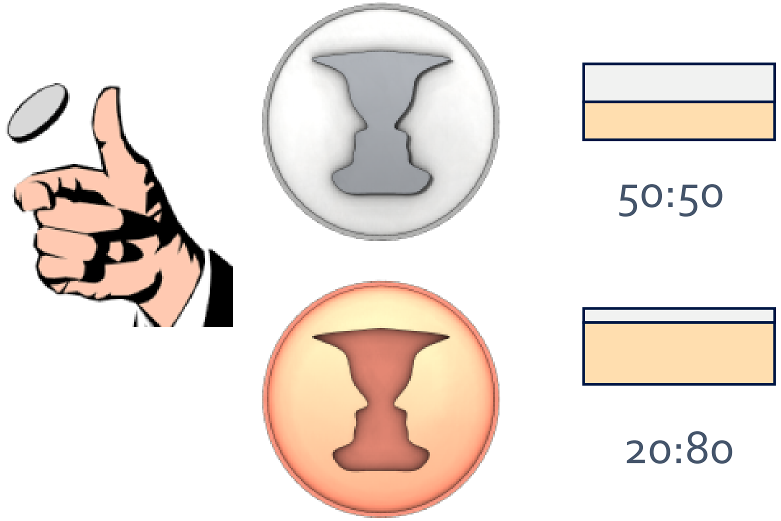
Prototype Probabilistic Neuromorphic System

Scalable and Tunable Stochastic Device Strategy

Tunable RNG – magnetic tunnel junctions & tunnel diodes

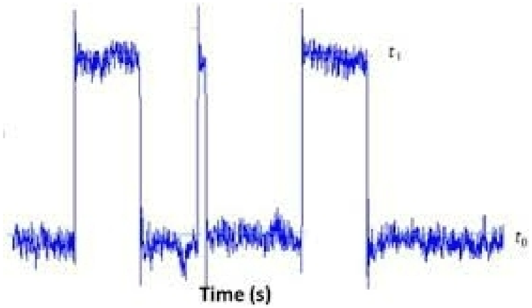
COINFLIPS

Tunable random number generator

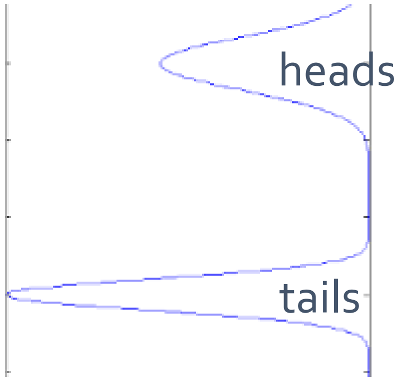


Why did we pick the devices we picked?

Large signals



Tunable

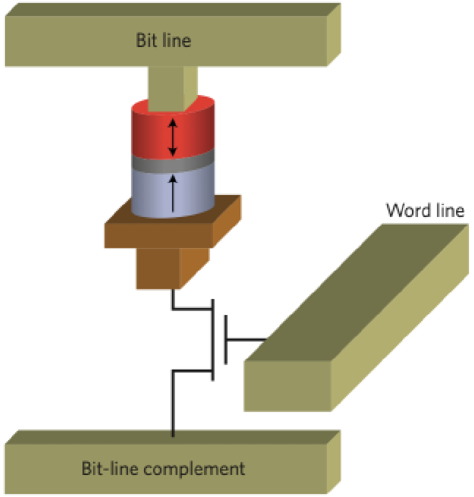


Integration

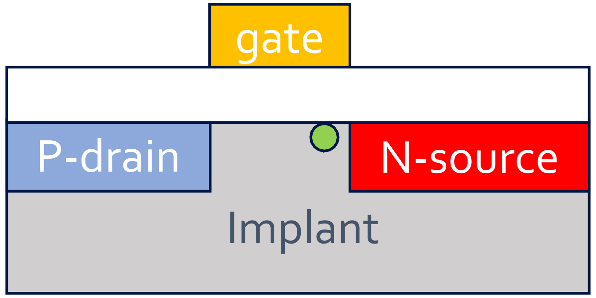
I. Magnetic tunnel junctions

Jean Anne Incorvia

Andy Kent

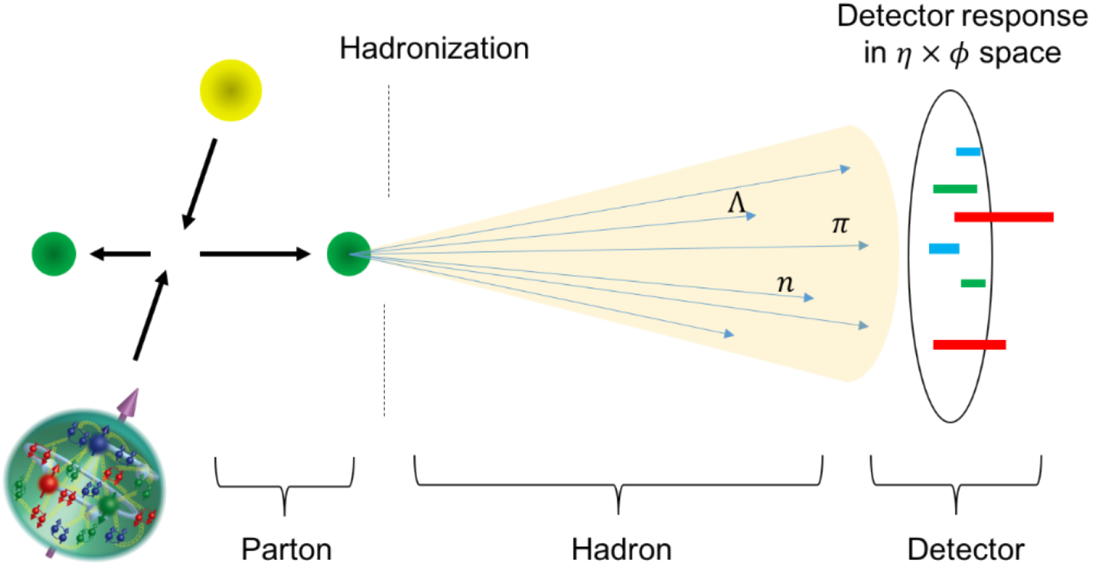


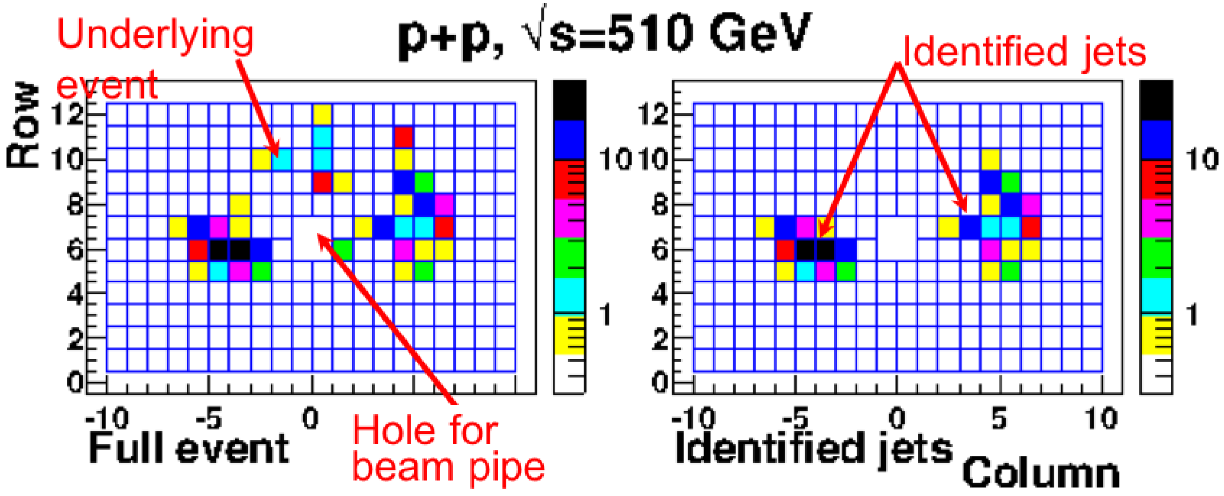
II. Tunnel diodes



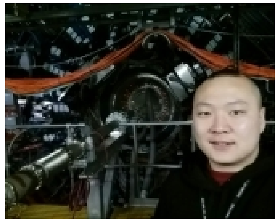
Shashank Misra & Tzu-Ming Lu

Jet detection in particle physics



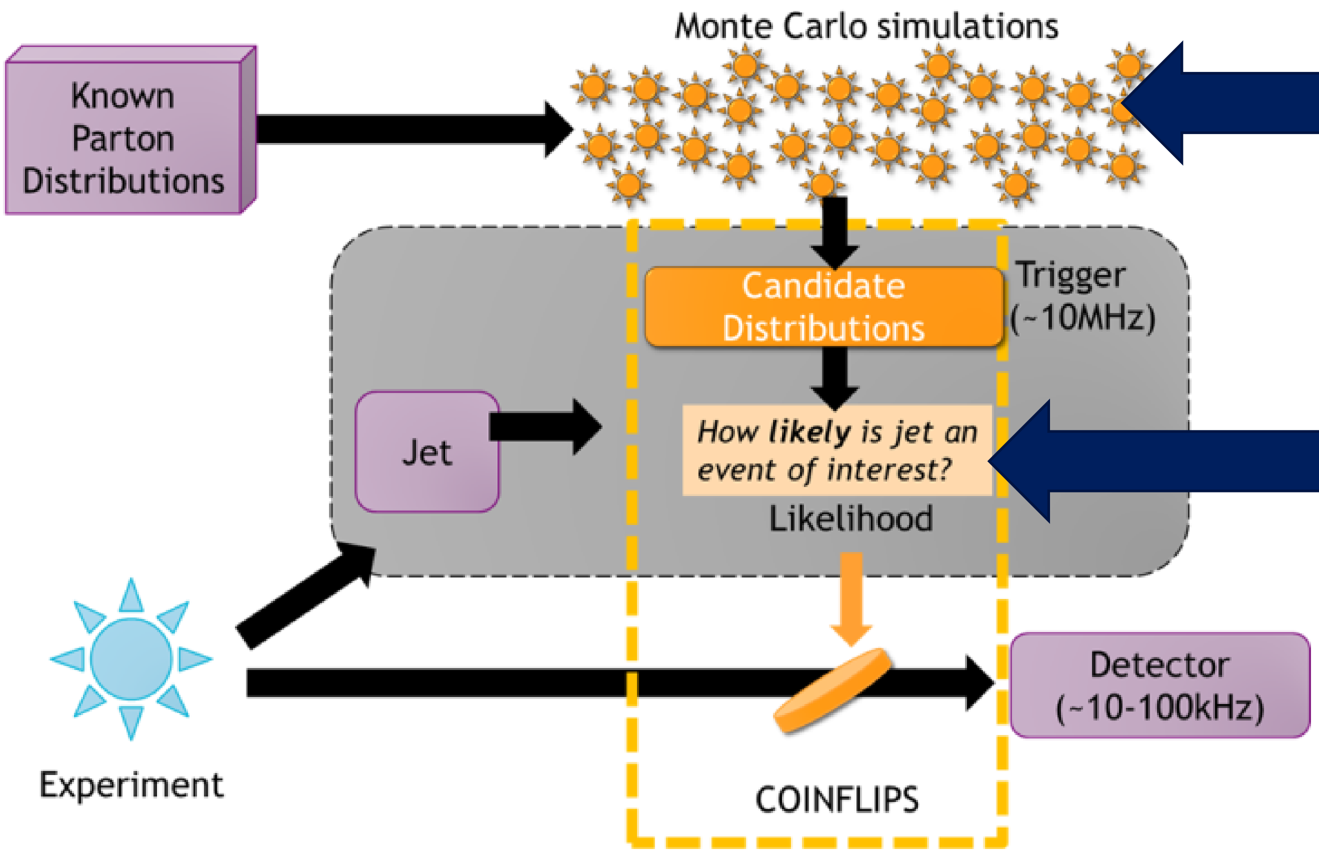






Les Bland, Bernd Surrow, Jae Nam

Opportunities for probabilistic neuromorphic computing in physics jet identification



Improved model generation of artificial data to train ML model

Real-time (<microsecond) Bayesian neural network identification of events

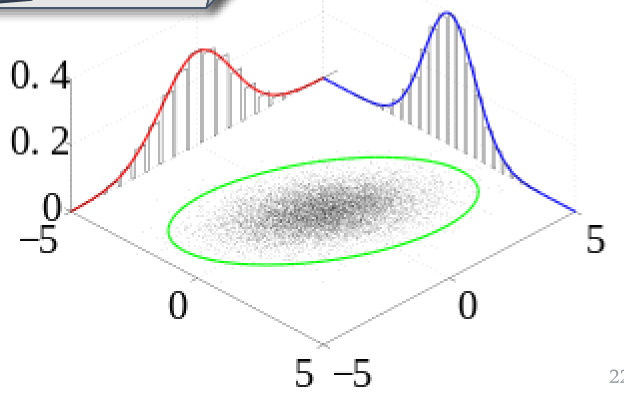
How do we use coinflips to sample from arbitrary distributions?

Biased random source to approximate uniform random numbers

Some literature here

Uniform random numbers to arbitrary distributions

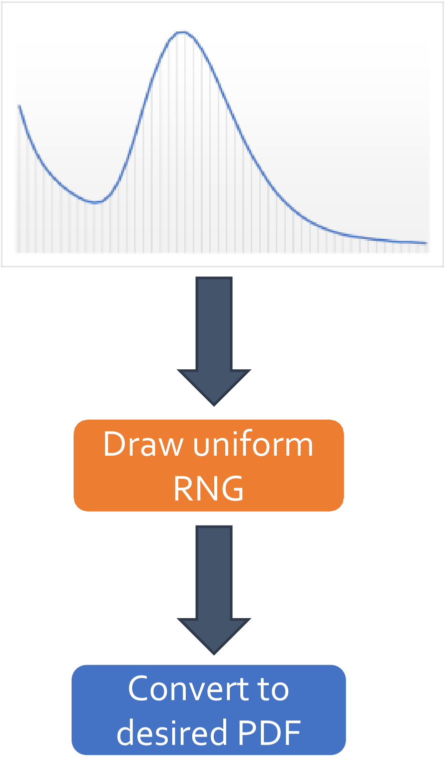
A major focus of numerical methods community



Biased random source to sample an arbitrary probability distribution

Relatively unexplored

Random numbers are a non-trivial computational cost today

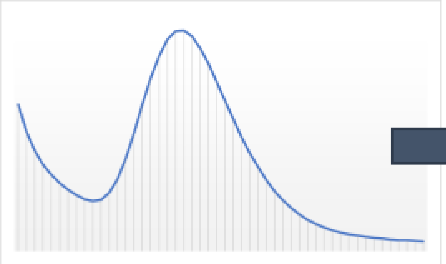


We want a RN pulled from some physics distribution

Software uses pseudo-RNG to pull uniform random number
- This is simple, but can be costly for volume and quality

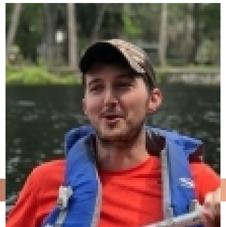
Numerical methods convert uniform RN to desired distribution
- Some distributions are easy (simple inverse CDF)
- Some distributions are challenging

It is possible to generate a random number from a desired statistical distribution



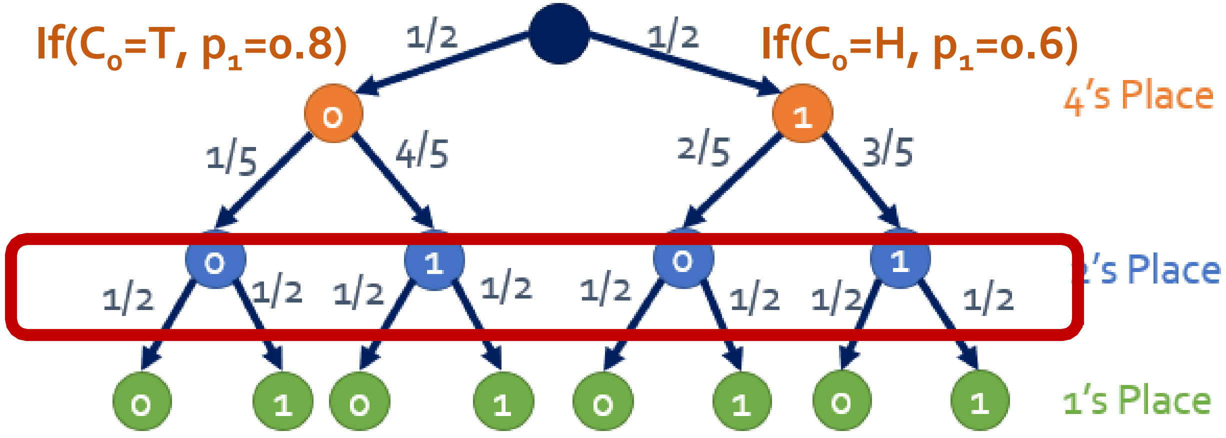
Draw uniform RNG

Convert to desired PDF



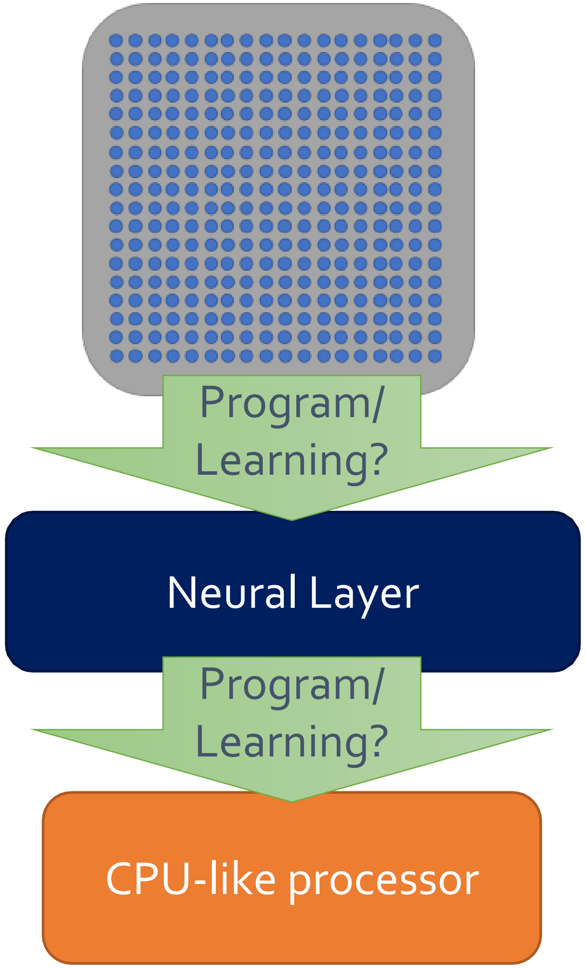
Darby Smith

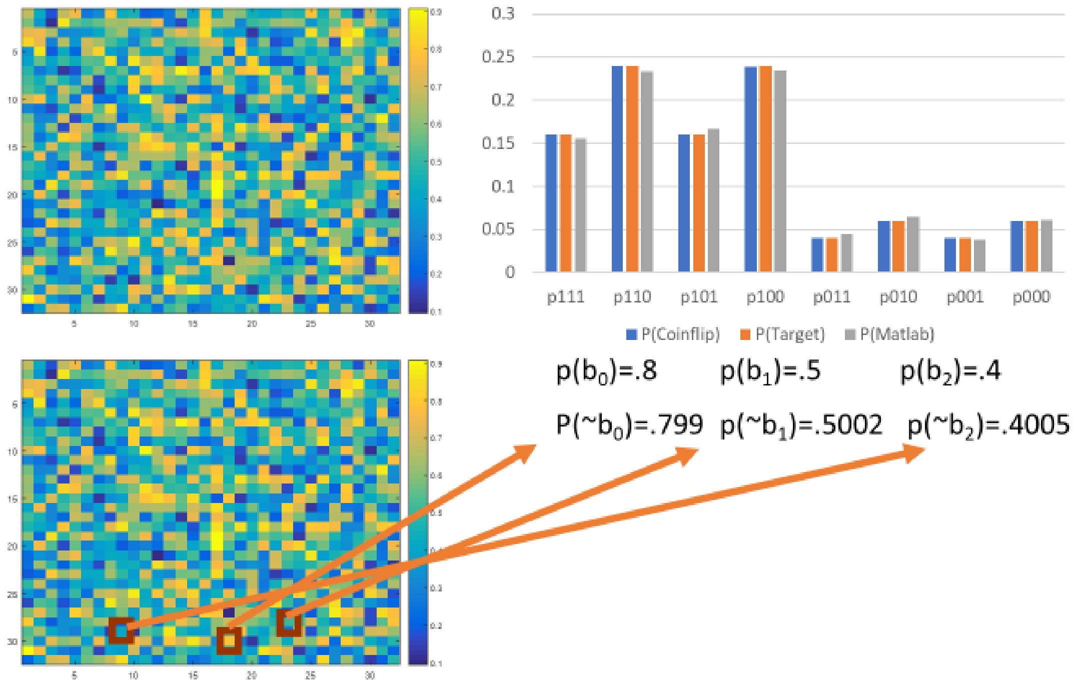
Expand Boolean tree of PDF and flip many coins for all branches in parallel



- Worst case, this is a exponentially large number of coins
- PDFs have structure and redundancies that can be leveraged
- Correlations from devices or built into neural circuits can similarly compress tree

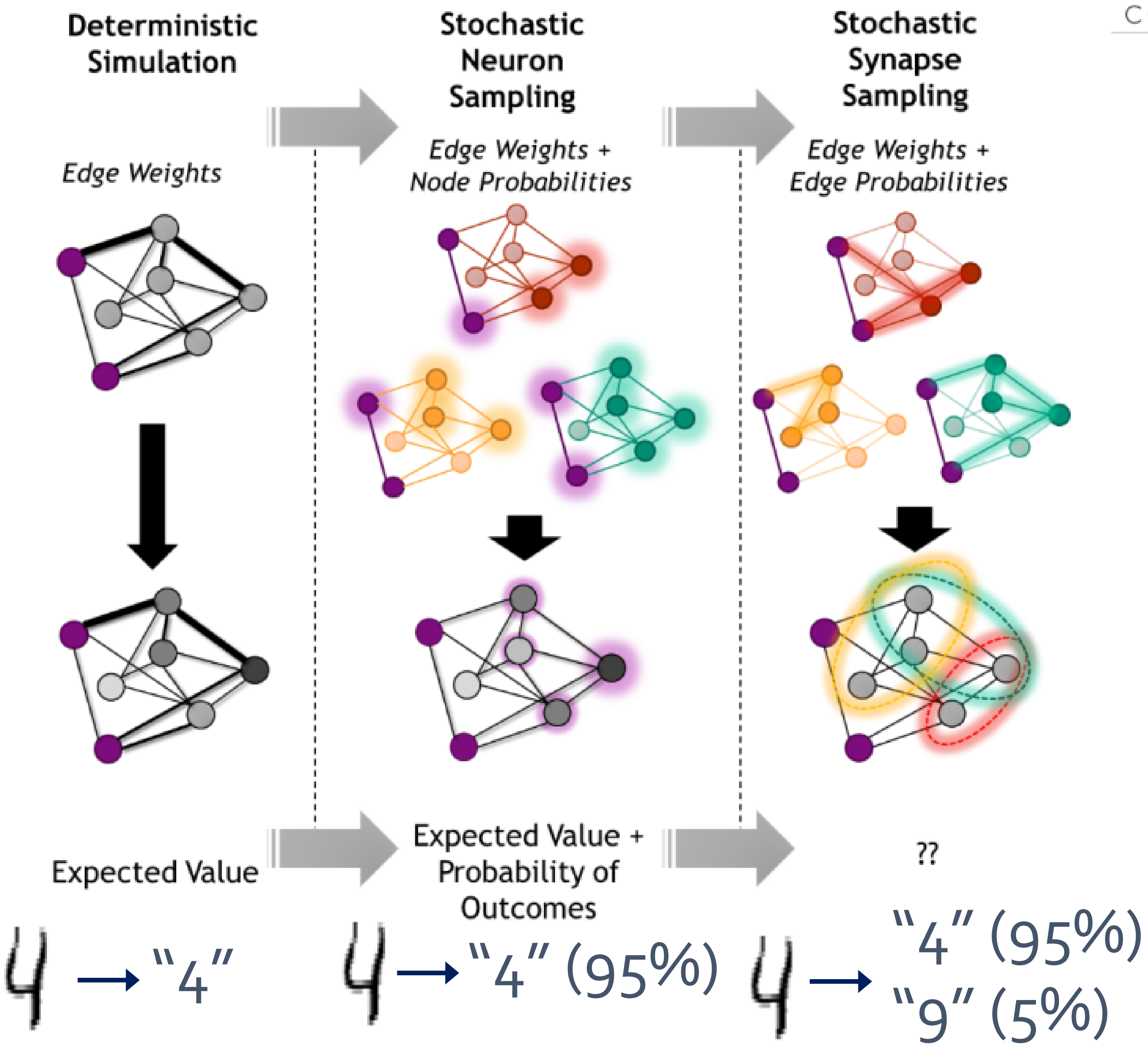
A potential COINFLIPS architecture for generating random numbers





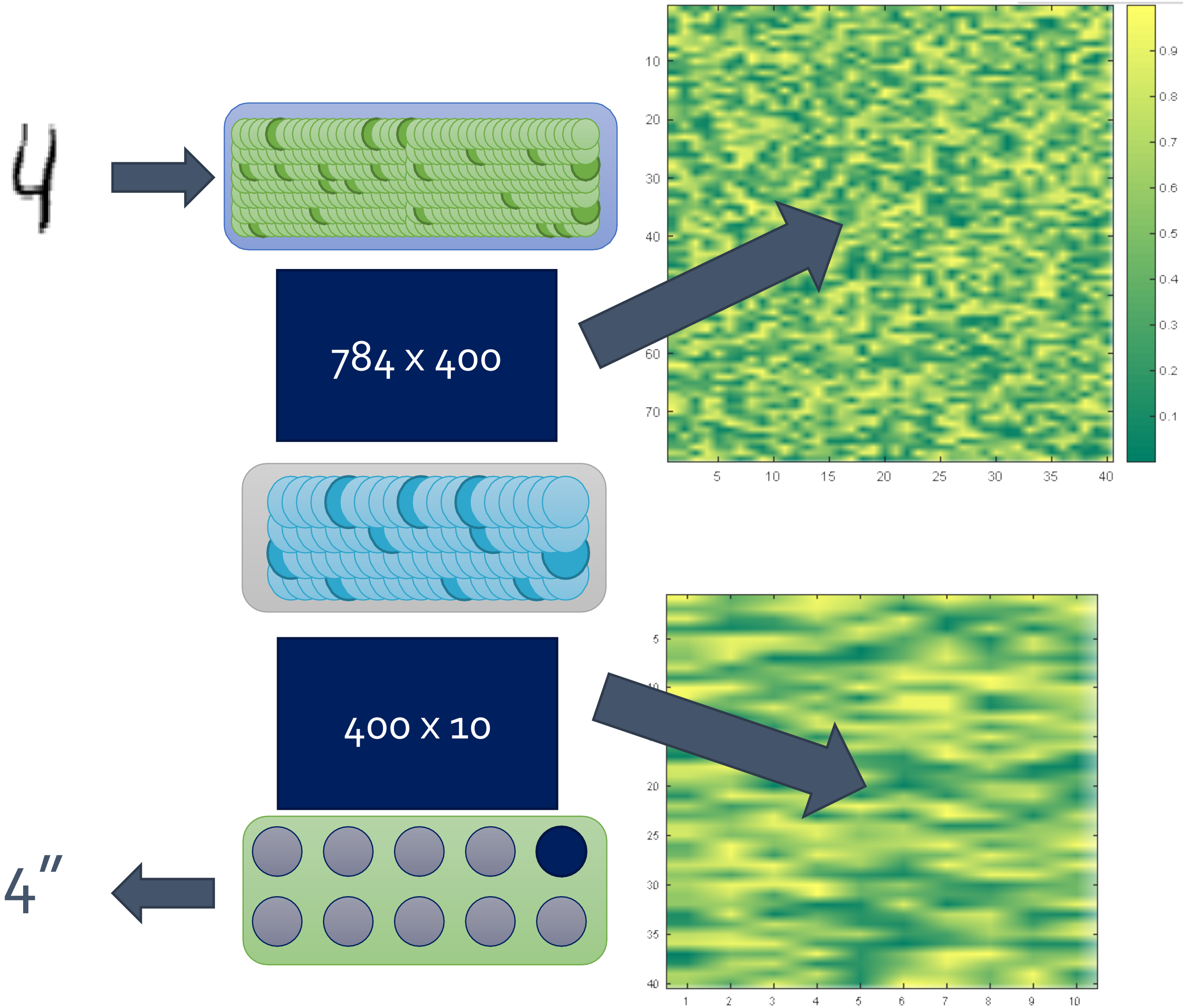
Establish a paradigm of computation around synaptic sampling

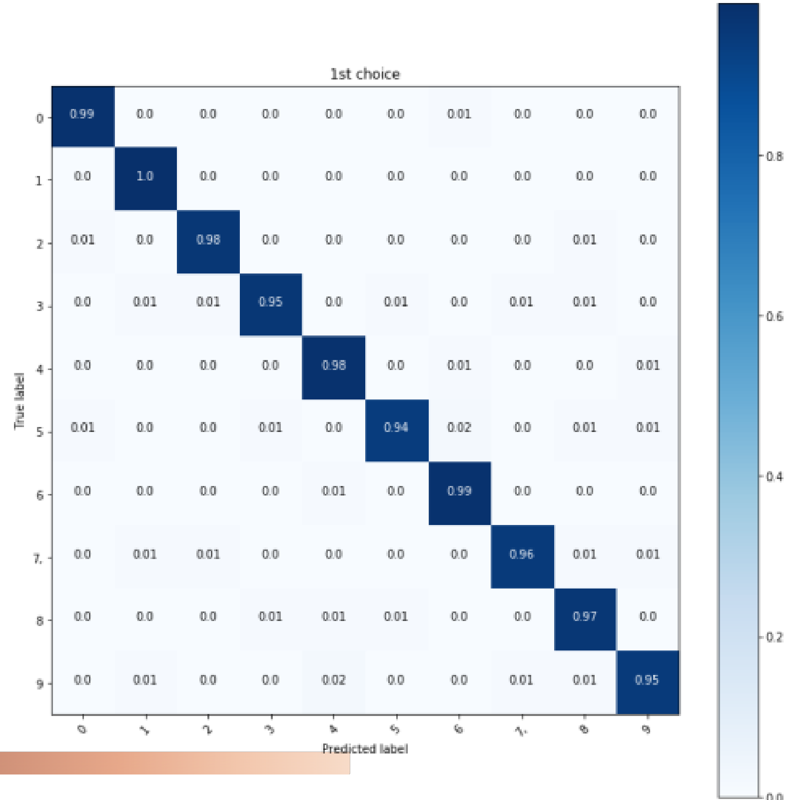
Can novel neural sampling algorithms be leveraged to provide more efficient and more powerful AI capabilities?



Sampling ANNs with stochastic synapses provides estimate of uncertainty

- Approach
  - Train simple neural network with only minor modificati
  - Simple network can achieve decent performance

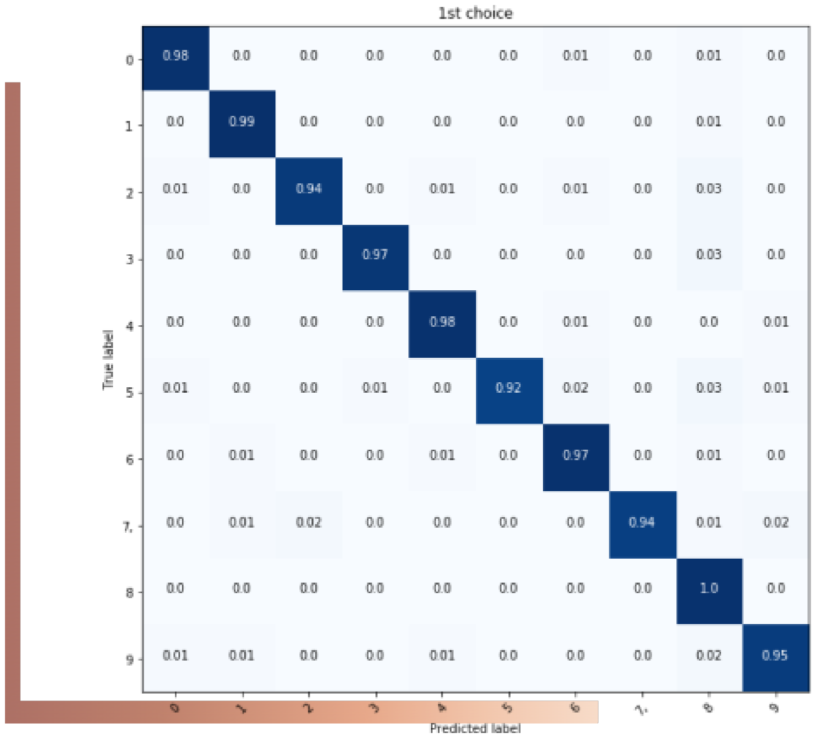




Weights continuous between 0 and 1

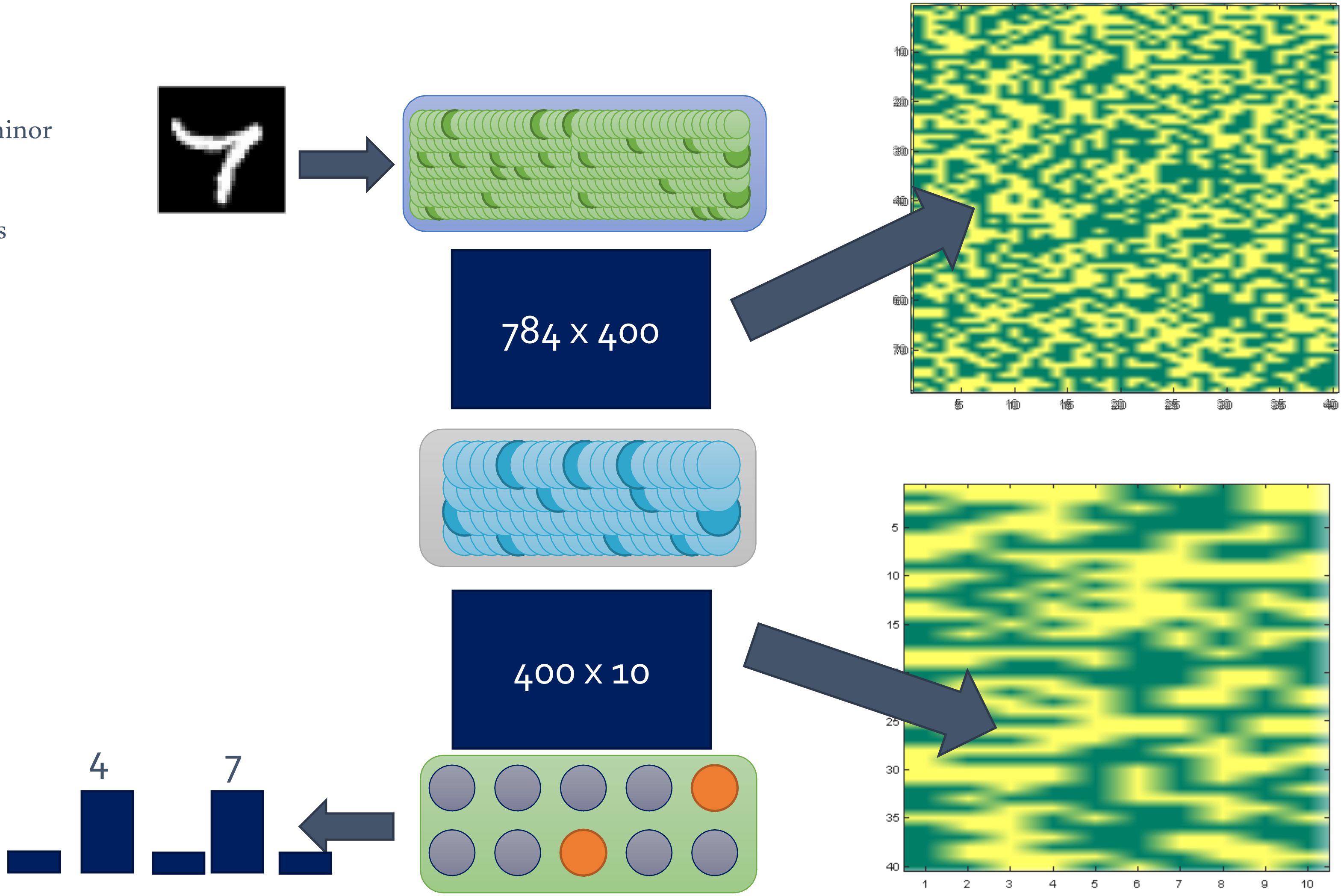
Sampling ANNs with stochastic synapses provides estimate of uncertainty

- Approach
  - Train simple neural network with only minor modifications
  - Convert weights to Bernoulli probabilities (weighted coinflips)
  - Sample network to identify what classes



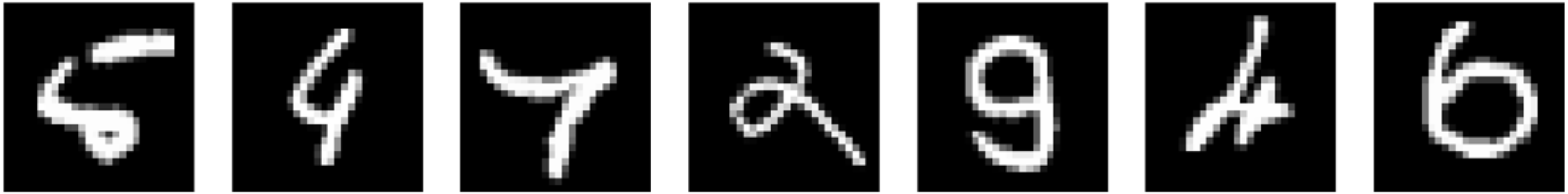
1st choice

True label \ Predicted label	0	1	2	3	4	5	6	7	8	9
0	0.98	0.0	0.0	0.0	0.0	0.0	0.01	0.0	0.01	0.0
1	0.0	0.99	0.0	0.0	0.0	0.0	0.0	0.0	0.01	0.0
2	0.01	0.0	0.94	0.0	0.01	0.0	0.01	0.0	0.03	0.0
3	0.0	0.0	0.0	0.97	0.0	0.0	0.0	0.0	0.03	0.0
4	0.0	0.0	0.0	0.0	0.98	0.0	0.01	0.0	0.0	0.01
5	0.01	0.0	0.0	0.01	0.0	0.92	0.02	0.0	0.03	0.01
6	0.0	0.01	0.0	0.0	0.01	0.0	0.97	0.0	0.01	0.0
7	0.0	0.01	0.02	0.0	0.0	0.0	0.0	0.94	0.01	0.02
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
9	0.01	0.01	0.0	0.0	0.01	0.0	0.0	0.0	0.02	0.95



Weights *sampled* as probability

2^{nd} choice of stochastic sampled networks is often the 'right' answer for misclassified results

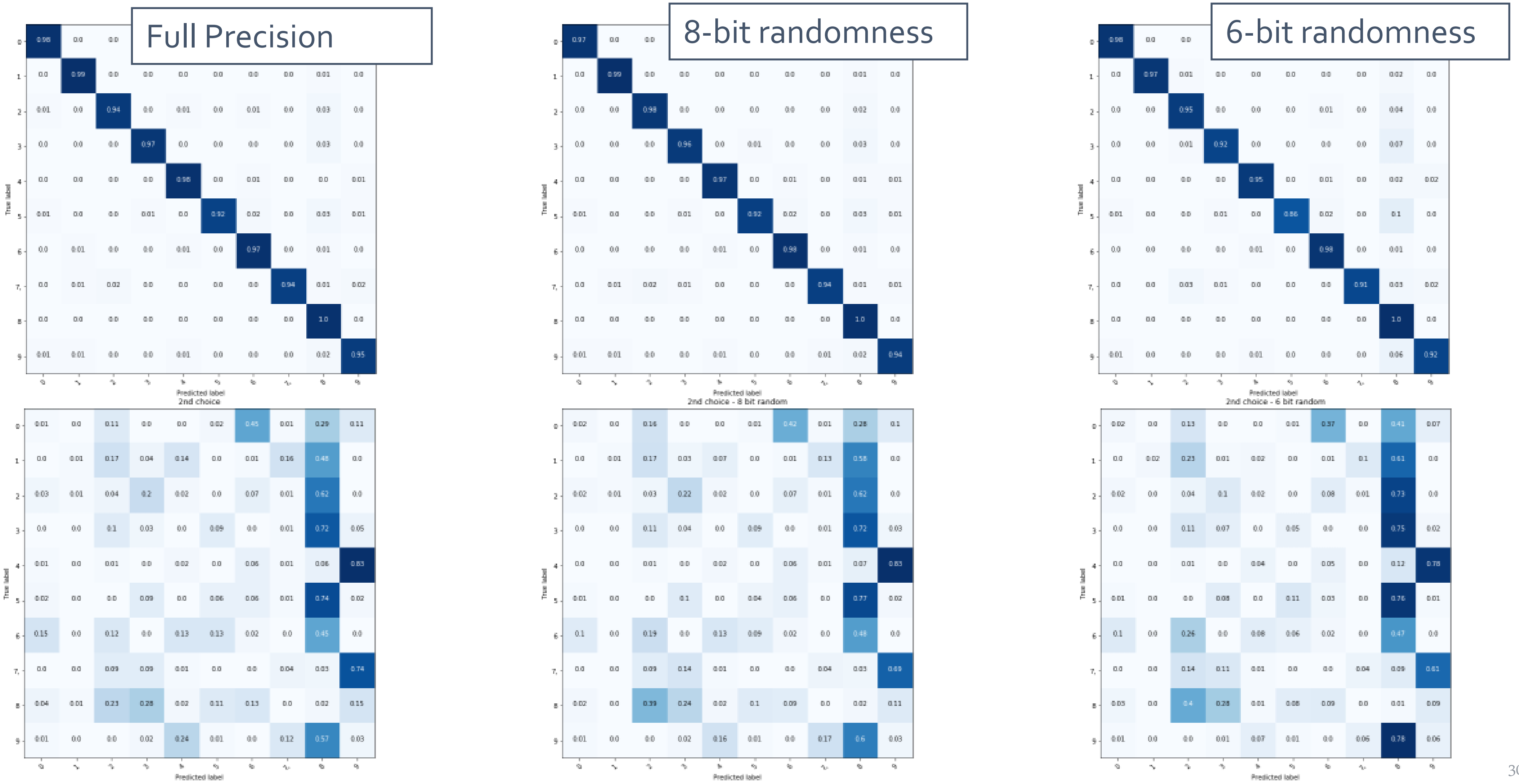


6 – 0.38	9 – 0.31	4 – 0.36	9 – 0.26	3 – 0.23	6 – 0.26	0 – 0.39
5 – 0.17	4 – 0.28	7 – 0.35	2 – 0.20	9 – 0.20	2 – 0.25	6 – 0.27

Sampling ANNs with stochastic synapses is robust to low precision synapses

1st choice

2nd choice



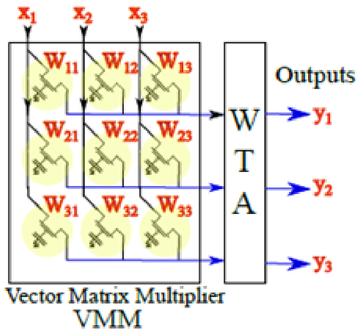
Next step: using AI to guide COINFLIPS neural circuit design

COINFLIPS

Data and Models

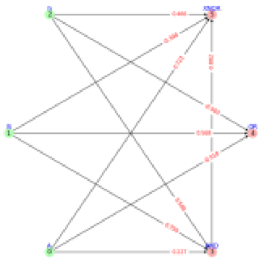
- Data Sweeps
- Device Models
- ASIC behavior models

Topological Analysis

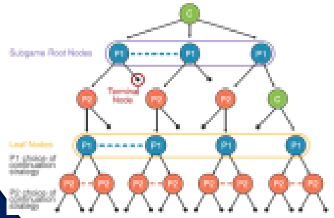


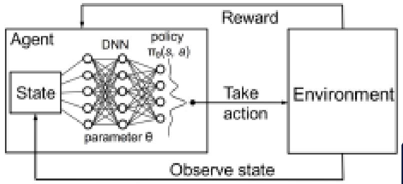
- Size constraints
- Discover novel circuit topologies

Neural Circuits & Architectures



Machine Learning





Hyper Parameters

- Learning Rate
- # of Epochs
- Hardware based constraints in architecture search

Device and Architectural Constraints

- Charge time
- Energy efficiency
- SWaP
- Connectivity
- Extreme Temperature environments

Our AI-enhanced framework would need inputs from algorithms, devices, architectures and ML-based hyper-parameters. The framework will enable new capabilities.

Katie Schuman (Tenn)
Suma Cardwell (Sandia)

COINFLIPS presents an opportunity to develop a *community of interest* to create a new computing paradigm

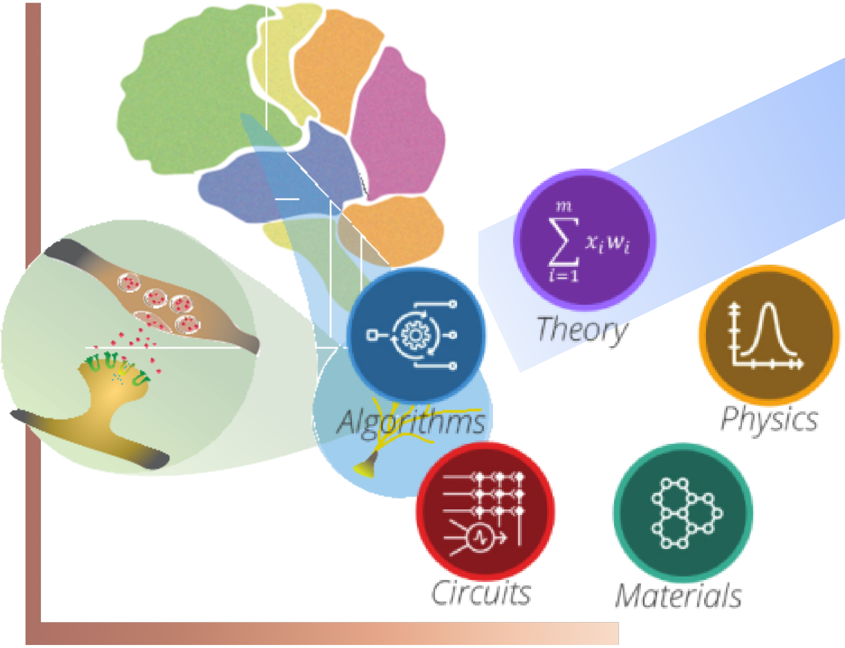
COINFLIPS

Jointly develop a programming model and theoretical framework with an emerging technology

Opportunity for computing to prioritize impact on different classes of applications

Factor in integration and system design from the onset of a new approach

Optimize non-CMOS devices for scalability and cost of reliability



Thanks!