SAND2024-04981C

Sandia National Laboratories

Exceptional service in the national interest

EXPRESSIVE DENDRITES IN SPIKING NETWORKS

Mark Plagge, Suma Cardwell, Frances S. Chance

ICRC 2023

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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.

DENDRITIC HARDWARE & COMPUTATION

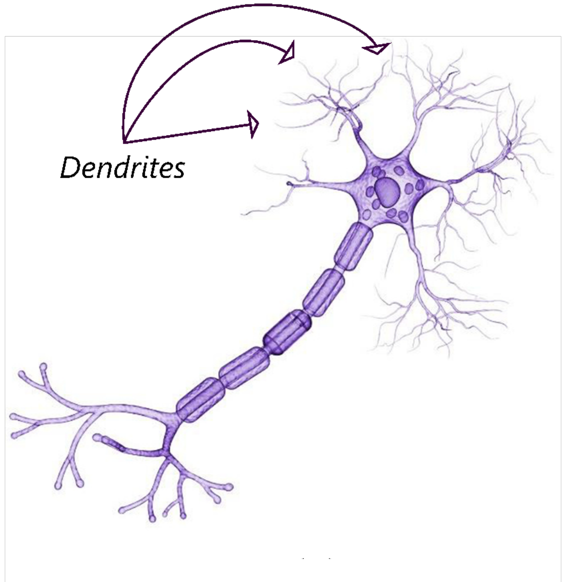
Dendrites are a nonlinear computational components

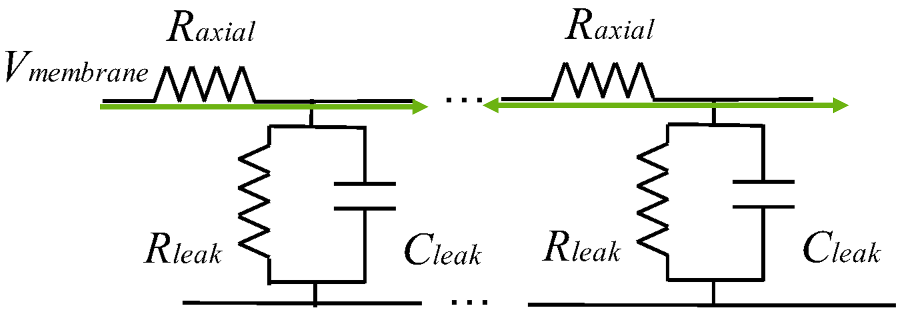
Provide a “pre-processing” computation

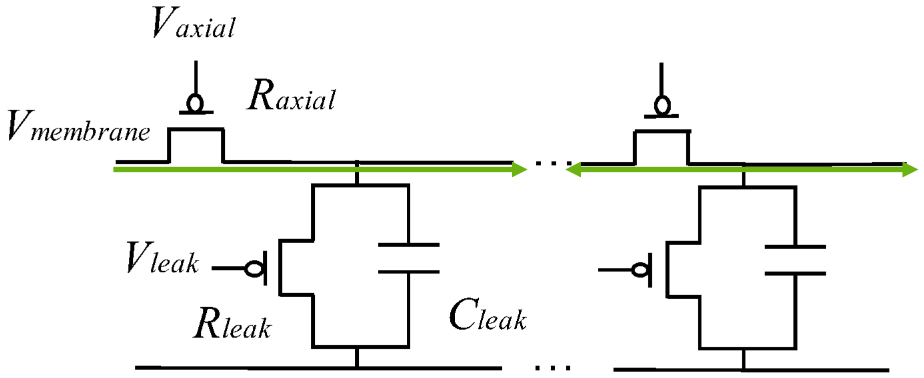
- Inputs travel to neighbors as well as output

Several methods to implement in hardware

Almost compute-on-wire







Using Resistors & Capacitors to implement Dendrite Chain

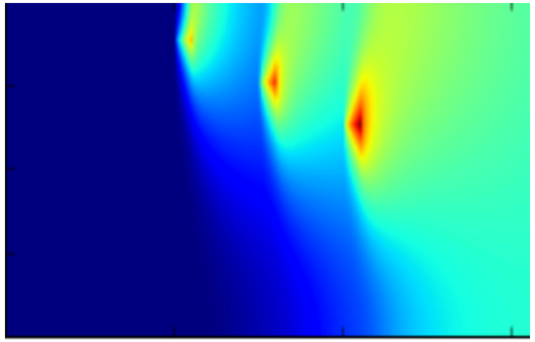
Using CMOS:
Subthreshold Transistors to Implement Dendrite Chain

DENDRITIC UTILITY

Custom hardware is needed to leverage Dendrites

- CMOS based subthreshold based dendrites have been demonstrated to work
- Exploration of beyond CMOS devices as well
  - Memristors, SONOS floating-gate, and more…

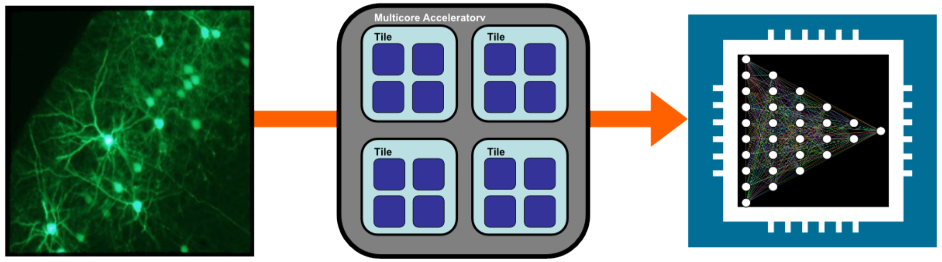
snnTorch



The chicken & egg of novel AI

- In order to justify novel hardware adoption, good software use-cases must exist
- But software developers will use the best current hardware and libraries

An easy-to use dendrite layer in a major ML library could help experimentation and development of dendrite and spiking networks



DENDRITE ENABLED SPIKING LIBRARY

Implemented Torch library with a dendritic chain

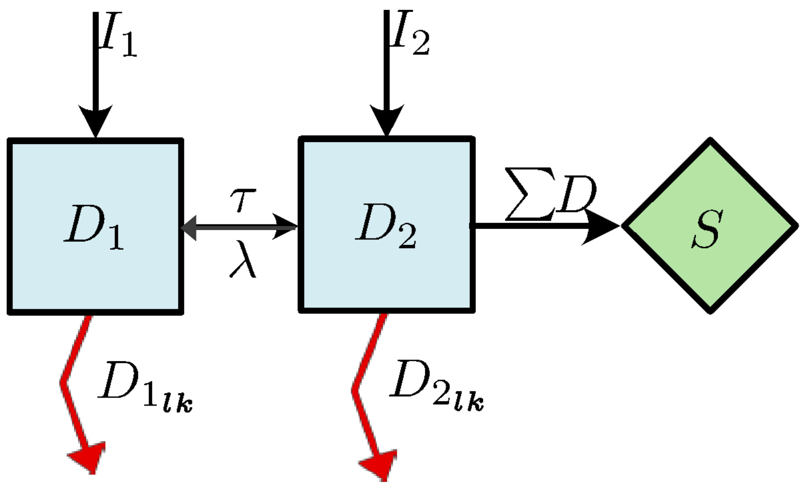
- Simplified version of the complex ODE dendrite solution
- Wrapped dynamics into a set of constants and parameters

Dendrites support SNNTorch & Non-Spiking Torch

Provides a trainable 1-D chain of dendrites

- Spiking or continuous output
  - Works with SNNTorch models or PyTorch models

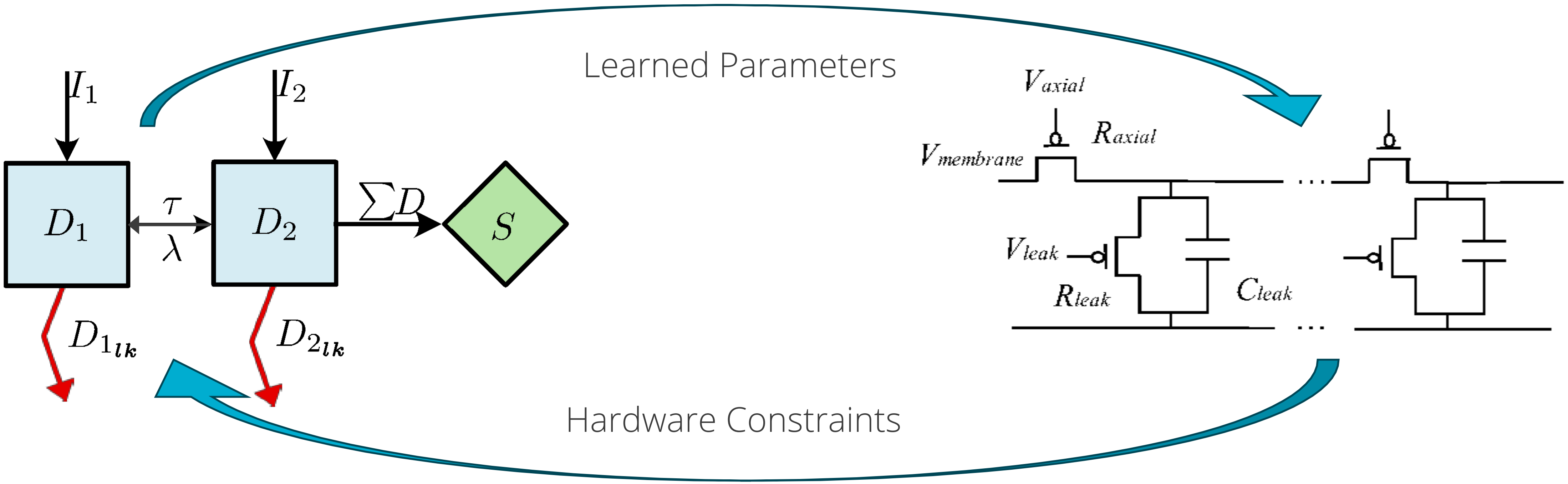
Value	Type
Lambda	"Spatial" constant: Represents Distance
Tau	"Temporal" constant: Capacitance and Resistance
Leak	Signal loss for each tap
Input Weight	Increases or Decreases signal strength



SNNTORCH DENDRITE LIBRARY

Abstract dendrite implementation

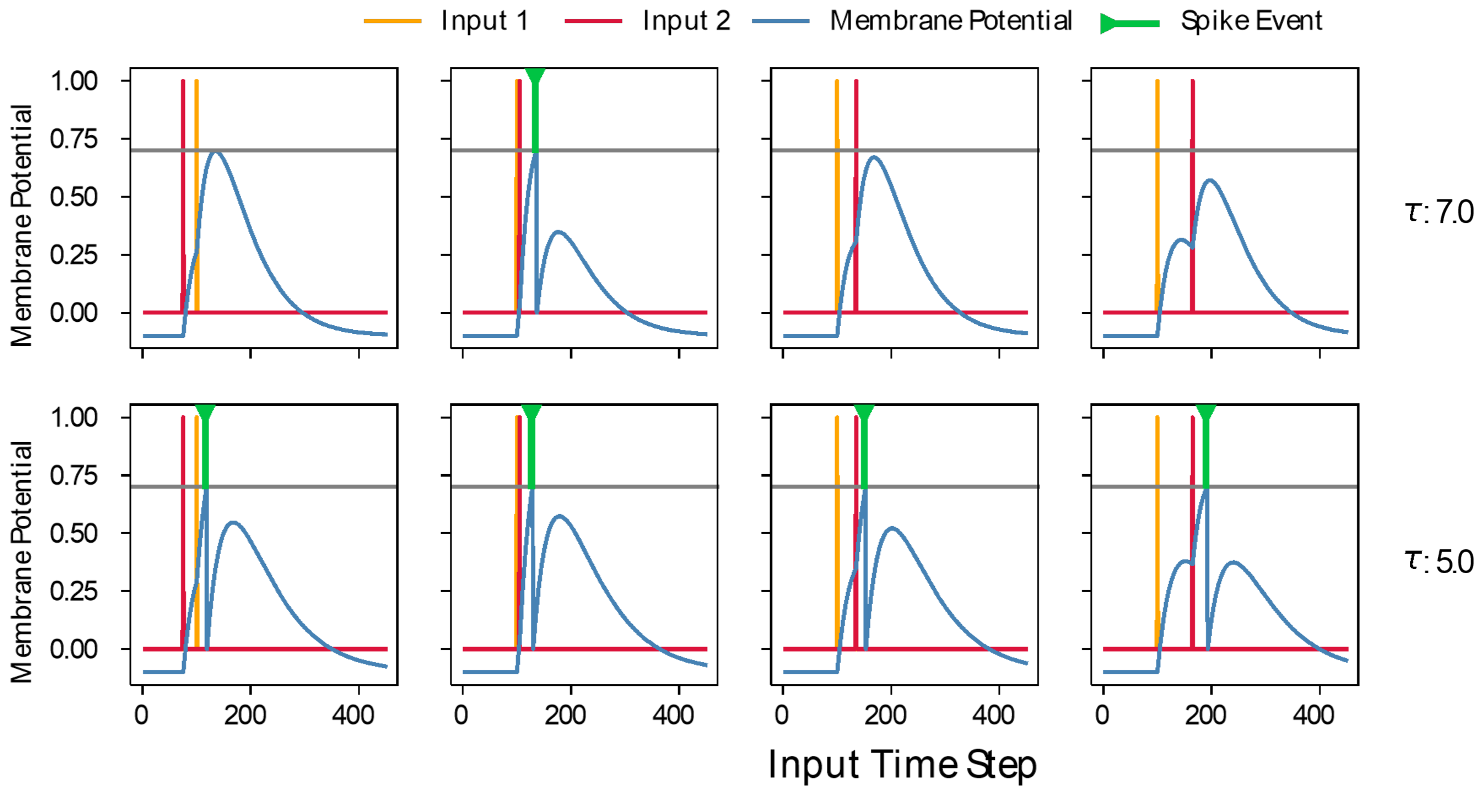
- Based on analog hardware design
- Goal to enable ML training that is transferrable to dendritic hardware
  - Hardware constraints (number of taps, possible fixed values, etc.) to software
  - Learned parameters to hardware

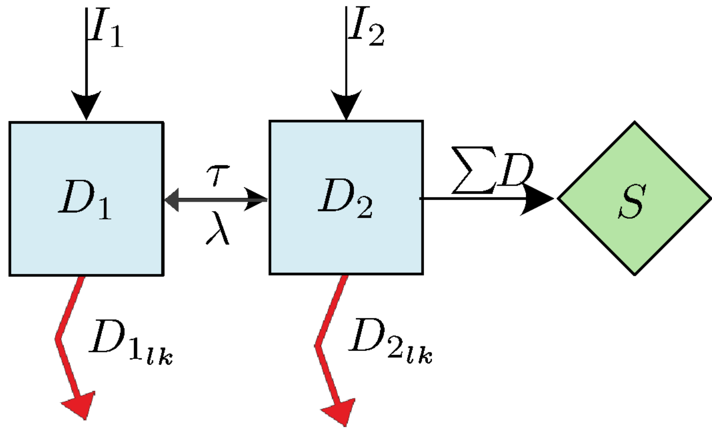


DENDRITIC COMPUTATION – BASIC COINCIDENCE NETWORK

A single dendrite-enabled neuron is capable of basic coincidence detection

The nonlinear temporal dynamics allow for a “time-based AND gate”

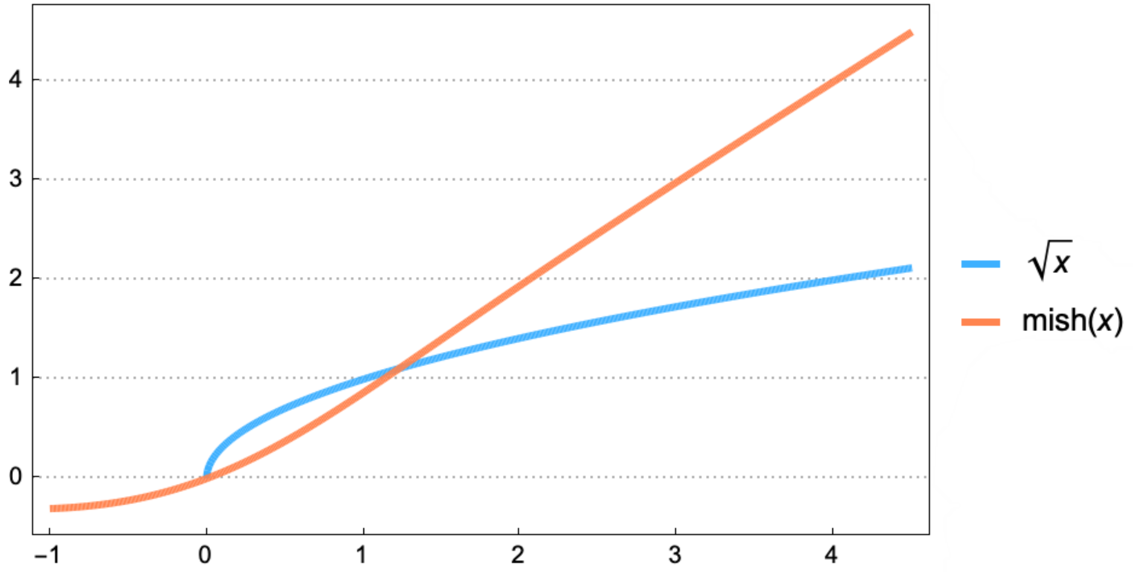




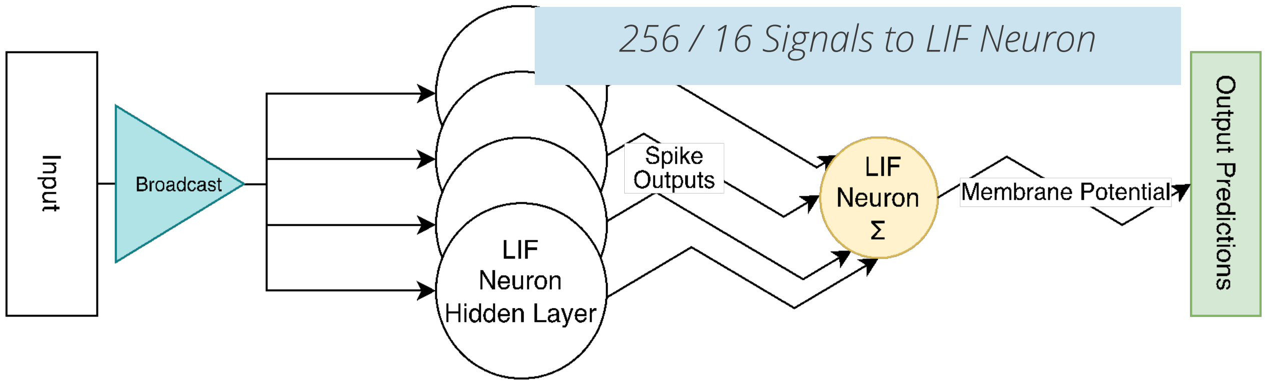
DENDRITE + SNN NETWORKS

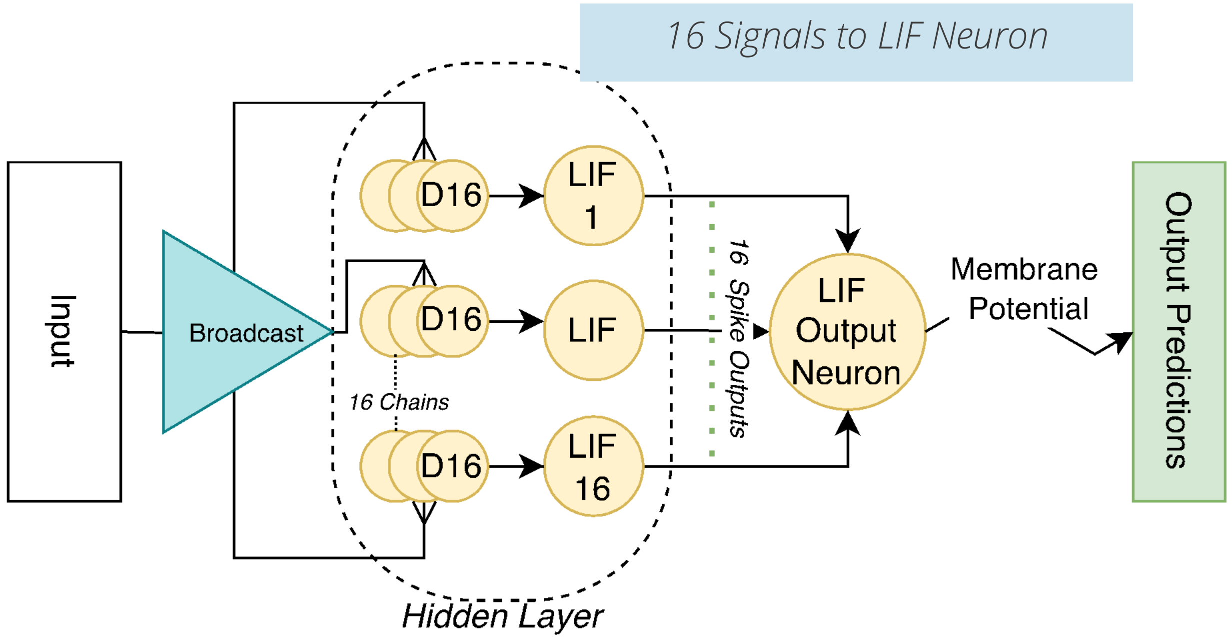
- Learn a pair of nonlinear functions using SNNs
- Based on example spiking networks from the SNNTorch library

- Learn two functions:
  - √x
  - Mish(x)
- Collected a set of 1,000 random samples of each function
  - *> 0 and ≤4*
- Trained all networks for 100 epochs



DENDRITE + SNN NETWORKS





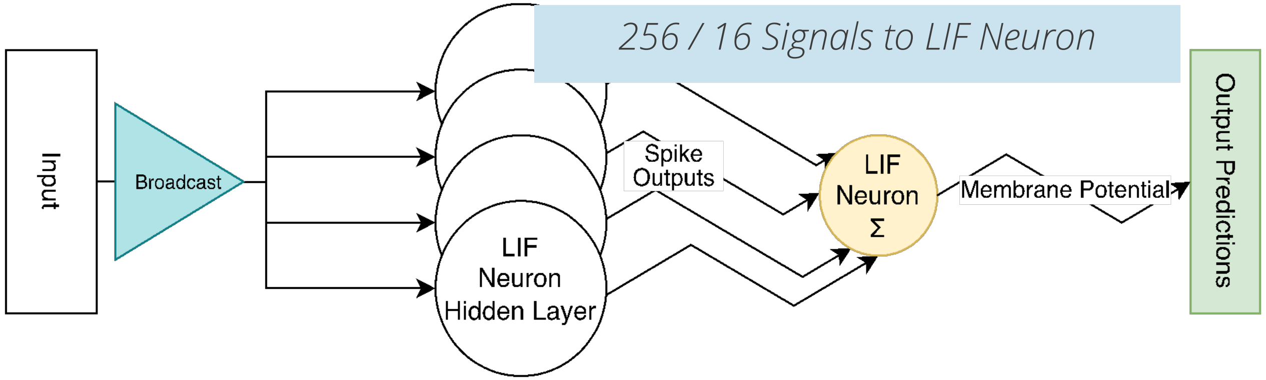
- Created three networks:
- 256 LIF Hidden Layer network
  - Sends 256 spikes to the output neuron
- 16 LIF Hidden Layer network
  - Sends 16 spikes to the output neuron
- 16x16 Dendrite Layer
  - Sends 16 spikes to the output neuron

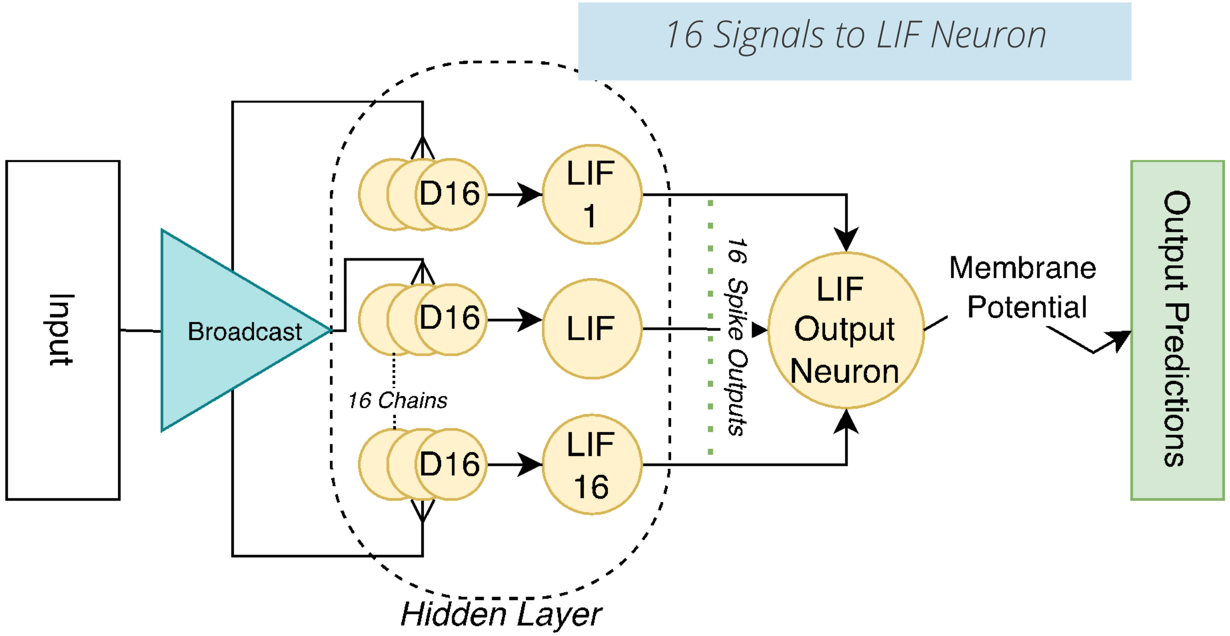
Compare signals sent to output layer against accuracy

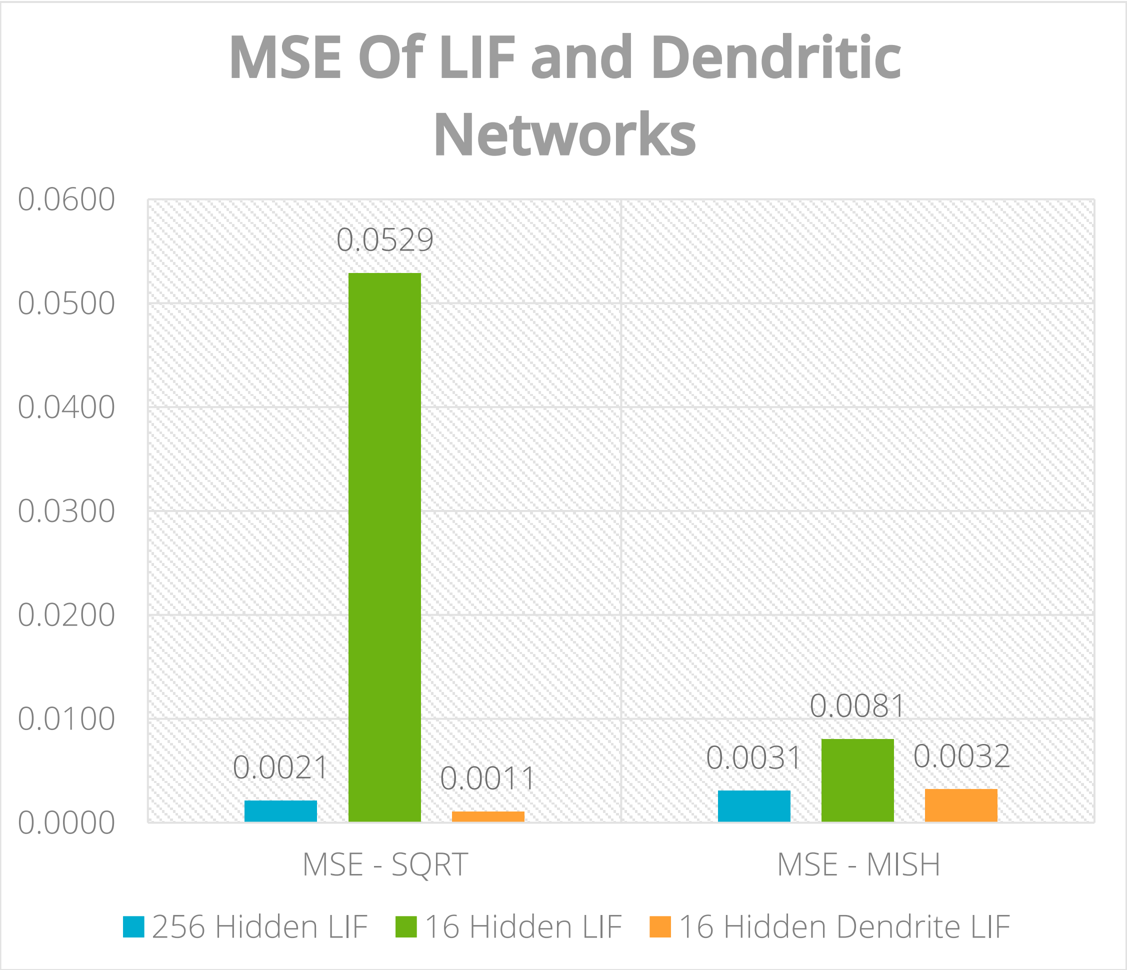
Tau and Lambda were learned:

- Maximum capacitance required in hardware ≤ 100pf

DENDRITE + SNN NETWORKS







FUTURE WORK

- Further develop links with Dendrite-SNN hardware simulations – SanaFe
- Work on a spiking self-attention network with dendrites:
  - Dendritic attention layer (Temporal coherence and context)
  - Dendritic pooling layers (More efficient summary layer)
- Other compelling network designs
- Release as stand-alone library or as SNNTorch add-on

SanaFe – Hardware Simulator

An in-progress tool to estimate timing and energy of neuromorphic systems. Currently supports Loihi. Dendrites are WIP

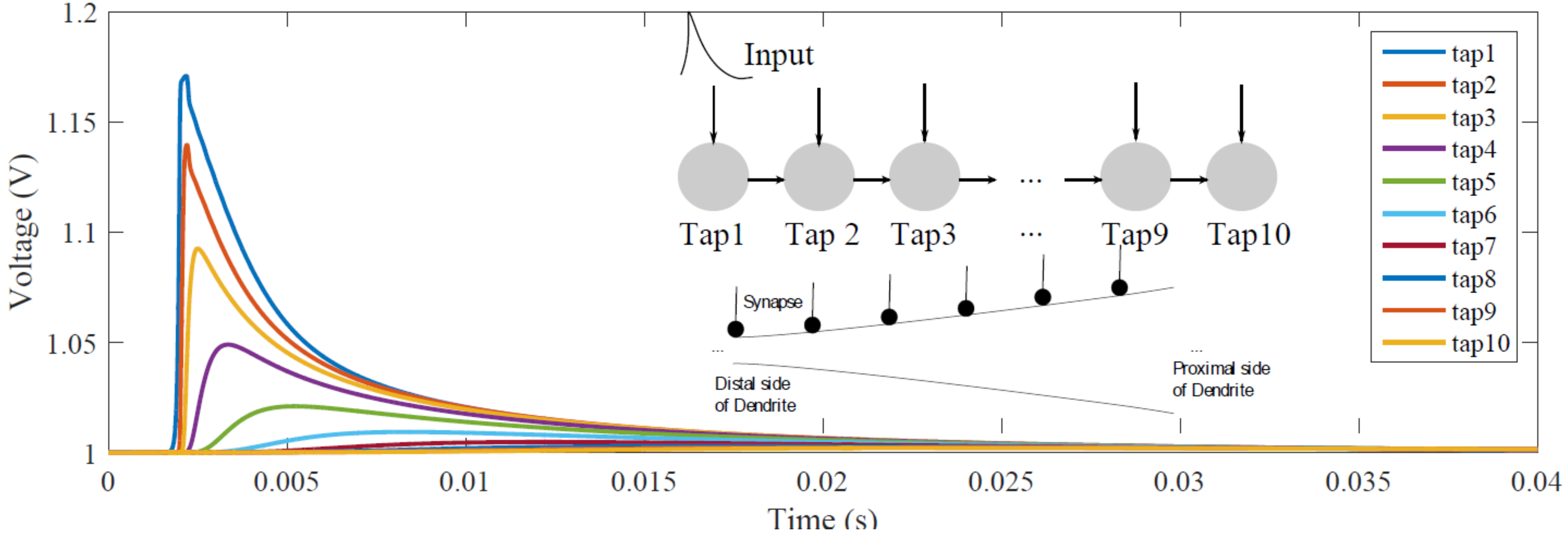
FUTURE WORK

- Further develop links with Dendrite-SNN hardware simulations – SanaFe
- Work on a spiking self-attention network with dendrites:
  - Dendritic attention layer (Temporal coherence and context)
  - Dendritic pooling layers (More efficient summary layer)
- Other compelling network designs
- Release as stand-alone library or as SNNTorch add-on

SanaFe – Hardware Simulator

An in-progress tool to estimate timing and energy of neuromorphic systems. Currently supports Loihi. Dendrites are WIP

DENDRITE BEHAVIORAL MODEL



Simulation model developed to characterize the subthreshold transistor characteristics for the circuit

<u>Power Estimate</u>

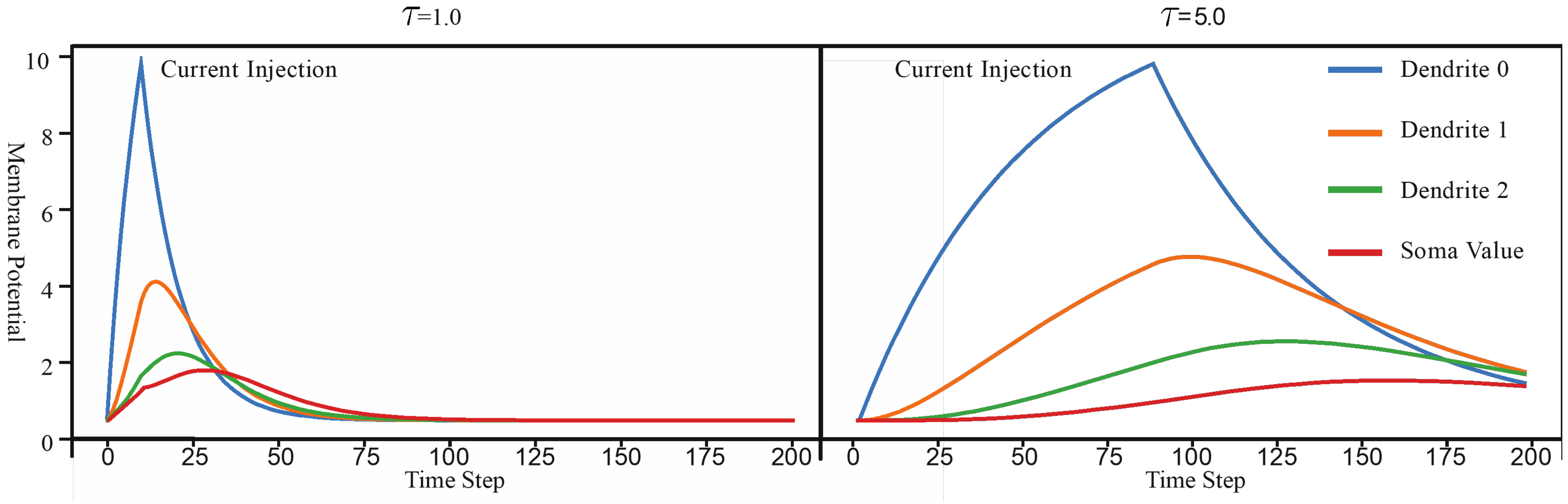
- The dendritic line performs 2MACs/node/0.1ms.
- This gives us efficiency of 20,000MAC/s for a single node.
- For the 350nm node with an FPAA dendrite grid we can achieve ~ 10MMACs/uW.

DENDRITIC COMPUTATION

Dendrites have complex interactions with signals

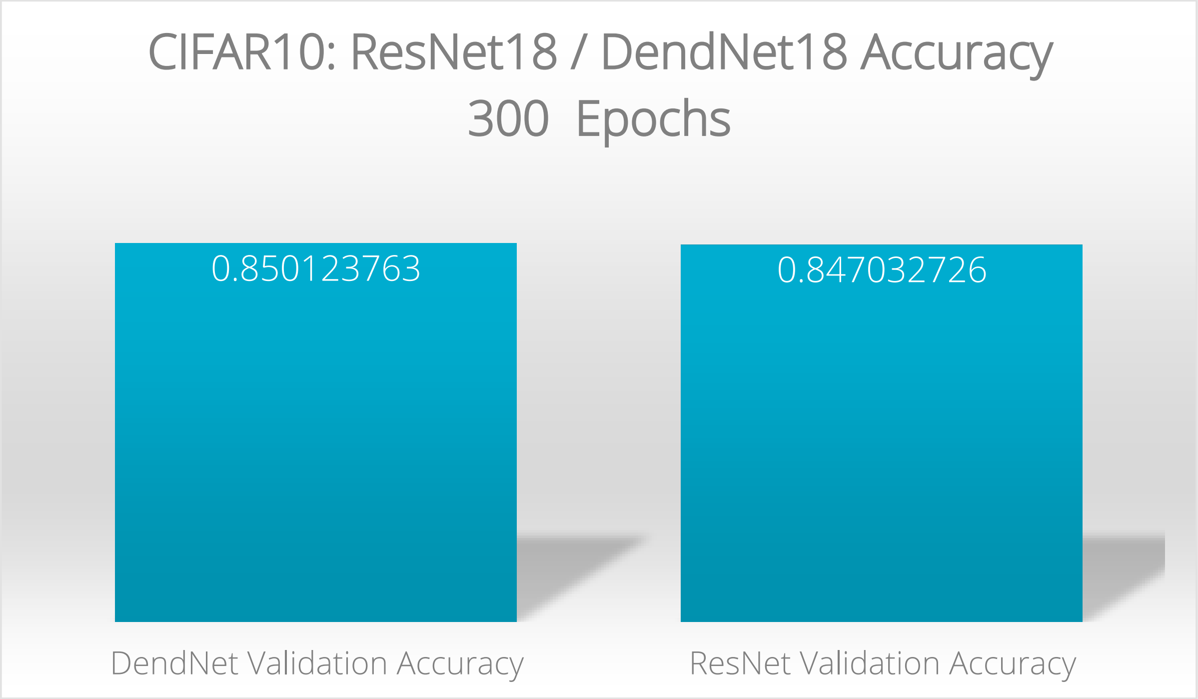
Able to both excite and shunt signals

Temporal component – signals take time to propagate down the wire to the Soma



ANN NETWORK APPLICATIONS – RESNET18

- Working with a graduate student intern at SNL
- Implemented a “Dendrite Pooling Layer” for use in AI ML
- Replaced traditional pooling layer with Dendrite Layer
- Trained ResNet18 on CIFAR-10 for 300 epochs
  - ResNet + Dendrite layer took significantly longer to train
  - Simplified ODE layer adds state and loops
- Found accuracy to be comparable
  - Dendritic pooling has potential in ANNs



Working with Priyam Mazumdar

ML NETWORK APPLICATIONS – RESNET18

- ResNet18 – Was slower to train with a dendritic layer
  - In hardware however, dendrites will be highly efficient

- Rough estimate of efficiency based on
  - Energy = $C(V_{mem} - Ek)V_{dd} = 500fJ$
    - C = 10pF
    - V_{dd}=2.5V
    - $V_{mem} - Ek$ = 100mV

  - Nvidia Jetson values from Rodrigues, et. al

Pooling Layer on Digital Nvidia Jetson	504.41 Micro Joules
Dendritic pooling	0.265 Micro Joules

Rodrigues, Crefeda Faviola, Graham Riley, and Mikel Luján.
"Energy predictive models for convolutional neural networks on mobile platforms."
arXiv preprint arXiv:2004.05137 (2020).