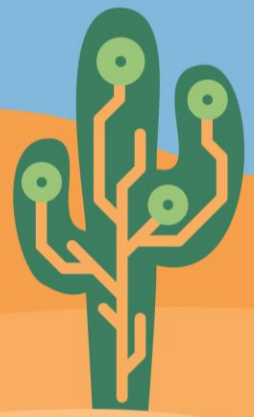


MWSCAS 2023
66TH IEEE INTERNATIONAL MIDWEST SYMPOSIUM ON CIRCUITS AND SYSTEMS

Efficient Implementation of a Fully Analog Neural Network on a Reconfigurable Platform

Afolabi Ige, Jennifer Hasler.

Georgia Institute of Technology

Analog ML & FPAAs

- Analog v Digital Computing

- Interesting points from history
  - Compute in Memory
  - Analog ML: XOR in a single layer
  - Field Programmable Analog Arrays

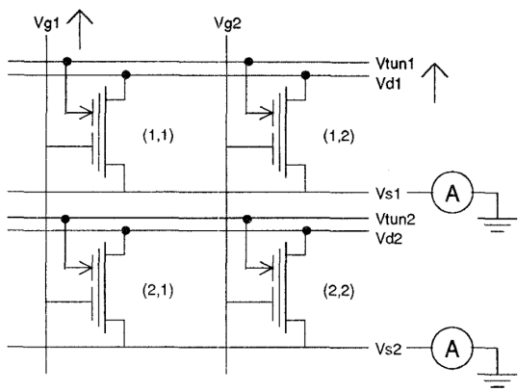
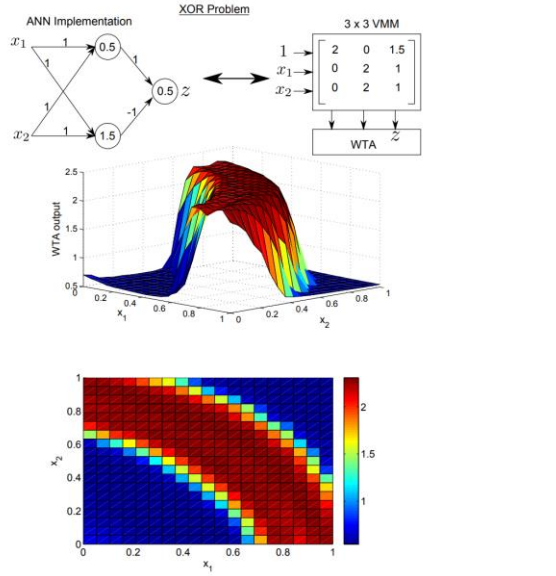
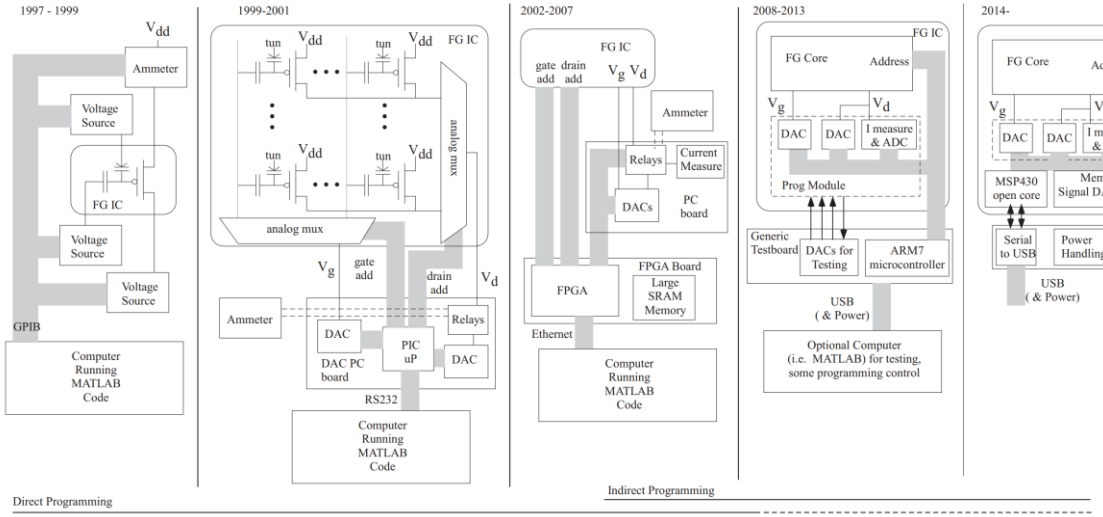


Figure 1: Circuit diagram of the single - transistor synapse array. Each transistor has a floating gate capacitively cou-

Hasler et al. Single transistor learning synapse with long term storage. 1995



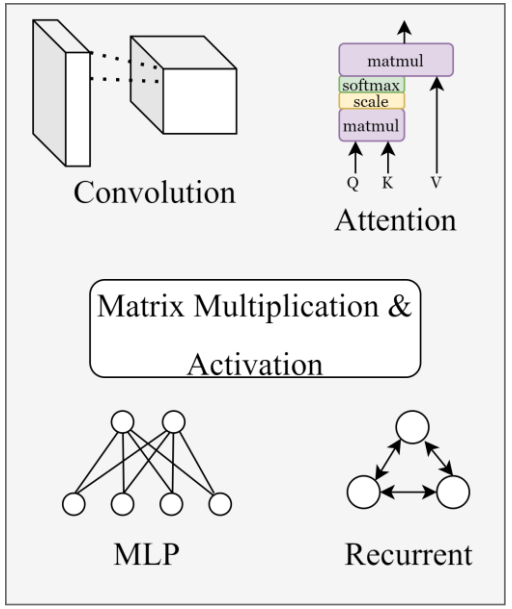
S, Ramakrishnan et al. A compact programmable analog classifier using a vmm + wta network. 2013.

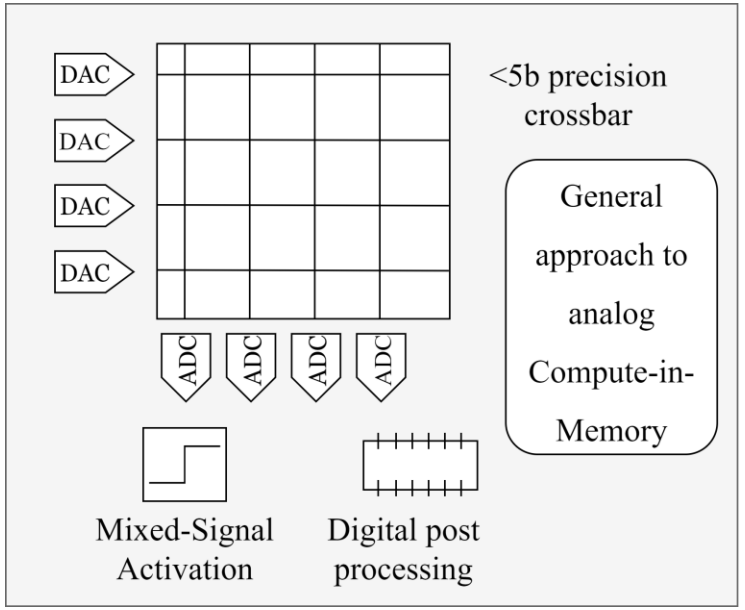


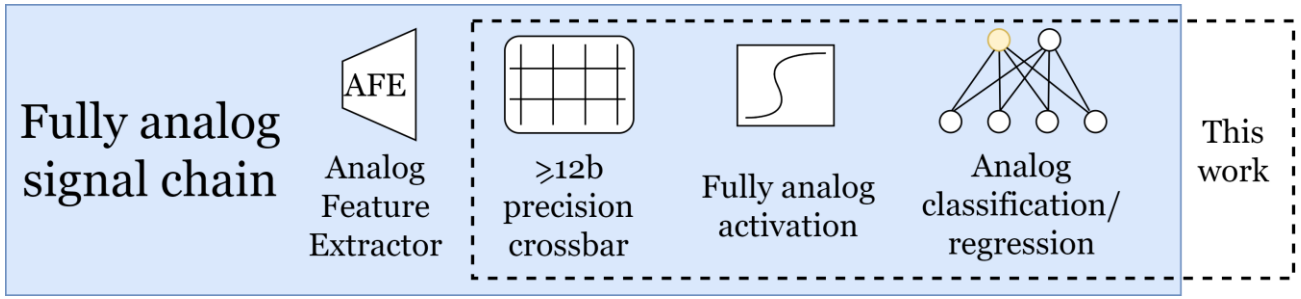
S. Kim, J. Hasler, S. George. Integrated Floating-Gate Programming Environment for System-Level ICs. 2015.

Overview

- Neural Networks are Matrices and Activations
- Data converter bottleneck in CIM.
- Fully Analog Networks break through.



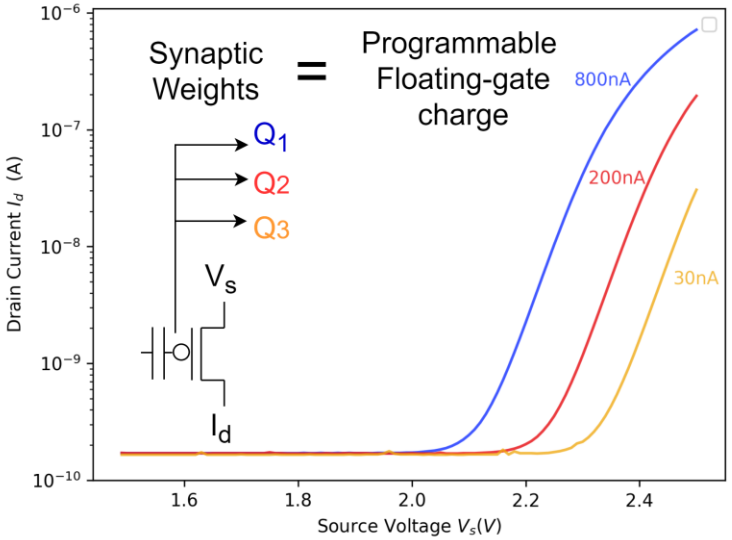


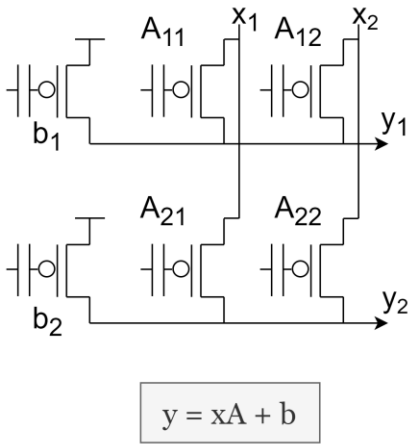


The Crossbar

- Floating Gates make great cross bars!
  - High precision
  - 10-year retention, no refresh
  - Established Fab Process
- Source v Gate Input
  - Source is an exponential reduction of weight. Better selectivity.
  - Gate is an exponential increase of input. FETs harder to shut off.

	Precision	Drift ?	Current draw	Fabrication Maturity
ReRAM [Wan, w. Nature '22]	~4 – 5 bits	Yes	mA	No
Digital	4/16 bits	No	mA	Yes
Phase Change memory [Gallo, M. J Phys. D Appl Phys '22]	~ 3 bits	Yes	uA	No
Floating Gates [Sihwan, K. TVLSI '16]	13 bits	No	pA to nA	Yes

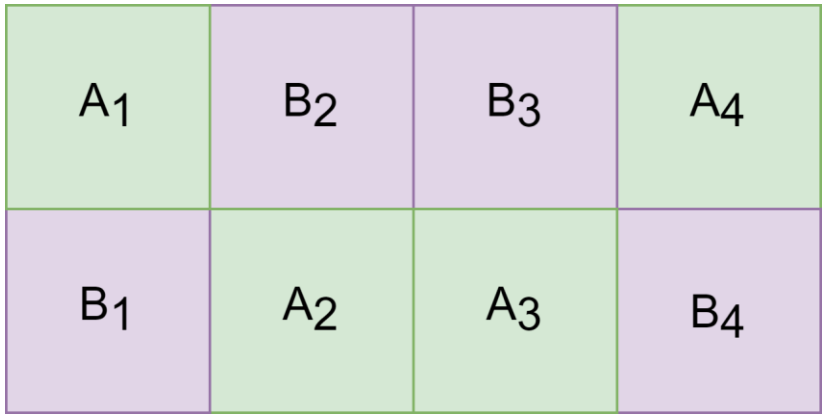




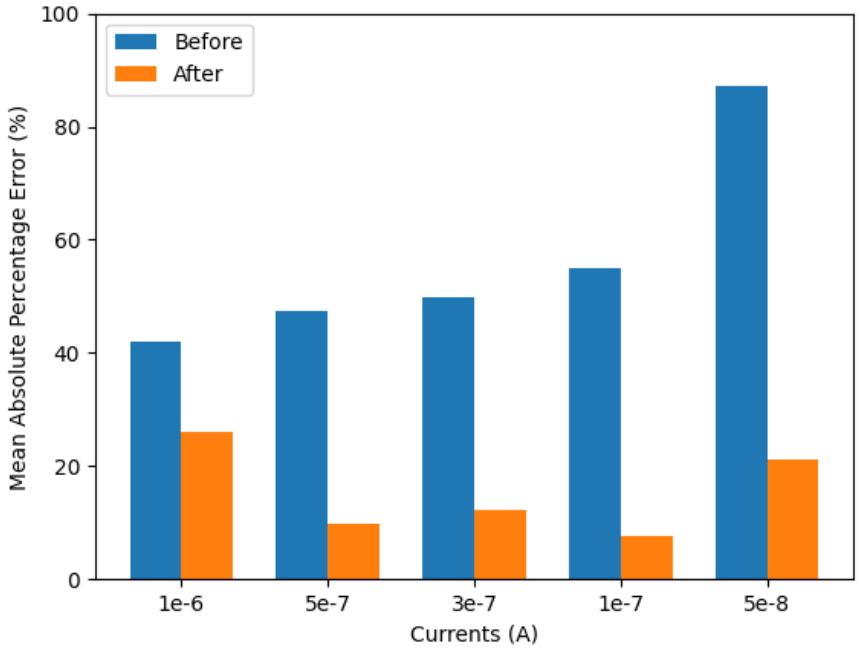
$y = xA + b$

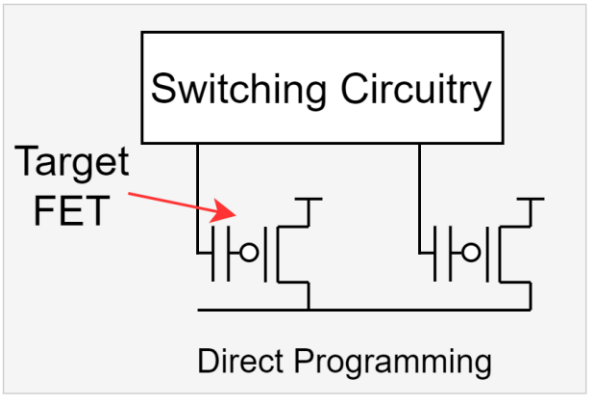
FG Calibration

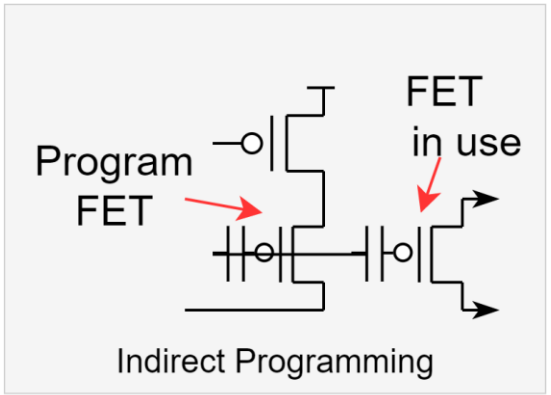
- Fabrication mismatch
- Program device mismatch
- Floating Gate Calibration



Techniques for combatting device variation

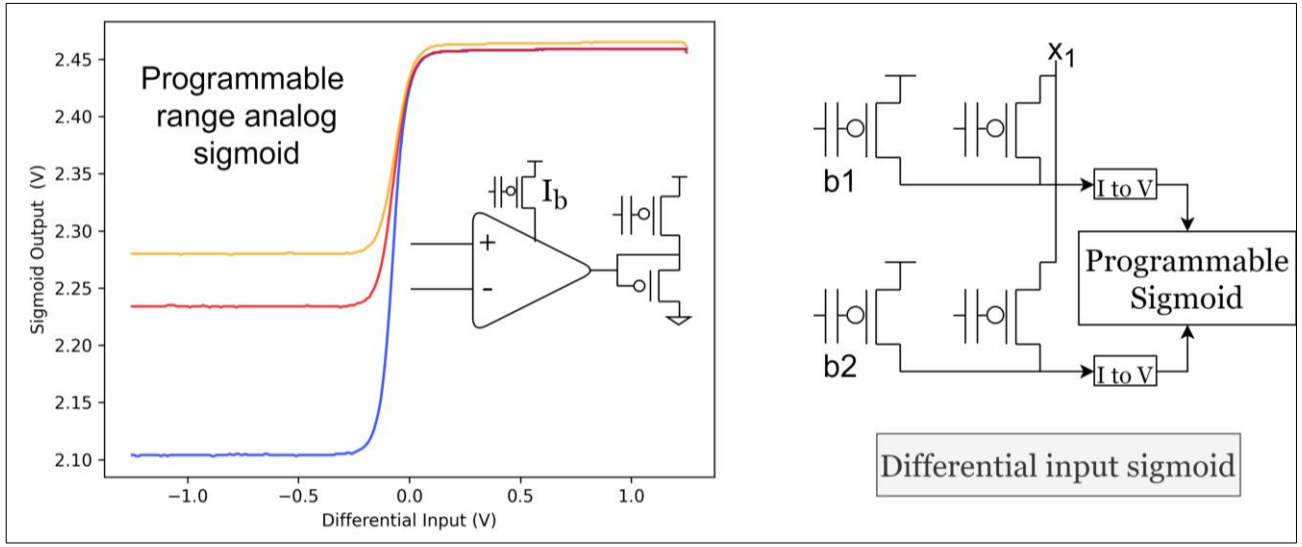


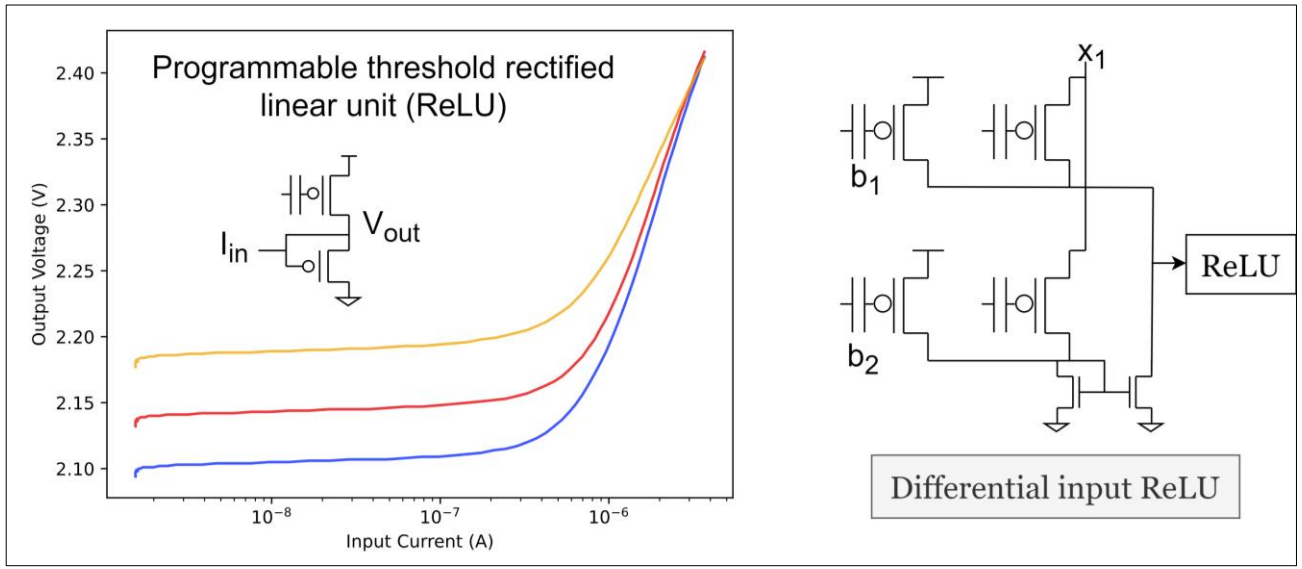




Analog Activations

- Activation functions need a programmable bottom.
- The Sigmoid and Rectified Linear Unit are well studied non-linear functions.
- Both handle differential weight schemes.

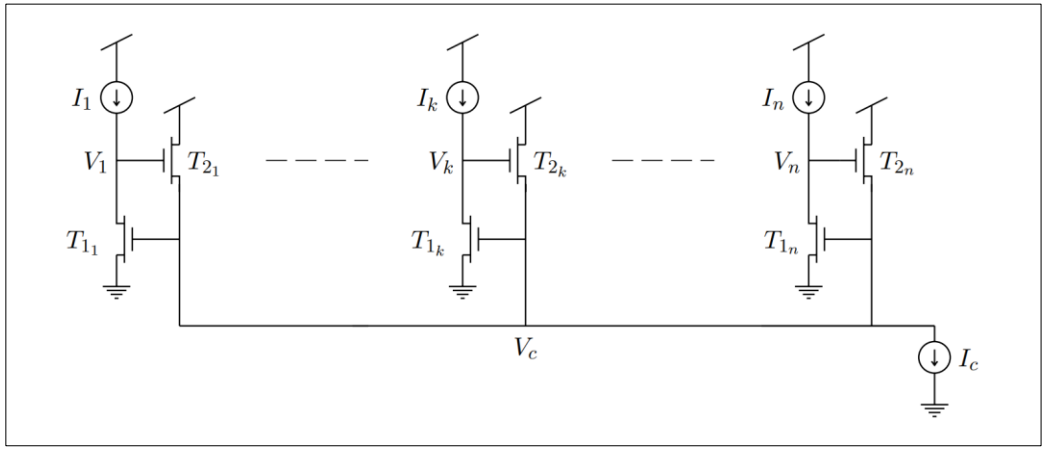




Analog Classification

- The softmax function normalizes inputs similar to the winner-take-all.

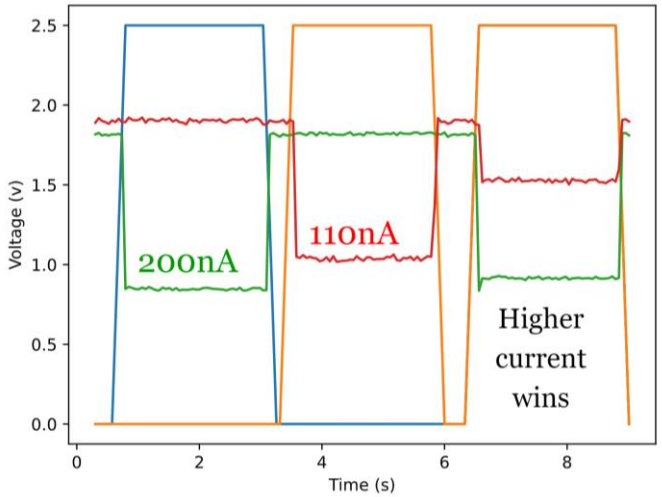
$$\sigma(x) = \frac{e^{x}}{\sum_{i=1}^{k} e^{x_i}}$$

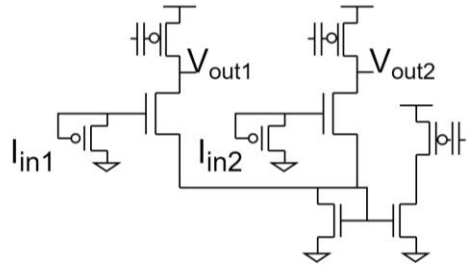


Original Lazzaro Winner-Take-All (WTA) Circuit. [J, Lazzaro et al. 1989]

- The WTA was modified for density and programmability on the FPAA.

- Last layer of the network measured on chip



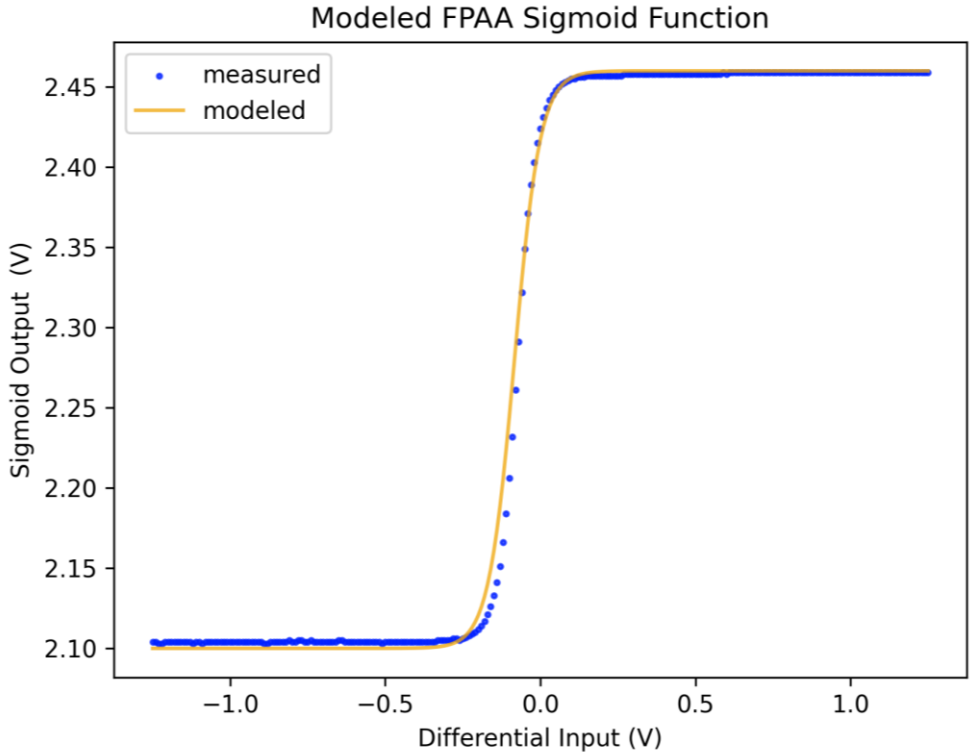


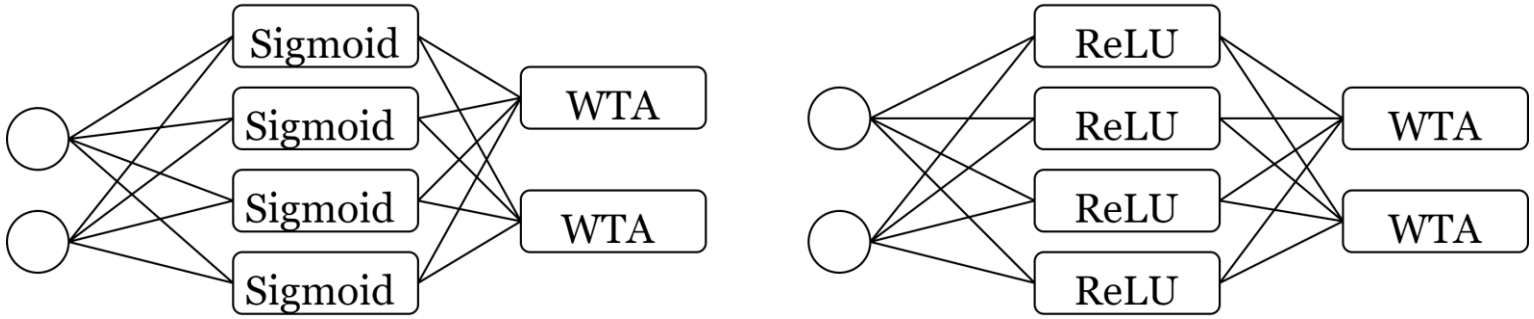
Winner-Take-All as a Softmax

Modified FPAA WTA Circuit

Training & Mapping

- Model analog activation
- Train digitally
- Scale weights to hardware
- Finetune for accuracy





FPAA Neural Network Structure

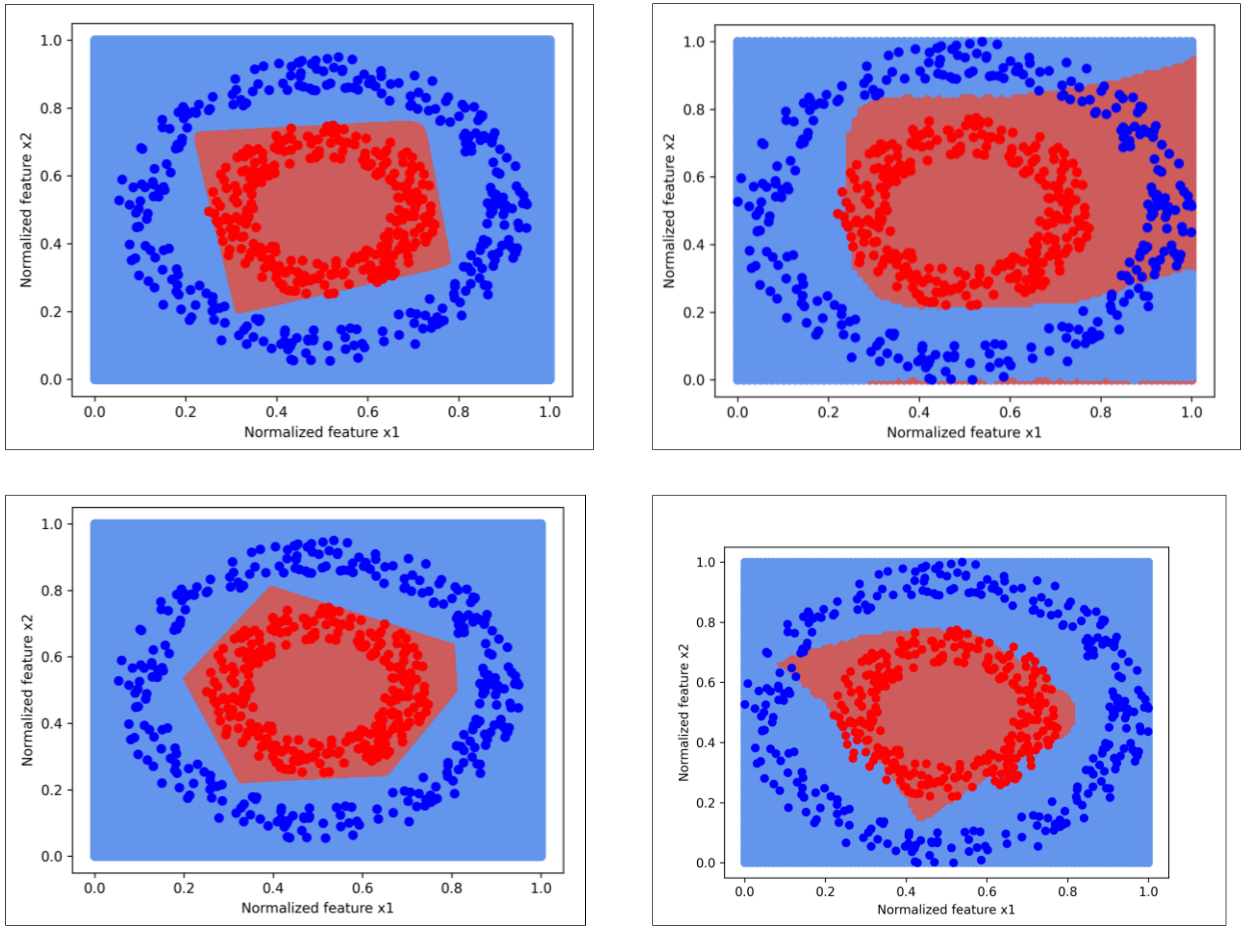
Analog Neural Network

- Task:
  - Solving the concentric circles problem [1]
- Architecture:
  - 2-Layer NN

	Accuracy	Power
Sigmoid	84.8 %	$20\ \mu W$
ReLU	94 %	$80\ \mu W$

Sigmoid

ReLU



Digital

Analog

[1] Visualize @ TF Playground: https://playground.tensorflow.org/

Questions?