Deep Learning with Analog Neuromorphic Hardware
1888 Neuron Doctrine
1917 LIF Neuron Model
1929 EEG

1943 McCulloch-Pitts
1958 Perceptron
1986 Backpropagation
2012 AlexNet

2001 HAGEN

2005 Spikey
2009 HICANN
2011 BSS1 Wafer
2015 HICANN-DLS
2019 BSS2 HICANN-X

1949 Hebbian Learning
1957 Hodgkin-Huxley Neuron Model
2005 AdEx Neuron Model
2018 Loihi
Biophysical Emulation

\[
C_{\text{m}} \frac{dV_{\text{m}}}{dt} = -g_{\text{leak}}(V_{\text{m}} - V_{\text{leak}}) + I_{\text{stim}}
\]

\[ C_{m} \Delta V_{m} \Delta t = -g_{\text{leak}} (V_{m} - V_{\text{leak}}) + I_{\text{stim}} \]

Biophysical Emulation

\[ C_m \frac{dV_m}{dt} = -g_{\text{leak}} (V_m - V_{\text{leak}}) + I_{\text{stim}} \]

Measure

Model

\[ C_m \frac{dV_m}{dt} = -g_{\text{leak}} (V_m - V_{\text{leak}}) + I_{\text{stim}} \]

Measure by Mohanty, Scholl, and Priebe (2012).
BrainScales2 – Overview

- Hybrid neuromorphic system, 65 nm CMOS
- $1000 \times$ speedup
- 512 multi-compartment AdEx neurons
- $512 \times 256$ synapse circuits
- Two general purpose SIMD processors
- 1024 columnar ADC channels (8 bit)
- 16 Gbit s$^{-1}$ (full duplex) I/O
BrainScales2 – Overview

- Hybrid neuromorphic system, 65 nm CMOS
- $1000 \times$ speedup
- 512 multi-compartment AdEx neurons
- $512 \times 256$ synapse circuits
- Two general purpose SIMD processors
- 1024 columnar ADC channels (8 bit)
- 16 Gbit s$^{-1}$ (full duplex) I/O

Figure adapted from Billaudelle, Stradmann, et al. (2019).
Simulation vs. Emulation

Figure adapted from Aamir et al. (2018).
Accelerated Emulation of Spiking Neural Networks

Experiment by Korbinian Schreiber (Billaudelle, Stradmann, et al., 2019).
Single Spike Coding – Time to First Spike

Göltz et al. (2019)
Single Spike Coding – Time to First Spike
On-chip Learning

520.8 s

On-chip learning rule:
- STDP
- Homeostasis
- Pruning

Experiment by Billaudelle, Cramer, et al. (2019).
BrainScaleS-2 – Spikes and Activations

### Spiking Mode

- **Pre-synaptic input:** spikes
- **Ion channel circuits:** normal synaptic time constant, normal leakage
- **Membrane capacitance:** PSP

### Perceptron Mode

- **Pre-synaptic input:** multi-valued
- **Ion channel circuits:** no leakage
- **Membrane capacitance:** ReLU
Inference on BSS2 – (Very) Early Results: Analog MAC

Unpublished, measurement designed and executed by Johannes Weis.

Input resolution: 5 bit
Weight resolution: 6 bit + sign
Activation resolution: 8 bit
Analog precision: ???
Inference on BSS2 – (Very) Early Results: MNIST

- **Simple Architecture**
  - One convolutional layer (10 × 10)
  - Two dense layers (128 units, 10 units)

- **Achieved accuracy**
  - Software: 98.42%
  - Hardware: 91.54% (without re-training)

Unpublished, measurement designed and executed by Johannes Weis.
Model Creation – Hardware in the Loop

Graph Transformations:
- Reduction to av. operations
- Mapping in place & time

Compute
Graph

Activations

CPU/GPU
Cluster

Model Development

Analog Inference
Backend
Model Application – Analog Inference

Graph Transformations:
- Reduction to av. operations
- Mapping in place & time

CPU/GPU Cluster

Compute Graph

Activations

Model Development

Analog Inference Backend
In the (near) future…

Software
- “Hardware in the Loop” ANN training
- TensorFlow/PyTorch integration
- SNN abstraction layer
- Compiler support for SIMD operations

Hardware
- Tape-Out in 02/2020
- Inference throughput: up to 131 GOPS
- Spike throughput: up to 250 MEvents s\(^{-1}\)
- Power consumption: \(\approx 1\) W

February 13, 2020
Yannik Stradmann
Slide 14
Summary

BrainScaleS-2 is an analog neural network accelerator

… manufactured in an affordable 65 nm CMOS process

… suitable for artificial neural networks

… suitable for spiking neural networks (1000× speedup)

… optimized for low-power applications

… embedding SIMD microprocessors for on-chip learning
References


