COMPARING NEUROMORPHIC SOLUTIONS IN ACTION

Alan Diamond, Thomas Nowotny, and Michael Schmuker

School of Engineering and Informatics
University of Sussex, Brighton, UK

7 March 2016, NICE Workshop, Berkeley
Motivation

- Neuromorphic hardware is ready to address real-world computing tasks today.

- Several platforms are available:
  - Dedicated digital hardware (e.g. SpiNNaker, TrueNorth)
  - Dedicated analog / mixed-signal hardware (Spikey, HICANN)
  - Generic hardware (CPU, GPU)

- Relevant constraints for choosing a substrate are
  - Computational speed
  - Power consumption
  - Convenience (programming / configuration)

- Question: “Given task X, which platform is best?”
  - What is the tradeoff between power, speed and convenience?
Platforms compared

- **SpiNNaker**
  - Specialised multicore system.
  - We used a “small” system (SpiNN-3).
  - Low power, small form factor, readily available.

- **Spikey**
  - Mixed-signal system with analogue neurons and digital routing.
  - 10,000 fold speedup over real-time.
  - Small footprint, low power, available remotely.

- **GeNN (GPU-Enhanced Neuronal Network)**
  - Meta-compiler that generates optimised CUDA-kernels.
  - Hardware: NVIDIA Titan Black, 2880 cores, 6GB, PCIExpress.
Task: Pattern recognition

- Classify handwritten digits, MNIST dataset.
- Well established, non-trivial classification task.
- High dimensional feature space (28x28=784 pixels).
- Up to 10 different classes (digits).
- Many samples to process (60,000 training, 10,000 testing)

The Virtual Receptor (VR) concept

- Problem: network requires firing rate input (positive & bounded), but multivariate data sets are generally real-valued.
- Solution: Virtual receptors (VRs).
- Example: 10 VRs, placed by self-organising, unsupervised algorithm.
- “Response” $r$ of VR $p$ depends linearly on distance $d$ to data point $s$:

$$r = 1 - \frac{d(s,p) - d_{\text{min}}}{d_{\text{max}} - d_{\text{min}}}$$

- Result: N-dimensional, positive, bounded representation of the multivariate data set. N is the “Number of VRs”.

Prof Thomas Nowotny
CENTRE FOR COMPUTATIONAL NEUROSCIENCE AND ROBOTICS, SUSSEX NEUROSCIENCE
SCHOOL OF ENGINEERING AND INFORMATICS
Conceptual network model, inspired by insect olfaction

- Spikey model follows closely this concept.
- 6 + 6 neurons per VR, 8+8 neurons per class
GeNN implementation

- Lightweight “map” neurons (similar to Izhikevich).
- Inhibitory interneurons abstracted out.
- Simulation on GPU individually stepped under workstation control.
- Plasticity implemented by external learning rule run on workstation (like on Spikey).
- Spikes extracted on every timestep. New weight matrix uploaded periodically.
- 30 neurons per VR, 30 neurons per class.
SpiNNaker implementation

- Lightweight LIF neuron model.
- Learning via Hebbian association on-device by combining teaching signal with STDP.
- On-chip spike-train “multiplication” using a shared Poisson population.
- Model invoked once for the whole learning stage with STDP, then rebuilt for testing with static weights & invoked once.
- Spike source inputs are generated *a priori* for the full learning and test runs and uploaded before the simulation.
- Considerable juggling with PyNN populations and connection matrices was required to efficiently use the resources.
Spike trains on the hardware

**Spikey**

- 1 s per stimulus (bio-time).
- 192 neurons available.
- Highly optimised for low neuron count: 6 to 8 neurons per population, long stimulus presentation, max. 10 VRs, max 3 classes.

**SpiNNaker**

- 120 ms per stimulus (real-time).
- Large neuron count allows larger populations, shorter presentation.
- Up to 6300 neurons used (200 VRs), 30 neurons per population, all 10 digits.
Classification performance

- Recognition performance increases with #VRs (SpiNNaker & GeNN: 200 VRs ≈ 6000 neurons)
- Neuron capacity is crucial for model performance (Spikey suffers from low neuron count)
- Performance decreases as more digits are learned (problem gets harder)
Classification results: SpiNNaker vs. GeNN

- Results from SpiNNaker and GeNN are comparable, in spite of the differences in implementation.

- This suggests that classification performance is determined by the model, rather than the platform *in spite of the any implementation differences*. 
Data preparation and transfer dominates run time for all platforms.

GPU looks good because of very high bandwidth.

SpiNNaker has constant simulation time, but uploading data is slow.

Spikey loses its speed advantage in transferring data and configuring the chip.
Power consumption

- Power was measured using an inline meter from the mains connection.
- The level graphed here includes the baseline power of devices but not the host Workstation.
Total energy used

- A lot of energy is used on the host workstation during data processing and communication.
- The GPU uses considerably more energy than SpiNNaker or Spikey.

Method

- Energy was calculated as power drawn integrated over time.
- Measurements were repeated for a small and a large model (10 and 200 VRs).
Conclusions

- The olfaction-inspired classifier scales to large data sets and large networks.
- **Classifier performance** is largely *independent* of the platform.
- In all cases, **speed** is largely determined by the **bandwidth** to/from the substrate, not by the neural computation: Neuromorphic hardware would benefit from higher bandwidth.
- Tuning the software interface may become crucial for large-scale applications.
  - Tune for throughput, but also provide better control over what is to be transferred.
- Neuromorphic algorithms should be designed to use as little communication between host and substrate as possible.
  - On-chip plasticity instead of host-based learning rules.
  - On-chip spike train generation instead of transferring spike times.
- Energy consumption is dominated by host for SpiNNaker and Spikey but GPU is hungry in its own right.
Acknowledgments

- Alan Diamond
- Michael Schmuker

Human Brain Project

GeNN
https://github.com/genn-team/genn

doi: 10.3389/fnins.2015.00491