The SpiNNaker project

Steve Furber
ICL Professor of Computer Engineering
The University of Manchester

200 years ago...

- Ada Lovelace, b. 10 Dec. 1815

"I have my hopes, and very distinct ones too, of one day getting cerebral phenomena such that I can put them into mathematical equations—in short, a law or laws for the mutual actions of the molecules of brain. .... I hope to bequeath to the generations a calculus of the nervous system."
70 years ago…

Bio-inspiration

• Can massively-parallel computing resources accelerate our understanding of brain function?

• Can our growing understanding of brain function point the way to more efficient parallel, fault-tolerant computation?
ConvNets - structure

- Dense convolution kernels
- Abstract neurons
- Only feed-forward connections
- Trained through backpropagation

The cortex - structure

- Spiking neurons
- Two-dimensional structure
- Sparse connectivity
convnets - GPUs

- Dense matrix multiplications
- 3.2kW
- Low precision

Cortical models - Supercomputers

- Sparse matrix operations
- Efficient communication of spikes
- 2.3MW
Cortical models - Neuromorphic hardware

- Memory local to computation
- Low-power
- Real time
- 62mW

Start-ups and industry interest

Samsung uses IBM’s brain-inspired chip to recognize gestures

Manchester-based MindTrace secures initial €1.5 million to create self-learning machines
SpiNNaker project

- A million mobile phone processors in one computer
- Able to model about 1% of the human brain...
- ...or 10 mice!

SpiNNaker system
**SpiNNaker chip**

**Chip resources**

- **Router**
  - routing tables
  - spike packet routing
  - system comms.

- **RAM port**
  - synapse states
  - activity logs

- **Instruction memory**
  - run-time kernel
  - application callbacks

- **Data memory**
  - kernel state
  - neuron states
  - stack and heap

- **Processor**
  - neuron and synapse
  - state computations
Multicast routing

High-level software flow
SpiNNaker machines

- HBP platform
  - 1M cores
  - 11 cabinets (including server)
- Launch 30 March 2016
  - then 500k cores
  - 93 remote users
  - 5,134 remote jobs run
  - >5 million local jobs run

SpiNNaker machines

- 100 SpiNNaker systems in use
  - global coverage
- 4-node boards
  - training & small-scale robotics
- 48-node boards
  - insect-scale networks
- multi-board systems
- 1M-core HBP platform

SpiNNaker board (864 ARM cores)

SpiNNaker chip (18 ARM cores)

SpiNNaker racks (1M ARM cores)
**SpiNNaker applications**

Computational Neuroscience

---

**Cortical microcolumn**

1st full-scale simulation of 1mm² cortex on neuromorphic & HPC systems

- 77,169 neurons, 285M synapses, 2,901 cores
- using as benchmark example, since improved:
  - run-time by x80: 10 hours → 7.5 minutes
  - run speed, from 20x slow-down to real time
  - efficiency, by 10x

Computational Neuroscience

• Serotonin modulates Pre Frontal Cortex
  • neurons express range of serotonin receptors
  • respond at different timescales

• Dorsal Raphe Nucleus stimulation modulates brain rhythms
  • releases serotonin

• Computational model to simulate serotonergic modulation
  • monitor local effects – firing rates
  • understand global effect on connected brain regions – oscillation in local field potential


Computational Neuroscience

• Explore chemistry modulating neuron behaviour
  – intracellular dynamics (ion channels)

• Simulate patch-clamp experiments from biology

• Incorporate findings at larger scales
  – study effect on consciousness
  – multiple brain regions

Jaakko Malmivuo and Robert Plonsey, 1995
Constraint satisfaction problems

Stochastic spiking neural network:
- solves CSPs, e.g. Sudoku
  - 37k neurons
  - 86M synapses
- also
  - map colouring
  - Ising spin systems

work by: Gabriel Fonseca Guerra
(PhD student)

G. A. Fonseca Guerra and S. B. Furber,
Using Stochastic Spiking Neural Networks on SpiNNaker to Solve Constraint Satisfaction Problems, Frontiers 2018.

Network plasticity for learning and memory
  – adjust synaptic connections
  – add/remove connections

HBP Co-Design Project 5
  – functional plasticity for learning on neuromorphic hardware

Bridge the gap from neuroplasticity to machine learning?


Structural plasticity
  • Create/remove connections to facilitate learning/consolidation
    – feedforward and recurrent
    – distance-dependent receptive field
    – pruning of weak connections
  • Computational challenge
    – update connection matrices on-the-fly
    – maintain network dynamics and computational performance


• Transfer machine learning concepts to brain-like spiking neurons
  – Long Short Term Memory (LSTM) units
  – BPTT & SGD

• Train SNNs via error back-propagation
  – recurrent spiking neural networks
  – pseudo differential to overcome discontinuity of gradient at spike

• First deployment on neuromorphic hardware
  – unlock scale and explore performance


Theoretical Neuroscience

- Basal Ganglia – biological decision making and action selection
  - Single channel model inspired by biology: neuron dynamics; numbers; and topology

- Dopamine is central to network function
  - Expressed via two receptor types
  - Explore how modulation relates to scale and disease

SpiNNaker applications

Neurorobotics

Computational Neuroscience

Theoretical Neuroscience
• Classification of electrical signals
  • real-time control of active prosthetics
  • low power
• Record electrical activity of
  participants during prescribed hand
  movements
• Classification with reservoir of spiking
  neurons
  • encode signals into spikes
  • train network (unsupervised)
  • readout to classify


• Study vestibular ocular reflex in iCub robot
  • SpiNNaker as neural substrate
• Learn control via cerebellum inspired
  spiking neural network
  • Range of learning kernels based on
    relative spike timing + error
• Research embodiment of neural
  control systems


Simulation
Computational Neuroscience
Theoretical Neuroscience
Neurorobotics

SpiNNaker applications

Dr Alok Joshi
Prof. Johan Storm
Dr. Ricardo Murphy
Dr Chiara Bartolozzi
Prof. Nikola Kasabov
Jan Behrenbeck
Zied Tayeb
Prof. Jorg Conradt
Prof. Wolfgang Maass
Dr Andre Grüning

SpiNNaker collaborators
**SpiNNaker2**

- **Approach: Neuromorphic Many Core System**
  - Processor based → flexibility
  - Fixed digital functionality as accelerators → performance
    - High quality random numbers (including stochastic rounding)
    - Exponential/Log functions
    - Machine Learning multiply-accumulate unit
  - Low voltage (near threshold) operation enabled by 22FDX technology and adaptive body biasing (ABB) → energy efficiency
  - Event driven operation with fine-grained dynamic power management and energy proportional chip-2-chip links → workload adaptivity

- **Scaling Target:**
  - >10 capacity compared to SpiNNaker1
  - Enabled by new hardware features and modern CMOS process

---

**SpiNNaker2 Processing Element**

- Dynamic Power Management for enhanced energy efficiency
- Memory sharing for flexible code, state and weight storage
- Multiply-Accumulate accelerator for machine learning
- Neuromorphic accelerators and random generators for synapse and neuron computation
- Network-on-Chip for efficient spike communication
- Adaptive Body Biasing for energy efficient low voltage operation

*Production ready layout in 22nm FDSOI technology*
Deep Rewiring

- Synaptic sampling as dynamic rewiring for rate-based neurons (deep networks)
- Ultra-low memory footprint even during learning
- Uses PRNG/TRNG, FPU, exp
  - \( \rightarrow \) speed-up 1.5
- Example: LeNet 300-100
  - 1080 KB \( \rightarrow \) 36 KB
  - training on local SRAM possible
  - \( \approx \) 100x energy reduction for training on SpiNNaker2 prototype (28nm) compared to X86 CPU
  - \( \rightarrow \) 96.6% MNIST accuracy for 1.3% connectivity

\( \rightarrow \) G. Bellec et al., “Deep rewiring: Training very sparse deep networks”, arXiv, 2018
\( \rightarrow \) Chen Liu et al., “Memory-efficient Deep Learning on a SpiNNaker2 prototype”, Frontiers in Neuromorphic Engineering

Reward-Based Synaptic Sampling

- Characteristics:
  - Spiking reward-based learning
  - Synaptic sampling of network configuration
- Benchmark: task-dependent routing
  - 200 input neurons, 20 stochastic neurons, 12k stochastic synapses
- Main results:
  - random, float\&exp, speed-up factor 2 of synapse update every time step
  - Use of Accelerators + local computation (no DRAM): 62% less energy
  - Modified version of synaptic rewiring “Random reallocation of synapse memory”: More efficient implementation, Faster exploration of parameter space

\( \rightarrow \) Yexin Yan et al., “Efficient Reward-Based Structural Plasticity on a SpiNNaker 2 Prototype”, IEEE Trans BioCAS

Reviewer: I rarely review papers like this that build so well on related work, that are comprehensive, and that present a significant result.
Adaptive Robotic Control with the Neural Engineering Framework

Theory:
Self-learning adaptive control algorithm realized through the Neural Engineering Framework (NEF)

Task:
Control of robotic arm
- Neural Adaptive Controller superior to PID Controller for simulated aging
- Low-latency between robot and chip required for real-time execution

Hardware Setup:
- FPGA-prototype / JIB-1 (planned) + Lego Mindstorms Ev3 + Host PC

Target:
- Demo for neuro-based processing in low-latency application
  - Evaluate use of Machine Learning Accelerator (MLA)
- > 10x speed-up from MLA

Challenges and new directions

Understanding biological neural systems
- despite accurate models, we still have little clue how, e.g., the cortex works
  - but we know it’s a lot better than any engineered system!

Learning for spiking neural nets
- currently much less established than backprop in ANNs
- promising progress from, e.g., TU Graz
  - e-Prop, BPTT, L2L, ...
- translating a trained ANN into an SNN is also possible
  - rate-based SNNs offer little advantage over ANNs?

Scale & energy efficiency
- useful networks are big, brains are very big
  - mouse ~100M neurons, 10^{12} synapses
  - human ~100B neurons, 10^{15} synapses
Conclusions

- **SpiNNaker:**
  - has been 20 years in conception…
  - …and 10 years in construction,
    - and is now ready for action!
  - ~100 boards with groups around the world
  - 1M core machine built
  - HBP is supporting s/w development

- **SpiNNaker2:**
  - 10x performance & efficiency
  - tape-out Q2 2020
    - prototype test-chips available now