The SpiNNaker project

Steve Furber
ICL Professor of Computer Engineering
The University of Manchester

1

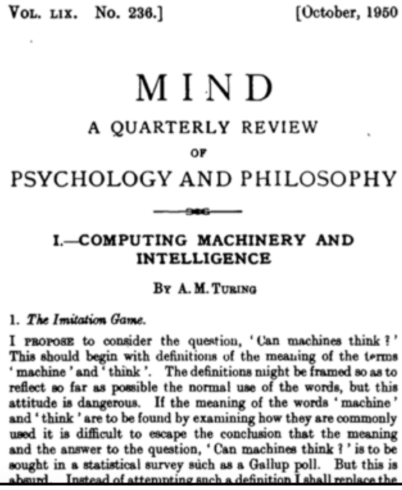
200 years ago…

- 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.”

2

70 years ago…



VOL. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think ?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think ?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the

3

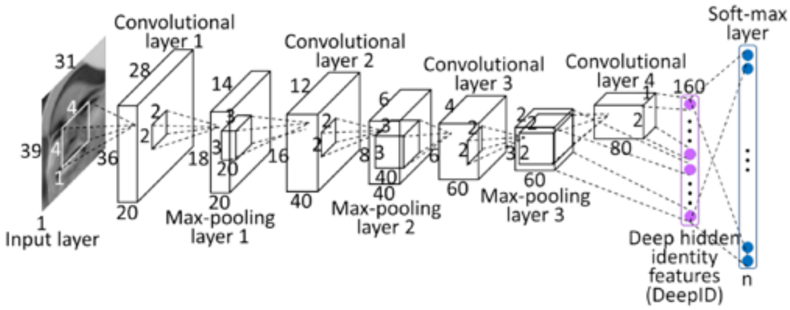
Bio-inspiration

- 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?

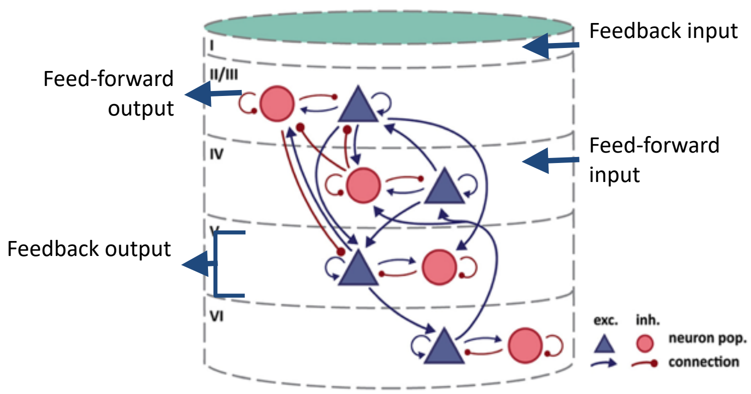
4

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

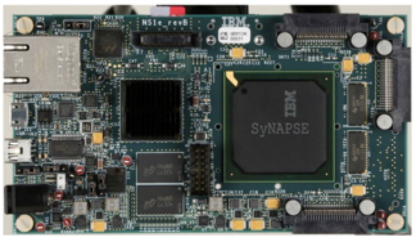
7

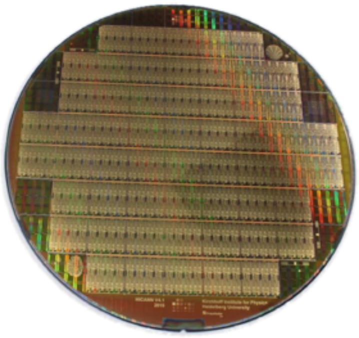
Cortical models - Supercomputers

- Sparse matrix operations
- Efficient communication of spikes
- 2.3MW

8

Cortical models - Neuromorphic hardware



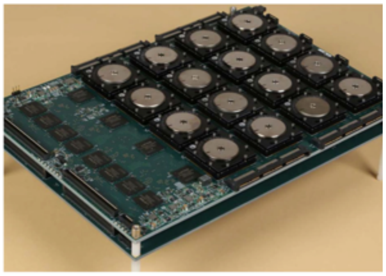


- Memory local to computation
- Low-power
- Real time
- 62mW

9

Start-ups and industry interest

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



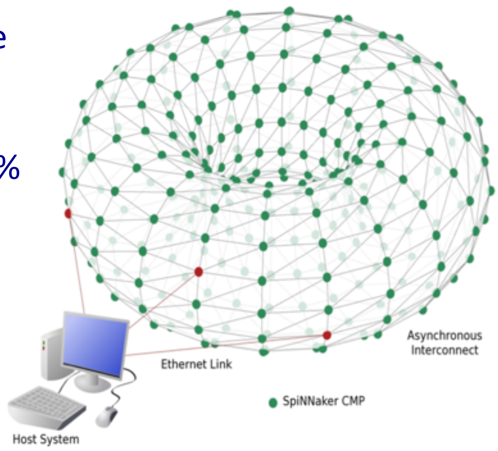
Funding UK-Startups

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

SpiNNaker

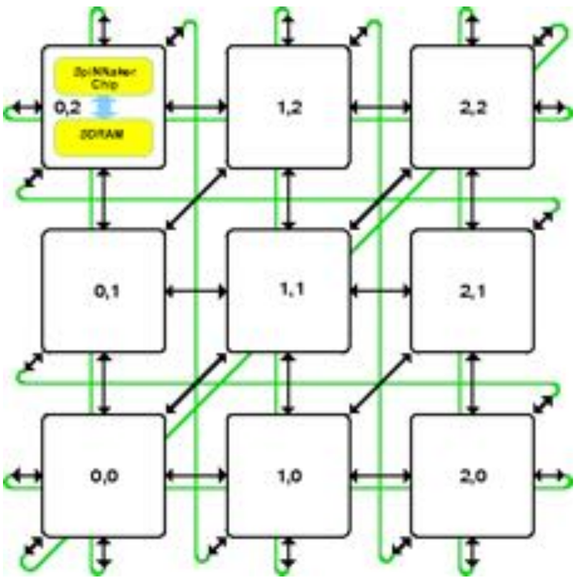
SpiNNaker project

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

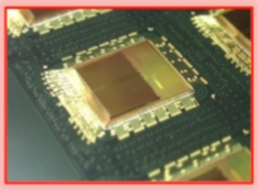


SpiNNaker

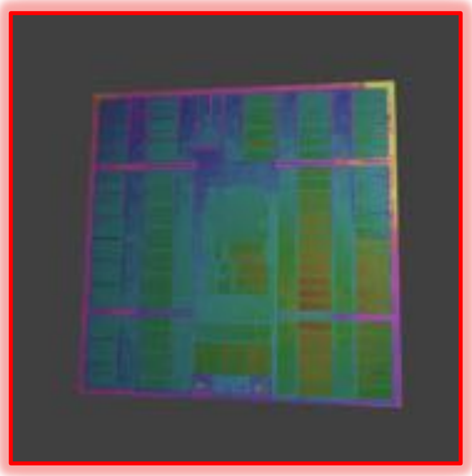
SpiNNaker system



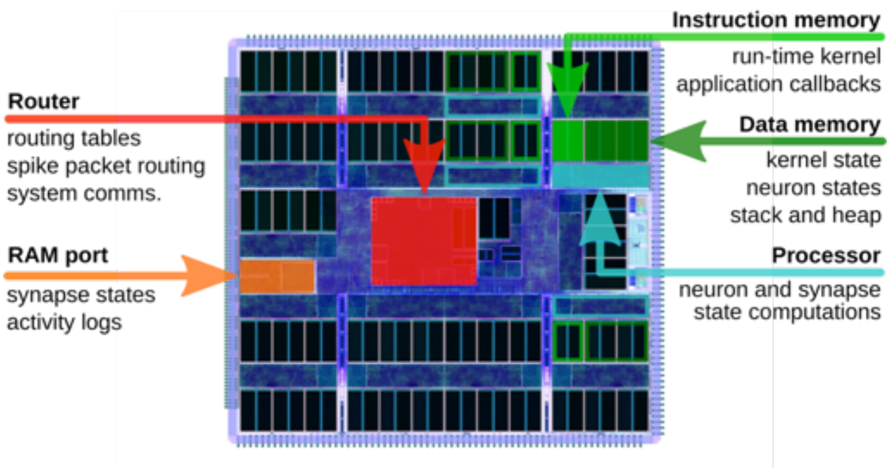
SpiNNaker chip



Multi-chip packaging by UNISEM Europe

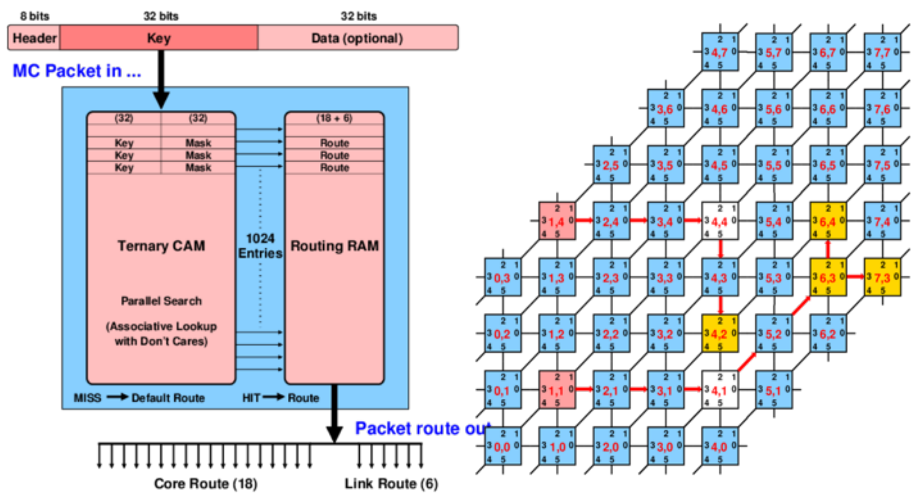


Chip resources



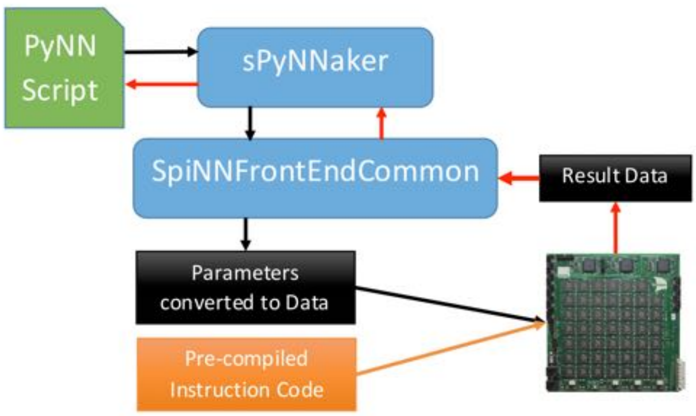
14

Multicast routing



15

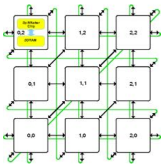
High-level software flow

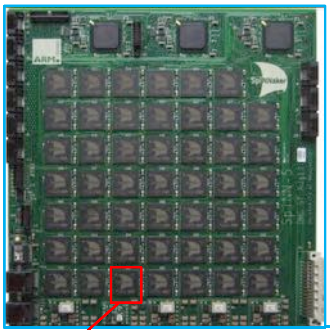


16

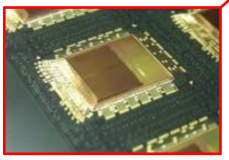
SpiNNaker machines

SpiNNaker board
(864 ARM cores)



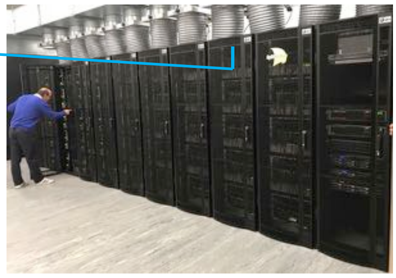


SpiNNaker chip
(18 ARM cores)



- 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

Human Brain Project

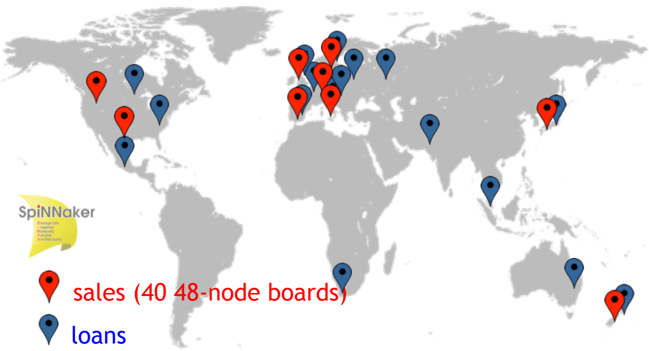


SpiNNaker racks
(1M ARM cores)

17

SpiNNaker machines

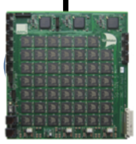
- 100 SpiNNaker systems in use
  - global coverage
- 4-node boards
  - training & small-scale robotics
- 48-node boards
  - insect-scale networks
- multi-board systems
- 1M-core HBP platform



18

SpiNNaker applications

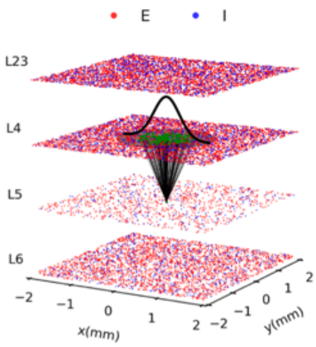
Computational Neuroscience

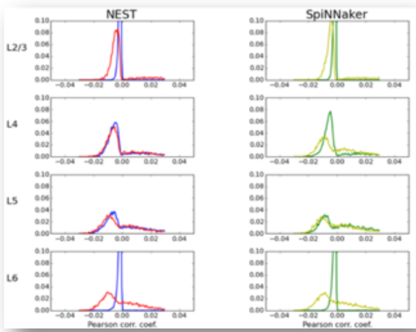


Cortical microcolumn

1st full-scale simulation of 1mm^2 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

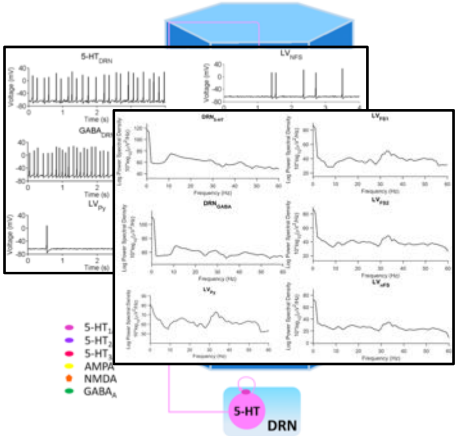




S.J. van Albada, A.G. Rowley, A. Stokes, J. Senk, M. Hopkins, M. Schmidt, D.R. Lester, M. Diesmann, S.B. Furber, *Performance comparison of the digital neuromorphic hardware SpiNNaker and the Neural network simulation software NEST for a full-scale cortical microcircuit model*. Frontiers 2018.

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



Celada, P., et al. *Serotonin modulation of cortical neurons and networks*. Frontiers in Neuroscience. 2013

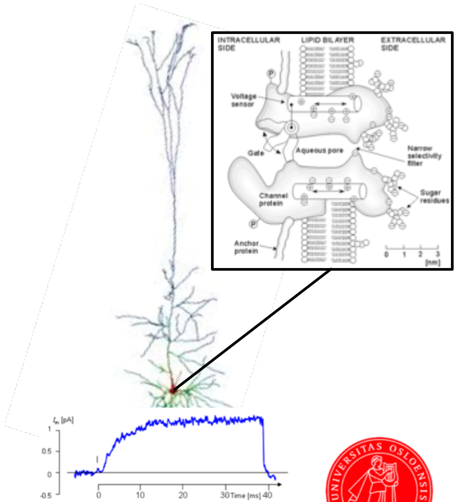
Joshi, A., & Rhodes, O,. et al. *Serotonergic modulation of cortical columnar dynamics: A large-scale neuronal network simulation study using SpiNNaker*. In prep.

Ulster University

21

Computational Neuroscience

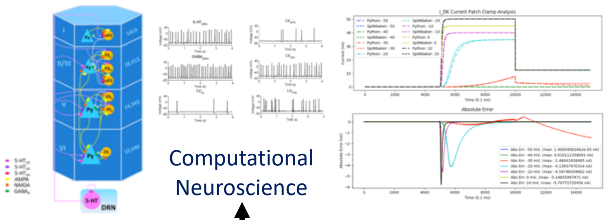
- Explore chemistry modulating neuron behaviour
  - – intracellular dynamics (ion channels)
- Simulate patch-clamp experiments from biology
- Incorporate findings at larger scales
  - – study effect on consciousness
  - – multiple brain regions

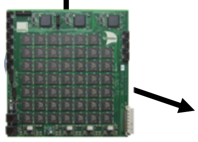


Jaakko Malmivuo and Robert Plonsey, 1995

22

SpiNNaker applications



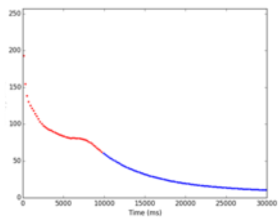


23

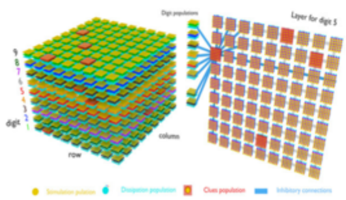
Constraint satisfaction problems

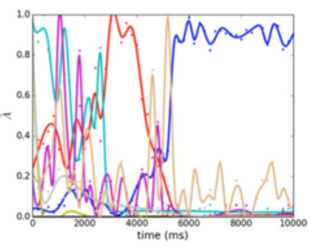
Stochastic spiking neural network:
- solves CSPs, e.g. Sudoku
  - 37k neurons
  - 86M synapses
- also
  - map colouring
  - Ising spin systems

4	6	1	9	5	8	6	7	3
3	8	5	6	7	7	4	9	1
7	8	9	3	3	1	6	8	5
6	9	6	8	1	3	5	5	7
5	3	8	7	9	2	1	4	6
1	2	4	7	6	5	8	3	9
3	5	3	1	8	6	9	9	4
8	1	6	4	4	9	3	6	2
9	4	2	5	3	6	7	1	8



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.

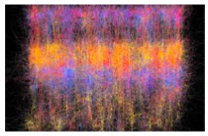
S. Habenschuss, Z. Jonke, and W. Maass, *Stochastic computations in cortical microcircuit models*, PLOS Computational Biology, 9(11):e1003311, 2013.

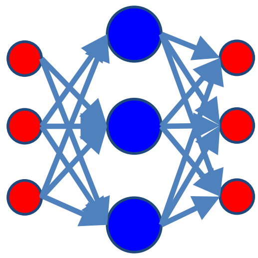
24

SpiNNaker

Theoretical Neuroscience

- 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?

Human Brain Project

Rhodes, O., et al. *How On-Chip Learning Impacts SpiNNaker Realtime Performance*. In prep.

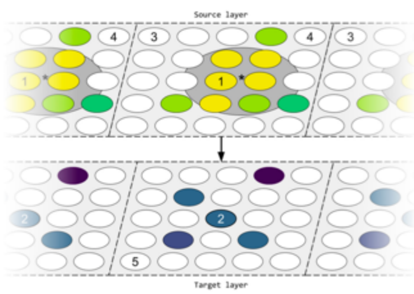
25

SpiNNaker

Theoretical Neuroscience

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



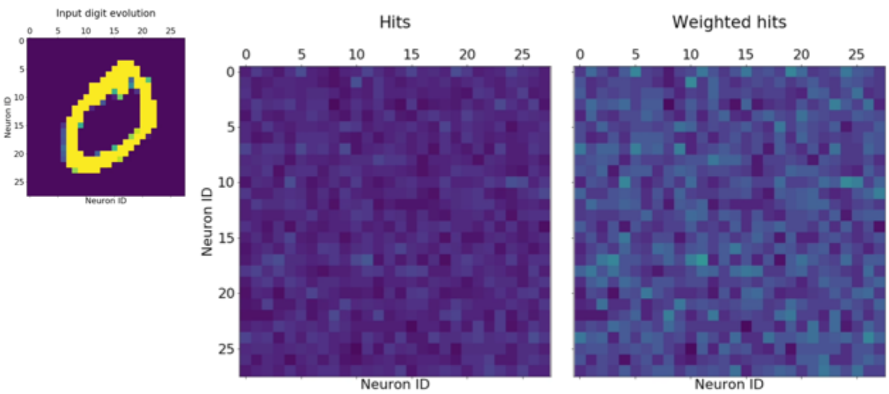
MANCHESTER 1824
The University of Manchester

Bogdan, P., et al. *Structural Plasticity on the SpiNNaker Many-Core Neuromorphic System*. Frontiers in Neuroscience. 2018

26

SpiNNaker

Theoretical Neuroscience



Hopkins, M., et al. *Spiking Neural Networks for Computer Vision*. Royal Society Interface Focus, 2018.

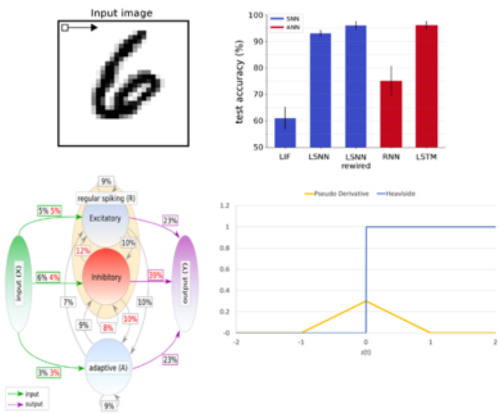
MANCHESTER 1824
The University of Manchester

27

SpiNNaker

Theoretical Neuroscience

- 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



Bellec, G., et al. *Long short-term memory and learning-to-learn in networks of spiking neurons*. NIPS 2018.

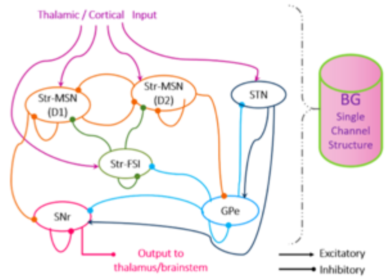
Rhodes, O., et al. *Gradient-based training of LSNNs on neuromorphic hardware*. In prep.

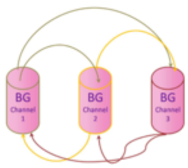
TU Graz

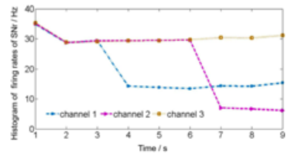
28

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





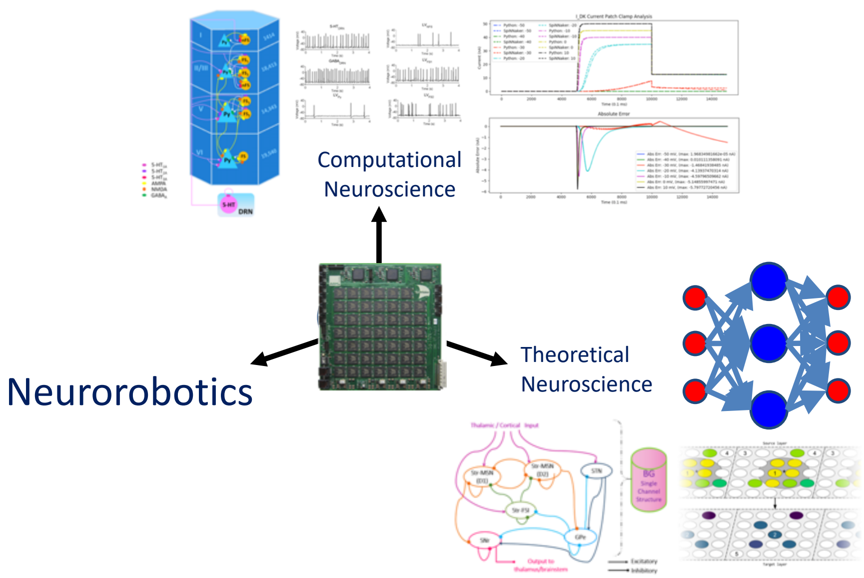


Sen-Bhattacharya, B., et al. *Building a Spiking Neural Network Model of the Basal Ganglia on SpiNNaker.* IEEE Transaction on Cognitive and Developmental Systems. 2018

MANCHESTER 1824
The University of Manchester

29

SpiNNaker applications

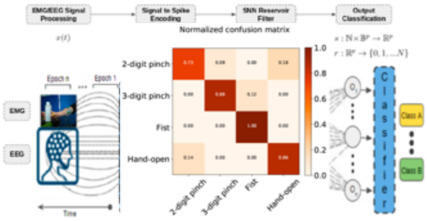


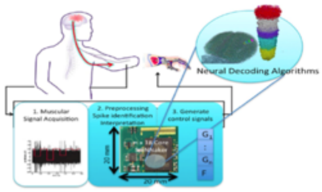
30

SpiNNaker

Neurorobotics

- 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





Behrenbeck, J. et al. *Classication and Regression of Spatio-Temporal Signals using NeuCube and its realization on SpiNNaker Neuromorphic Hardware.* Journal of Neural Engineering. 2018

AUT UNIVERSITY TUM

31

SpiNNaker

Neurorobotics

- 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

Francisco Naveros, Jesús A. Garrido, Angelo Arleo, Eduardo Ros, Niceto R. Luque. *Exploring vestibulo-ocular adaptation in a closed-loop neuro-robotic experiment using STDP. A simulation study.*

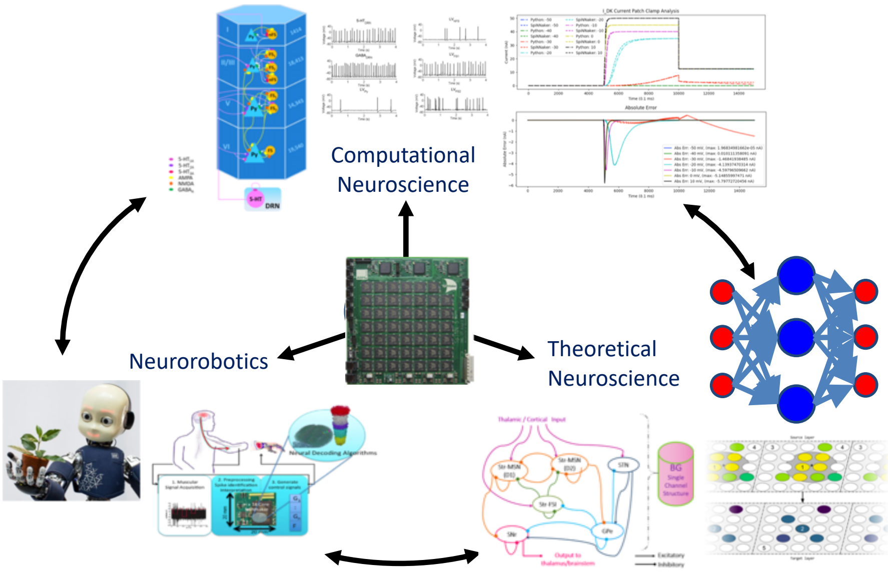
Bartolozzi, C., et al. *A Cerebellum Inspired Vestibular Occular Reflex in and iCub Robot with SpiNNaker as the Neural Substrate.* In Prep

iit ISTITUTO ITALIANO DI TECNOLOGIA

32

SpiNNaker
Biologically Inspired Massively Parallel Architectures

SpiNNaker applications



33

SpiNNaker
Biologically Inspired Massively Parallel Architectures

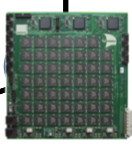
SpiNNaker collaborators

JÜLICH Forschungszentrum
Prof. Markus Diesmann
Dr. Sacha van Albada
Prof. Abigail Morrison

Ulster University
Dr Alok Joshi

Prof. Johan Storm
Dr. Ricardo Murphy

Computational
Neuroscience



ISTITUTO ITALIANO DI TECNOLOGIA
Dr Chiara Bartolozzi

Neurorobotics

Theoretical
Neuroscience

MANCHESTER 1824 The University of Manchester

AUT UNIVERSITY
Prof Nikola Kasabov

TUM
Jan Behrenbeck
Zied Tayeb
Prof. Jorg Conradt

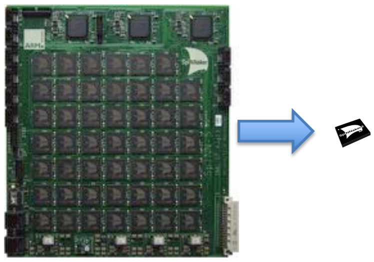
TU Graz
Prof. Wolfgang Maass

UNIVERSITY OF SURREY
Dr Andre Grüning

34

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:**
  - >x10 capacity compared to SpiNNaker1
  - Enabled by new hardware features and modern CMOS process



35

SpiNNaker2 Processing Element

Dynamic Power Management for enhanced energy efficiency

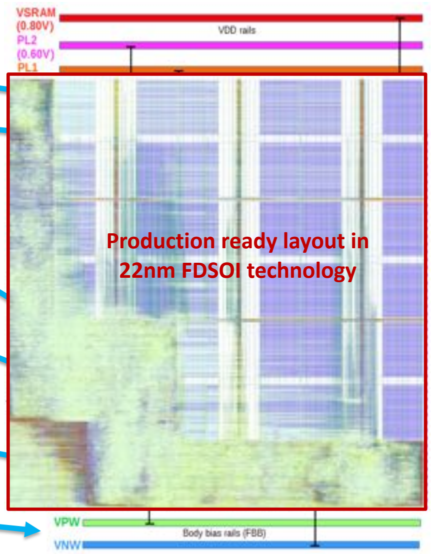
Memory sharing for flexible code, state and weight storage

Multiply-Accumulate accelerator for machine learning

Neuromorphic accelerators and random generators for synapse and neuron computation

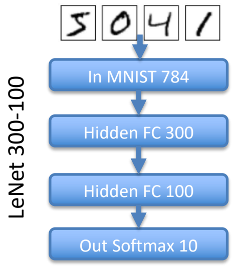
Network-on-Chip for efficient spike communication

Adaptive Body Biasing for energy efficient low voltage operation

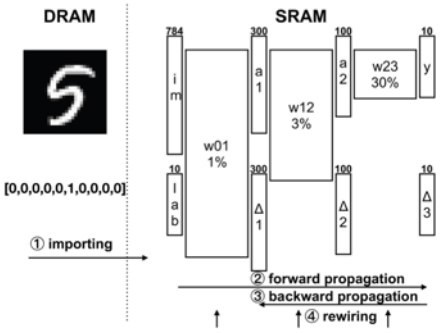


36

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
  - → speed-up 1.5
- Example: **LeNet 300-100**
  - 1080 KB → 36 KB
  - training on local SRAM possible
  - ≈ 100x energy reduction for training on SpiNNaker2 prototype (28nm) compared to X86 CPU
  - **→ 96.6% MNIST accuracy for 1.3% connectivity**

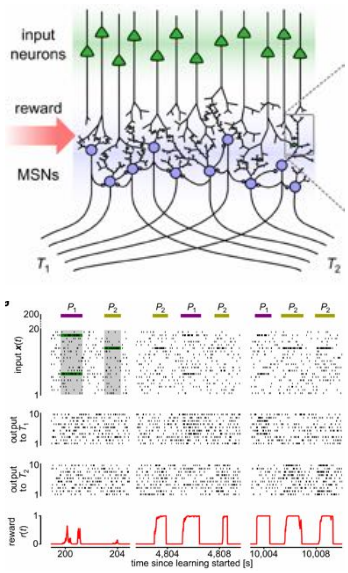


→ *G. Bellec et al., "Deep rewiring: Training very sparse deep networks", arXiv, 2018*
→ *Chen Liu et al., "Memory-efficient Deep Learning on a SpiNNaker 2 prototype", Frontiers in Neuromorphic Engineering*

37

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



→ *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.

38

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

39

Challenges and new directions

Understanding biological neural systems

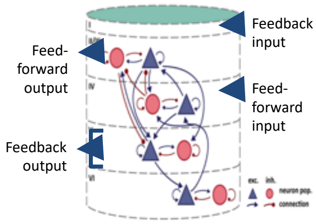
- despite accurate models, we still have little clue how, e.g., the cortex works
  - but we know it's a lot better than any engineered system!

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

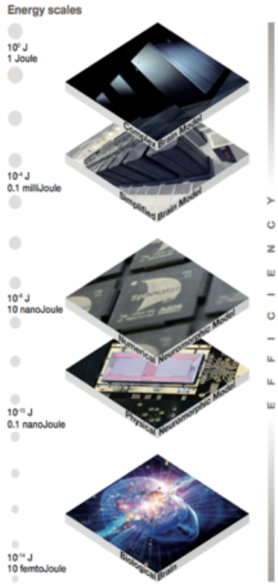


40

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Conclusions

Human Brain Project



- ***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

41