

Neuronal Modelling & Large-Scale Neuromorphic Systems (SpiNNaker)

Steve Furber CBE FRS FREng

Professor Emeritus
The University of Manchester



The University of Manchester



EPSRC **ARM**

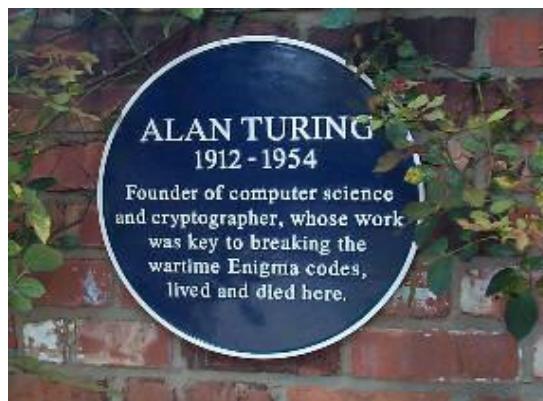
200 years ago...

- 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...



VOL. LIX. No. 236.]

[October, 1950



M I N D
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

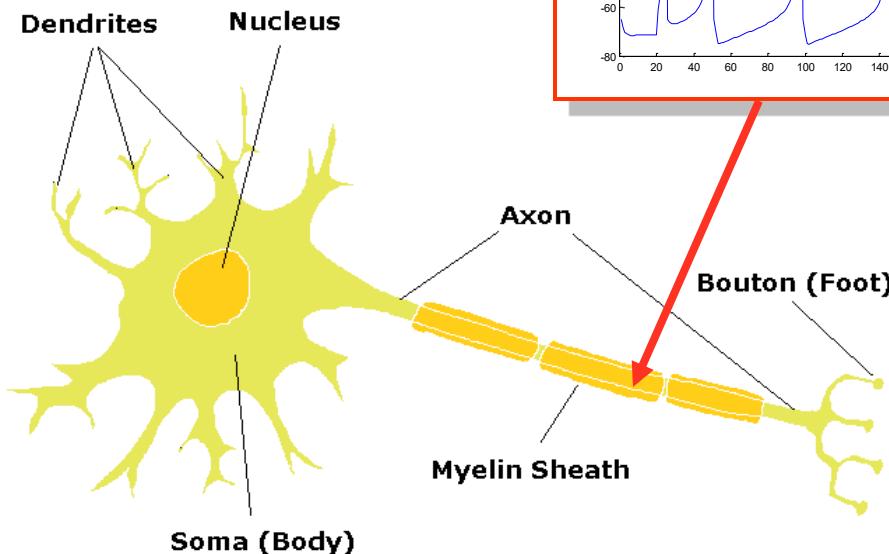
Outline



- Brains
- Neuromorphic computing
- The SpINNaker project
- SpINNaker applications
- SpINNaker2
- SpINNaker2 applications
- Conclusions

Neurons

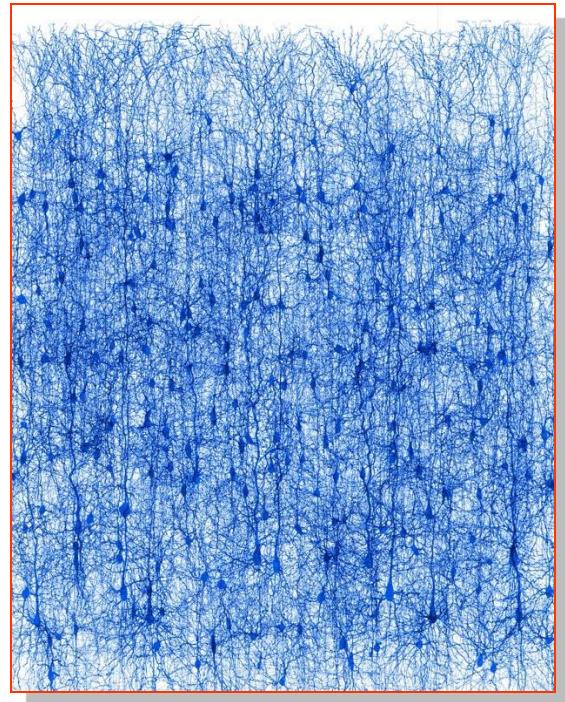
- Multiple inputs (dendrites)
- Single output (axon)
 - digital “spike”
 - fires at 10s to 100s of Hz
 - output connects to many targets
- Synapse at input/output connection



(www.ship.edu/~cgoeree/theneuron.html)

Neurons

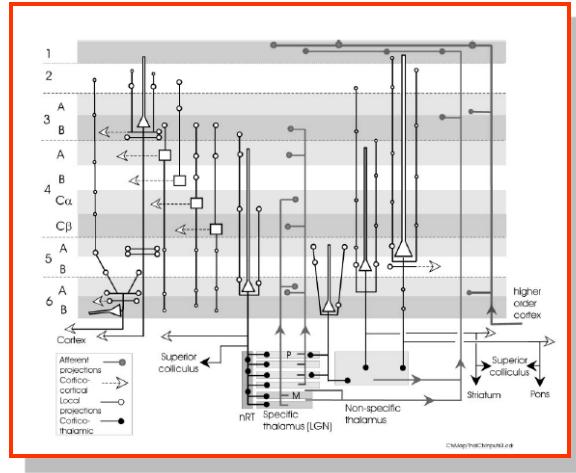
- A flexible biological control component
 - very simple animals have a handful
 - bees: 850,000
 - humans: 10^{11}



(photo courtesy of the Brain Mind Institute, EPFL)

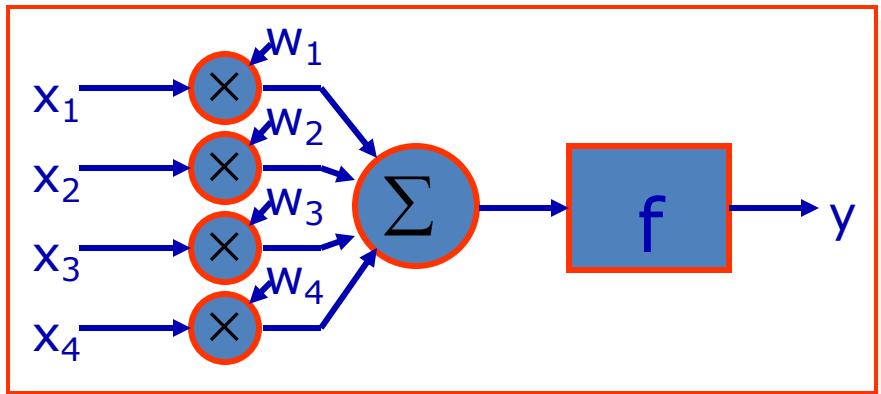
Neurons

- Regular high-level structure
 - e.g. 6-level cortical microarchitecture
 - low-level vision, to
 - language, etc.
- Random low-level structure
 - adapts over time



Neural computation

- To compute we need:
 - *Processing*
 - *Communication*
 - *Storage*
- Processing:
abstract model
 - linear sum of weighted inputs
 - ignores non-linear processes in dendrites
 - non-linear output function
 - learn by adjusting synaptic weights



Processing

- Leaky integrate-and-fire model
 - inputs are a series of spikes
 - total input is a weighted sum of the spikes
 - neuron activation is the input with a “leaky” decay
 - when activation exceeds threshold, output fires
 - habituation, refractory period, ...?

$$x_i = \sum_k \delta(t - t_{ik})$$

$$I = \sum_i w_i x_i$$

$$\dot{A} = -A / \tau_A + I$$

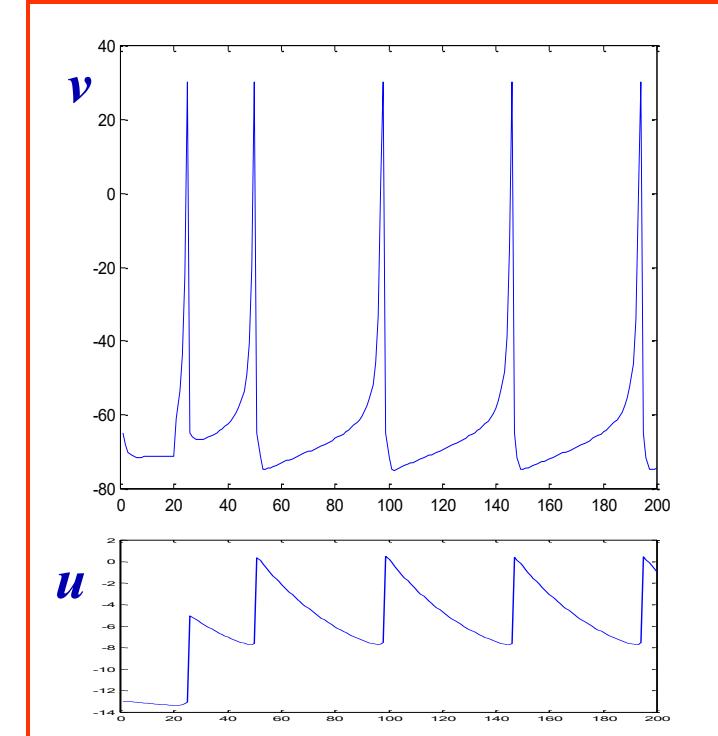
if $A > \vartheta_A$ fire

& set $A = 0$

Processing

- Izhikevich model
 - two variables, one fast, one slow:
$$\dot{v} = 0.04v^2 + 5v + 140 - u + I$$

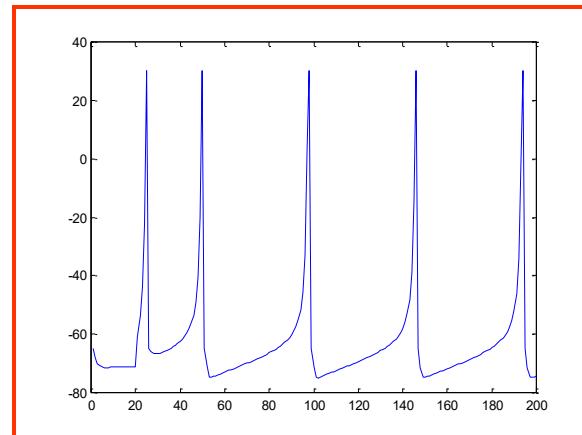
$$\dot{u} = a \cdot (bv - u)$$
- neuron fires when $v > 30$; then:
 - $v = c$
 - $u = u + d$
- a, b, c & d select behaviour



(www.izhikevich.com)

Neural communication

- Spikes
 - biological neurons communicate principally via ‘spike’ events
 - asynchronous
 - information is only:
 - which neuron fires, and
 - when it fires
 - ‘Address Event’ Representation (AER)



Neural storage

- Synaptic weights
 - stable over long periods of time
 - with diverse decay properties?
 - adaptive, with diverse rules
 - Hebbian, anti-Hebbian, LTP, LTD, ...
- Axon ‘delay lines’
- Neuron dynamics
 - multiple time constants
- Dynamic network states
- Structural plasticity
- Neurogenesis
- ...?

Outline



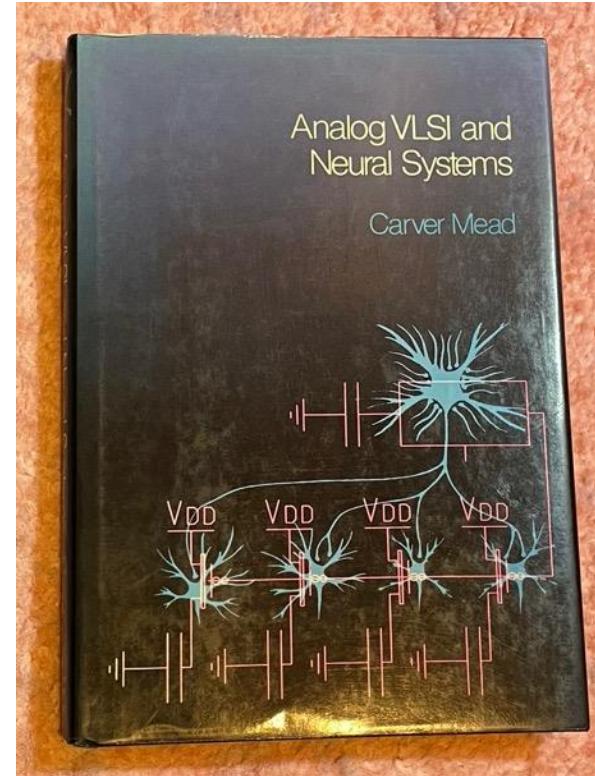
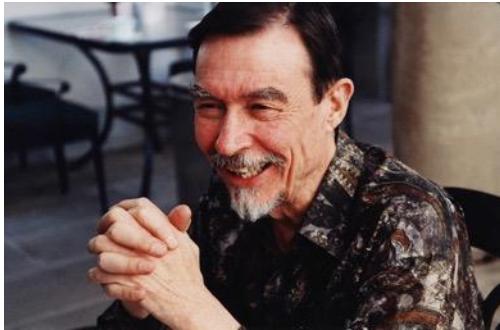
- Brains
- Neuromorphic computing
- The SpINNaker project
- SpINNaker applications
- SpINNaker2
- SpINNaker2 applications
- Conclusions

Neuromorphic Computing

Carver Mead

Caltech, 1980s

- observed analogy between ion channels in neurons and sub-threshold analogue transistor behaviour
- neuromorphic touch, hearing & vision sensors



The Human Brain Project

Why focus on the brain ? Three Reasons

– Understanding the brain (Unifying Science Goal)

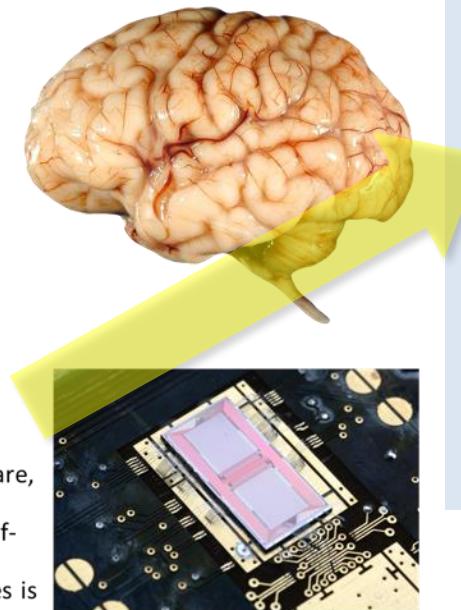
- Underpins what we are,
- Data & knowledge are fragmented,
- Integration is needed,
- Large scale collaborative approach is essential.

– Understanding brain diseases (Society)

- Costs Europe over €800 Billion/year,
- Affects 1/3 people,
- Number one cause of loss of economic productivity,
- No fundamental treatments exist or are in sight
- Pharma companies pulling out of the challenge.

– Developing Future Computing (Technology)

- Computing underpins modern economies,
- Traditional computing faces growing hardware, software, & energy barriers,
- Brain can be the source of energy efficient, robust, self-adapting & compact computing technologies,
- Knowledge driven process to derive these technologies is missing.



Neuromorphic Computing

- Neuromorphic Machines
- Algorithms and Architectures for Neuromorphic Computing
 - Theory
 - Applications



Co-funded by
the European Union

The HBP Neuromorphic Computing Strategy

Next generation of NMC is more biology driven

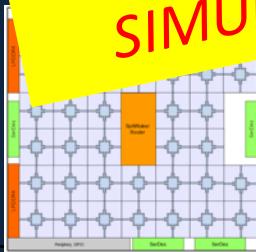
1st generation SpiNNaker-1 Machine



Many-core
Architecture
SIMULATION

Many-core system
on ARM cores
time simulator

Towards 2nd gen



152 Cortex M4F per chip
36 GIPS/Watt per chip
x10 with constant power

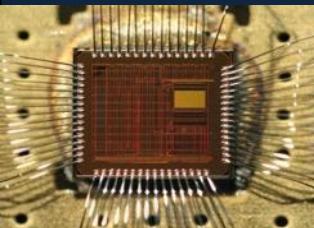
1st generation BrainScaleS-1 Machine



Physical mode
EMULATION

Physical model system
4M neurons, 1B plastic syn.
accelerated emulator

Towards 2nd gen

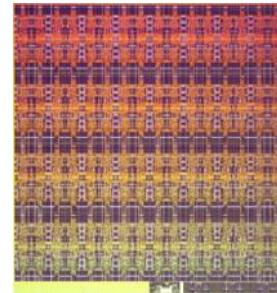


On-chip plasticity processors
Flexible hybrid plasticity
Active dendrites

Designed and built from the transistor up !

Co-designed with (theoretical) neuroscience

Neuromorphic systems worldwide



Biological realism

Ease of use

Many-core (ARM) architecture
Optimized spike
communication network
Programmable local learning
x0.01 real-time to x10 real-time

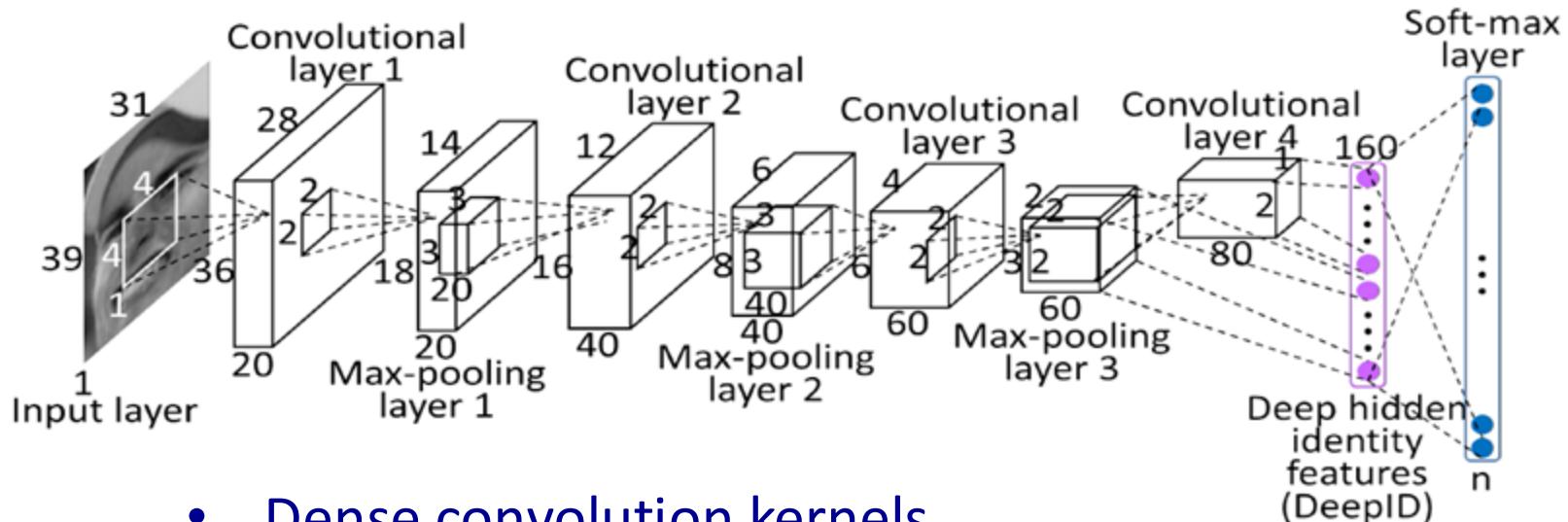
Full-custom-digital neural circuits
No local learning (TrueNorth)
Programmable local learning (Loihi)
Exploit economy of scale
x0.01 real-time to x100 real-time

Analog neural cores
Digital spike communication
Biological local learning
Programmable local learning
x10.000 to x1000 real-time

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?

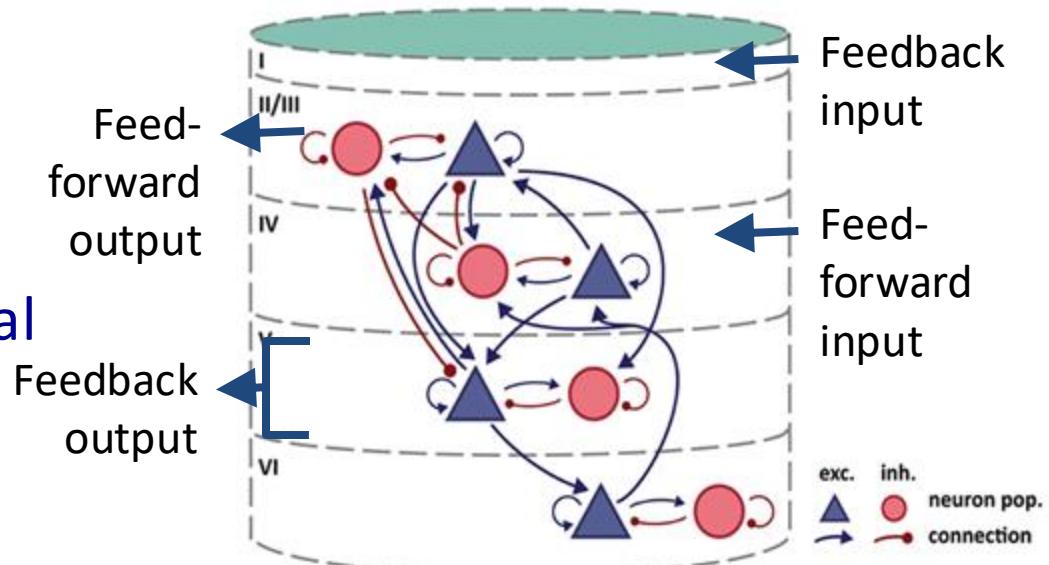
ConvNets - structure



- Dense convolution kernels
- Abstract neurons (no spikes)
- Feed-forward connections (mainly)
- Trained through backpropagation

The cortex - structure

- Spiking neurons
- Complex information flow
- Two-dimensional cortical structure
- Sparse connectivity
 - < 10%



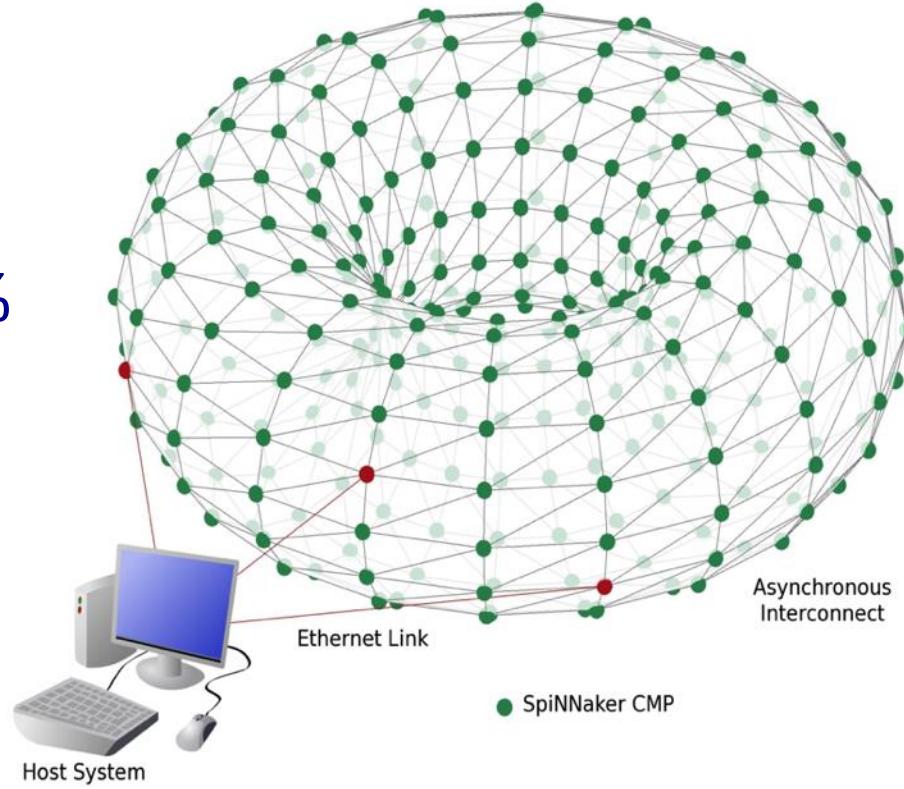
Outline



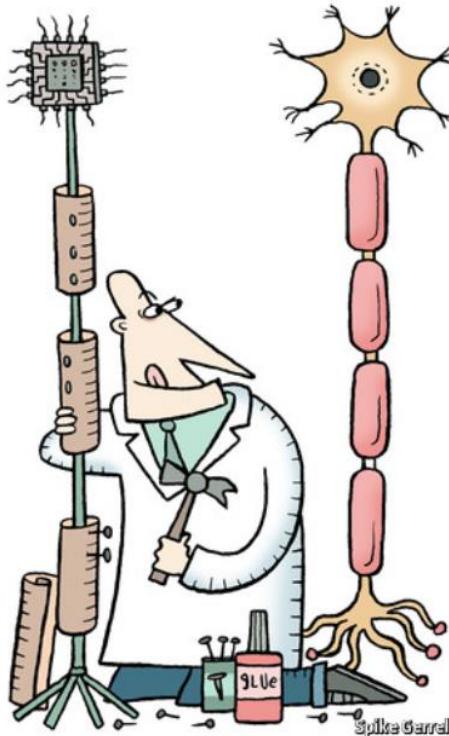
- Brains
- Neuromorphic computing
 - The SpINNaker project
 - SpINNaker applications
 - SpINNaker2
 - SpINNaker2 applications
 - Conclusions

SpiNNaker project

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

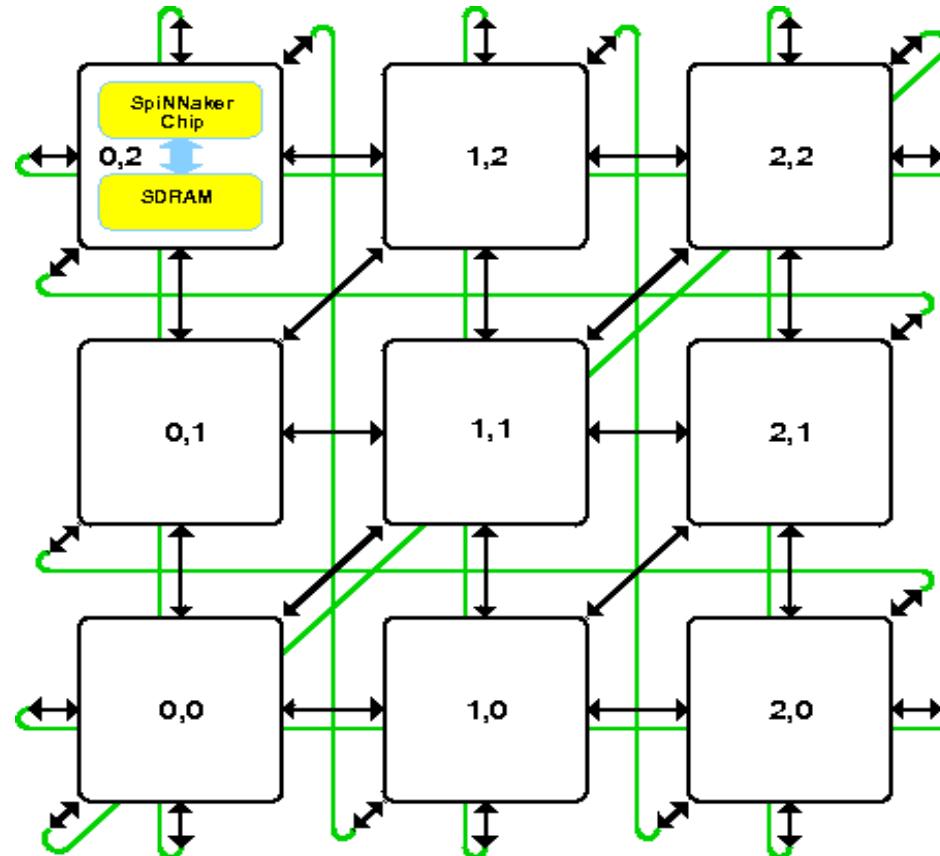


Design principles

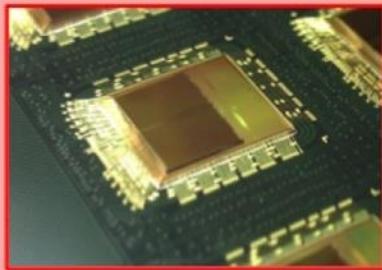


- *Virtualised topology*
 - physical and logical connectivity are decoupled
- *Bounded asynchrony*
 - time models itself
- *Energy frugality*
 - processors are free
 - the real cost of computation is energy

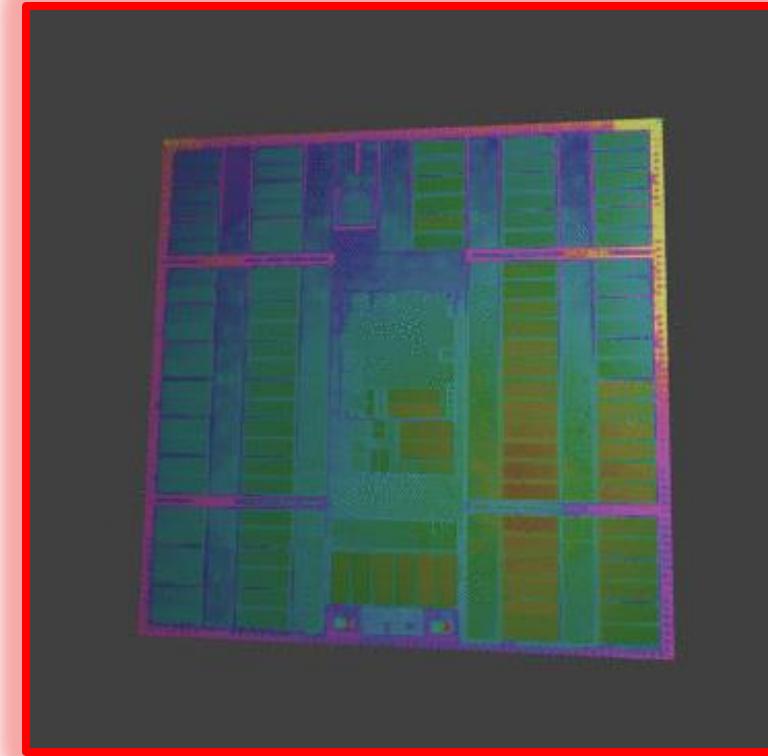
SpINNaker system



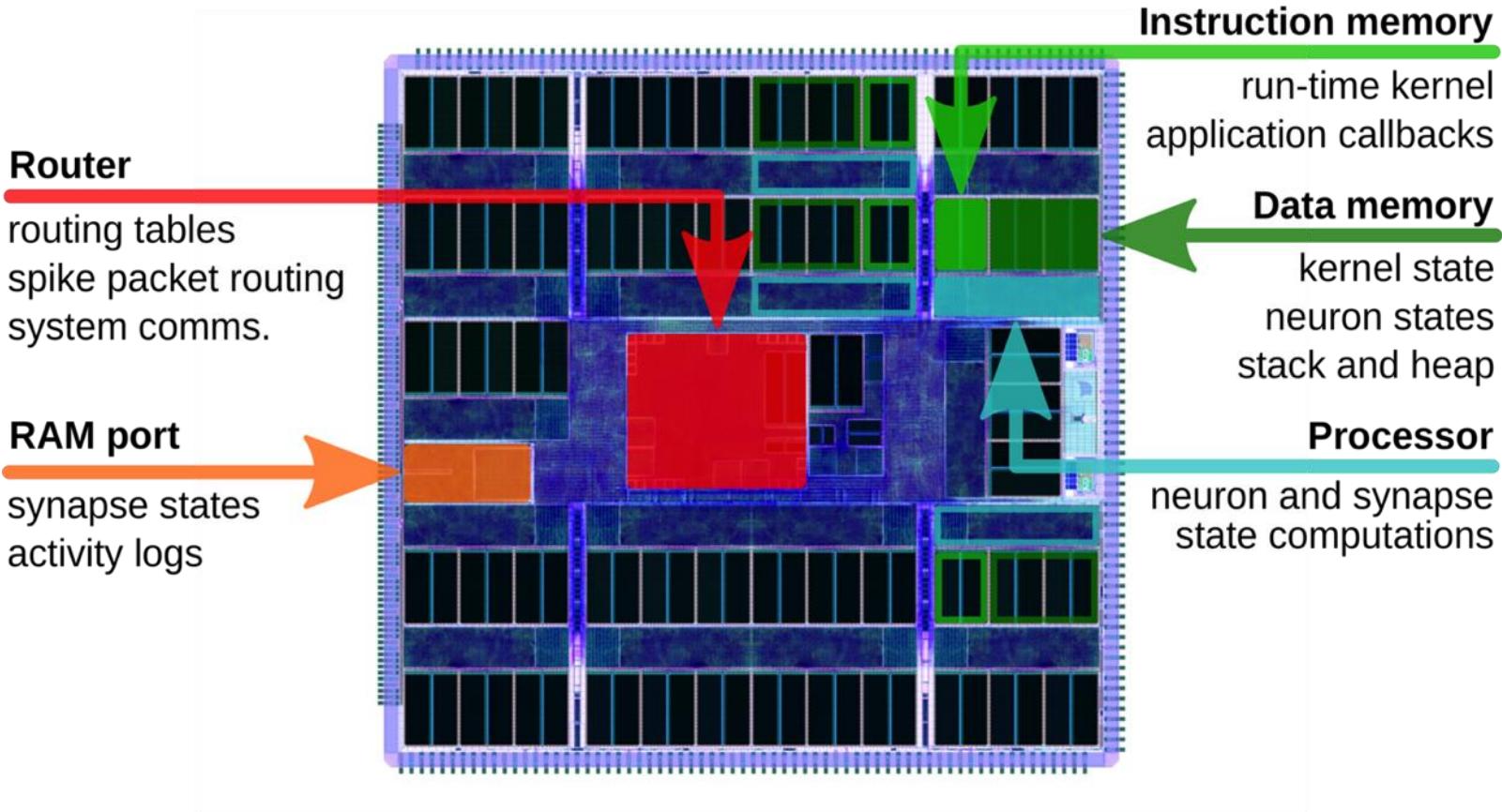
SpINNaker chip



Multi-chip
packaging by
UNISEM

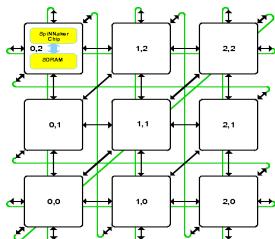


Chip resources

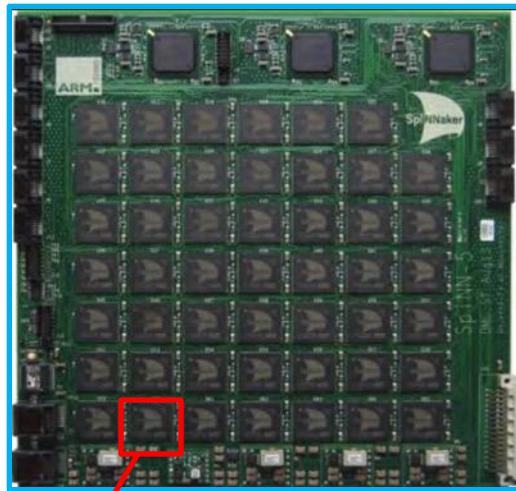
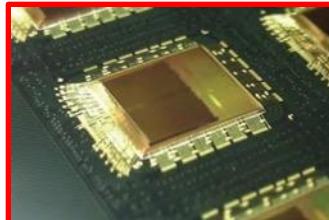


SpiNNaker machines

SpiNNaker board
(864 ARM cores)



SpiNNaker chip
(18 ARM cores)



SpiNNaker racks
(1M ARM cores)

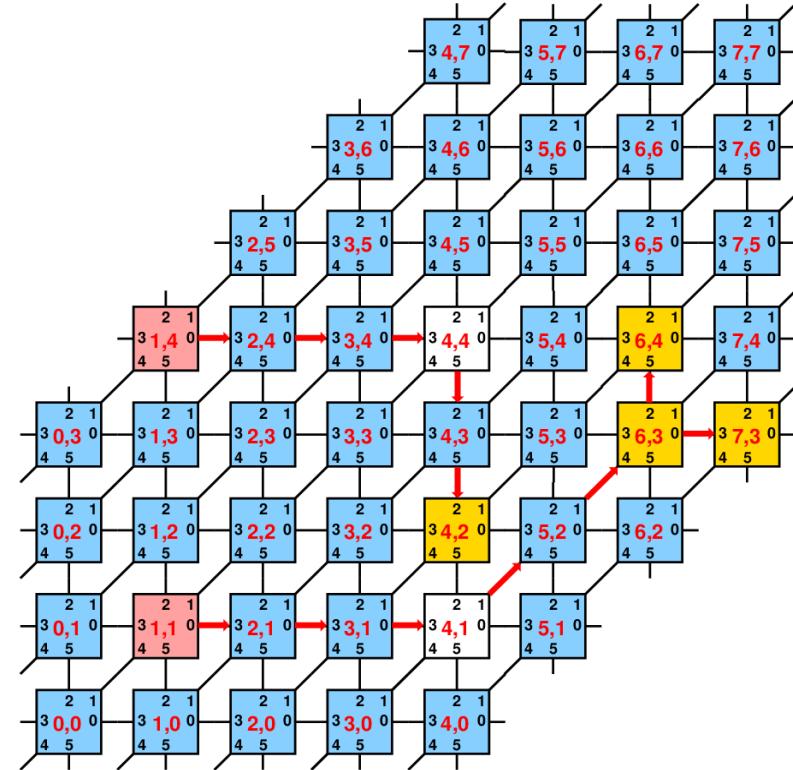
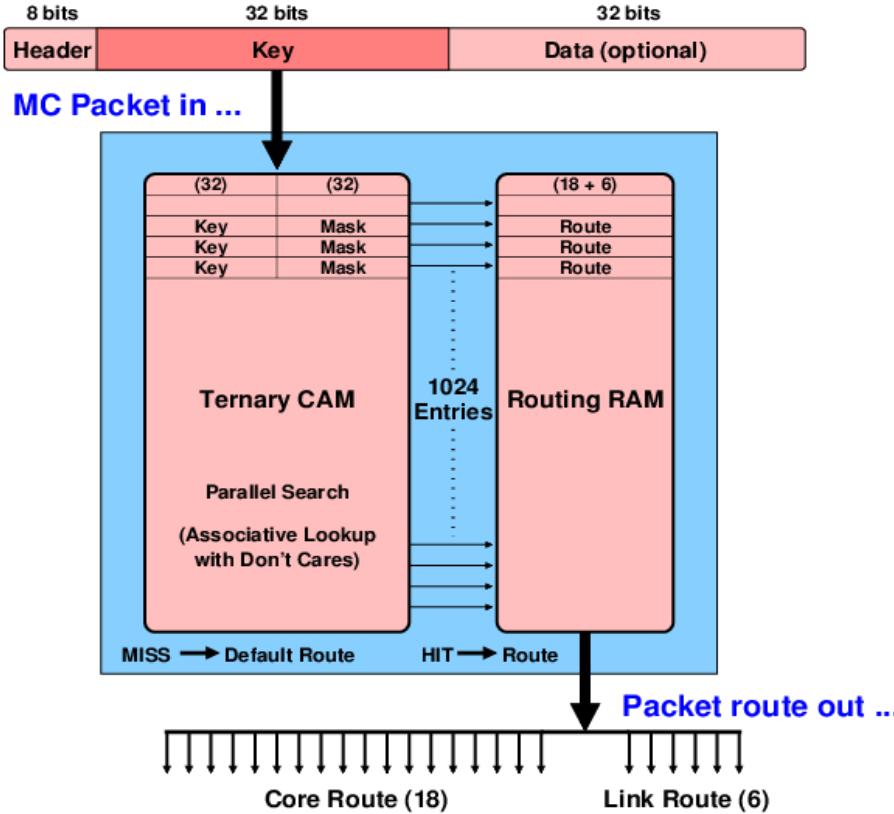
- HBP platform
 - 1M cores
 - 11 cabinets (including server)
- Launch 30 March 2016
 - then 500k cores
 - ~450 remote users
 - 5M SpiNNaker jobs run



Human Brain Project



Multicast routing



Building the HBP machine



The University of Manchester

Building and wiring up the
518,400 core SpiNNaker machine

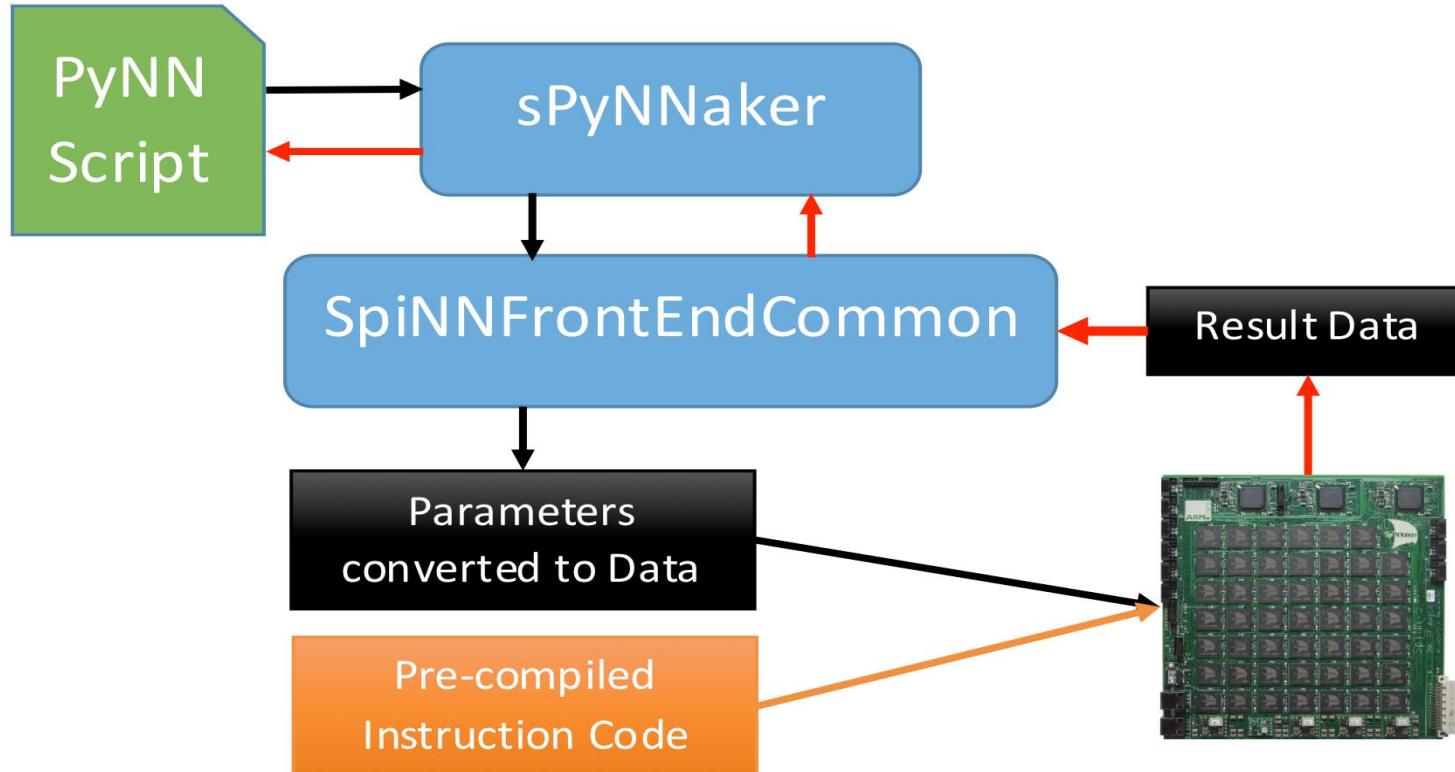
.....



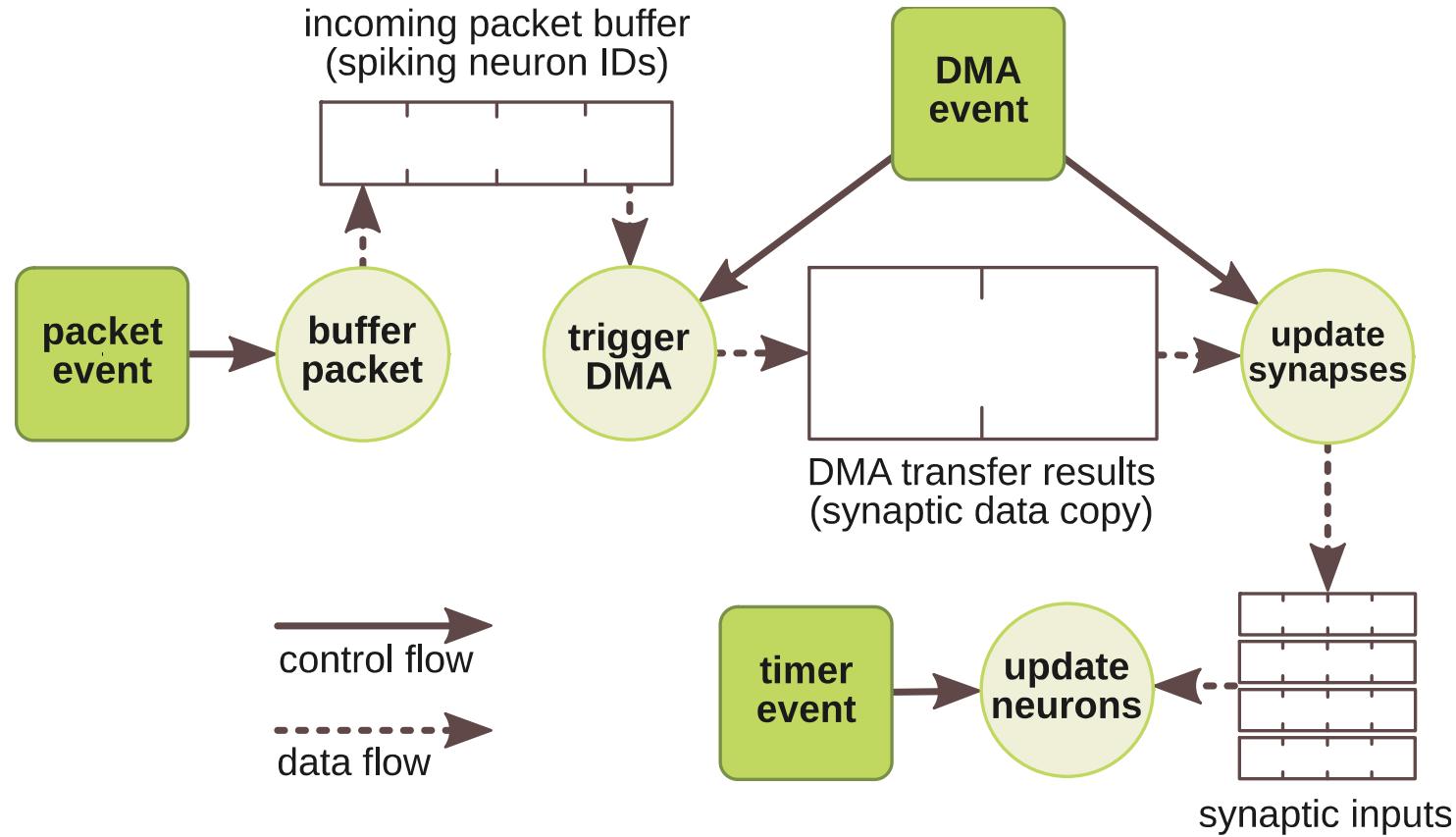
Human Brain Project



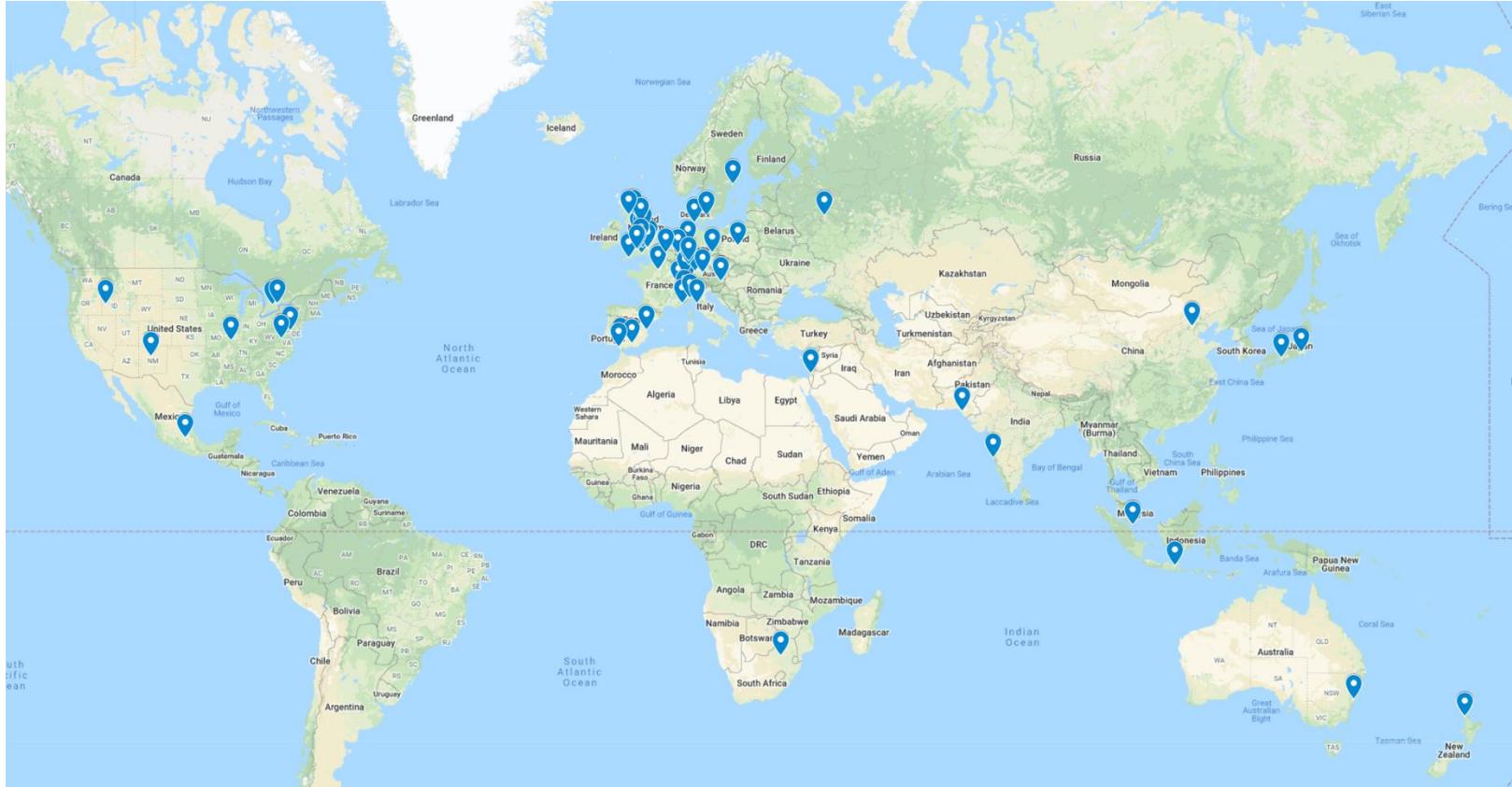
High-level software flow



Event-driven software model



SpINNaker machines

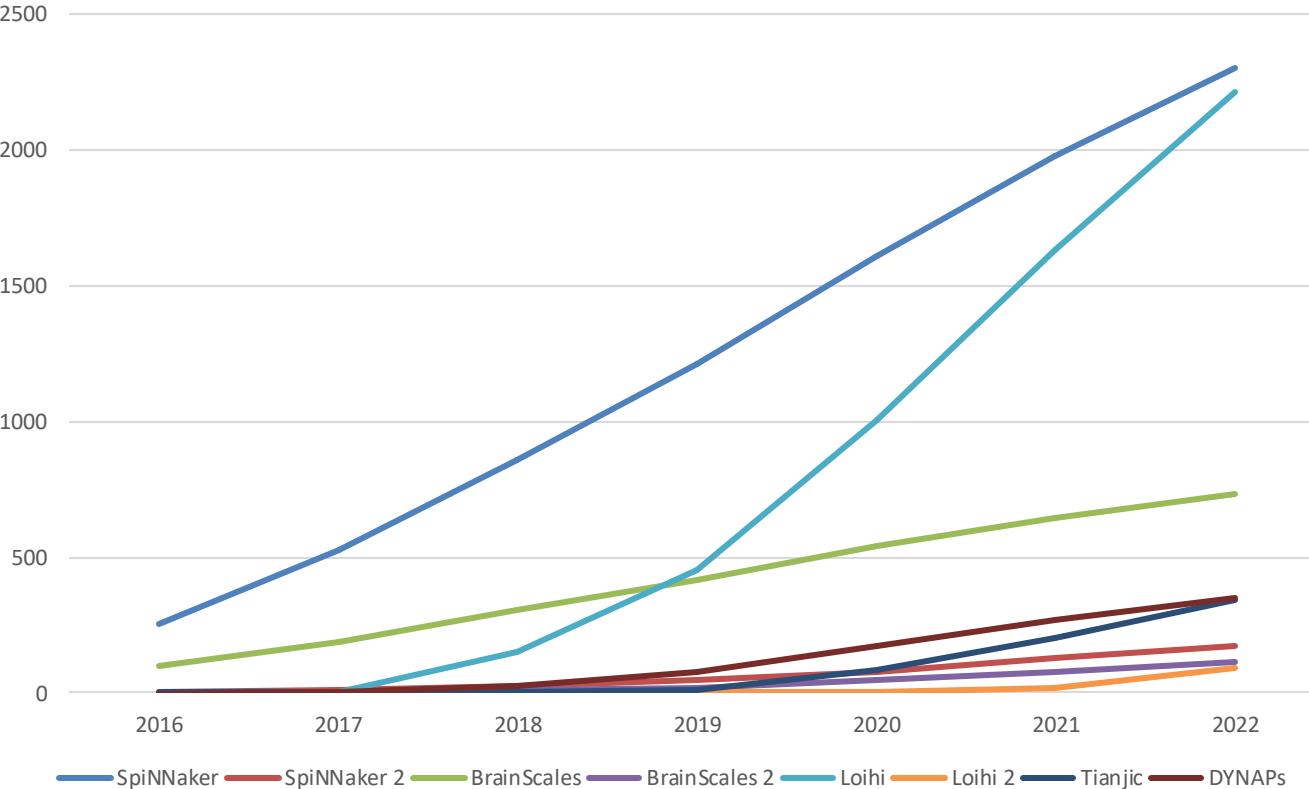


Publications mentioning neuromorphic chips

Mentioned papers per year (cumulative)

SpiNNaker GS citations:

- 1,033: *The SpiNNaker Project*, Proc. IEEE, 2014.
- 664: *Overview of the SpiNNaker system architecture*, IEEE Trans. Computers, 2012.
- 367: *SpiNNaker: A 1w 18-core SoC for...*, IEEE JSSC, 2013.
- 356: *SpiNNaker: mapping neural networks...*, IJCNN 2008.
- 342: *Large-scale neuromorphic computing systems*, J. Neural Engineering, 2016.



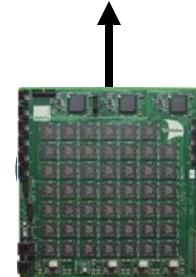
Outline



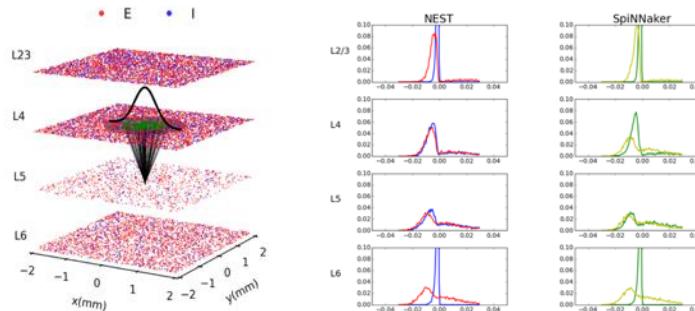
- Brains
- Neuromorphic computing
- The SpINNaker project
- SpINNaker applications
- SpINNaker2
- SpINNaker2 applications
- Conclusions

SpiNNaker applications

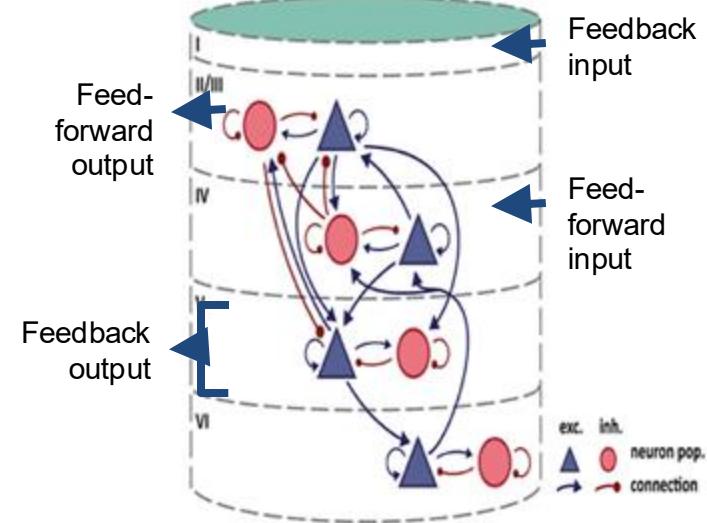
Computational
Neuroscience



Cortical microcircuit



- **Realtime execution of cortical model**
 - 1mm² cortex
 - 77k neurons
 - 285M synapses
 - 0.1 ms time-step
- **Best previous versions of this model**
 - HPC: 3x slow-down
 - GPU: 2x slow-down
- **Will scale to 100mm² without slow-down**
 - on current machine, simply by using more boards

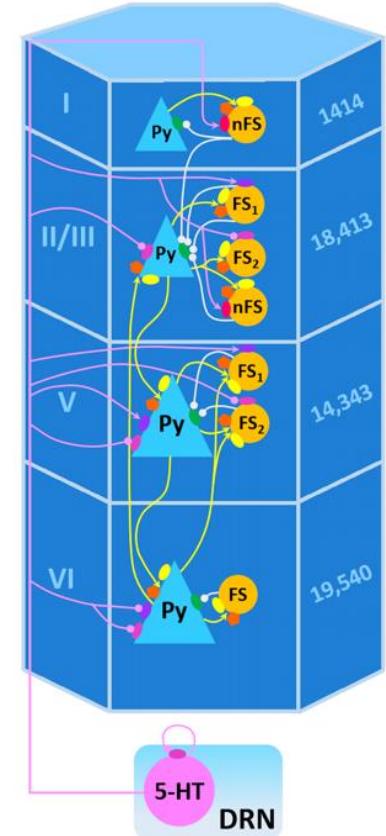
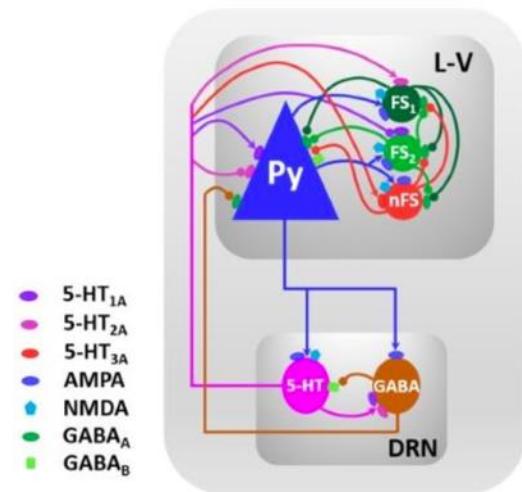


S.J. van Albada, A.G. Rowley, A. Stokes, J. Senk, M. Hopkins, M. Schmidt, D.R. Lester, M. Diesmann, S.B. Furber, "Performance comparison of the digital neuromorphic hardware SpiNNaker and the Neural network simulation software NEST for a full-scale cortical microcircuit model", Frontiers 2018.

Oliver Rhodes, Luca Peres, Andrew G. Rowley, Andrew Gait, Luis A. Plana, Christian Brenninkmeijer & Steve.B. Furber, "Real-time cortical simulation on neuromorphic hardware", Phil Trans Roy Soc A, December 2019.

Pre-Frontal Cortex, Dorsal Raphe Nucleus

- 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

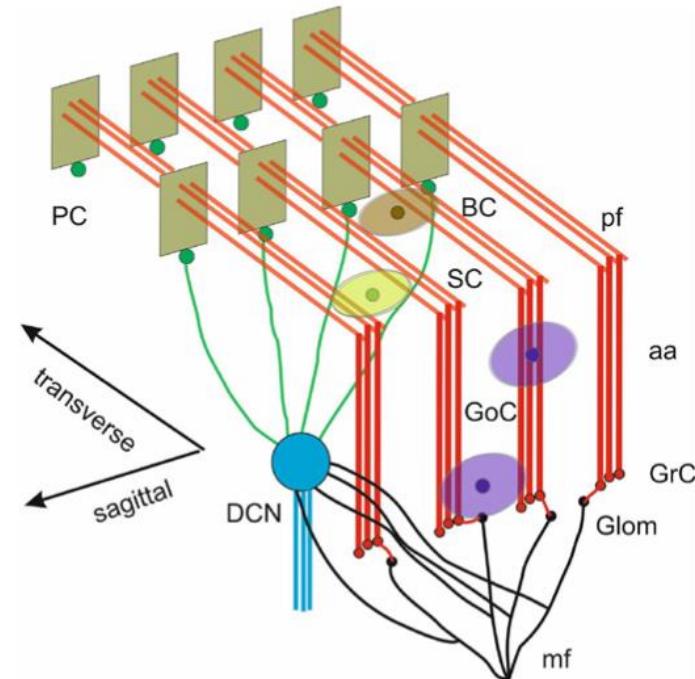


Joshi, A., Rhodes, O., et al. *Large-scale network modelling of interactions between pre-frontal cortex and dorsal raphe nucleus using SpiNNaker*. 2017.

Joshi, A., et al. *Long- and short-range connectivity and neuronal types affect prefrontal dorsal raphe circuit dynamics differently*. Neuromatch 2020.

Towards a bio-inspired real-time neuromorphic cerebellum

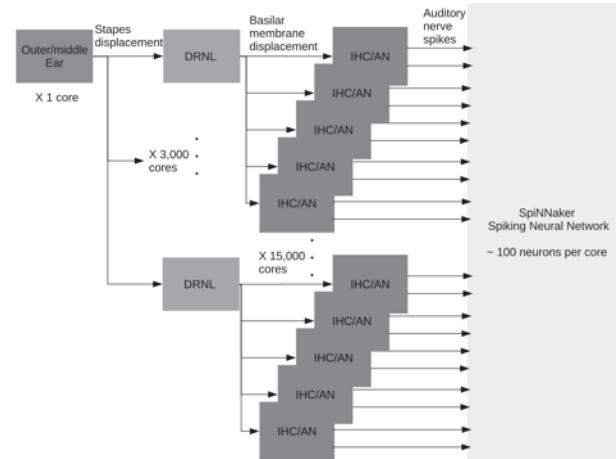
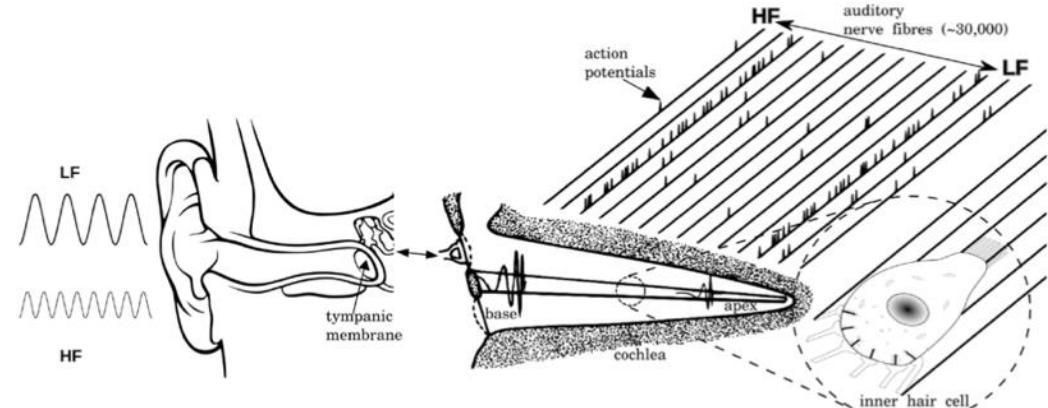
- HBP collaboration with U. Pavia
- 97,000 neurons
- 4.2 million synapses
- SpiNNaker model validated against NEST



mf	mossy fibres
Glom	Glomerulus
GrC	Granule cell
GoC	Golgi cell
SC	Stellate cell
BC	Basket cell
PC	Purkinje cell
DCN	Deep Cerebellar Nucleus
aa	ascending axons
pf	parallel fibres

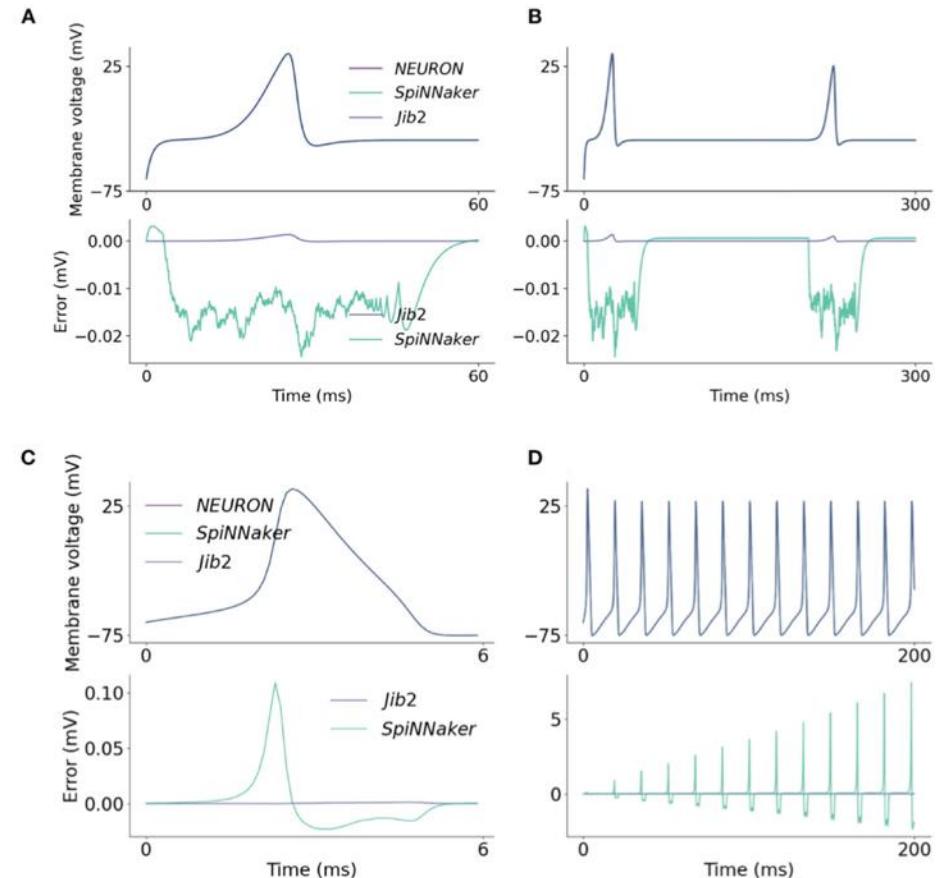
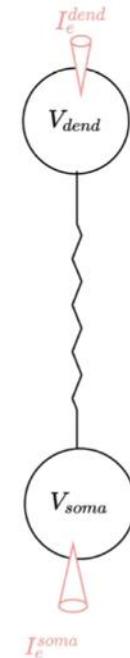
Inner Hair Cell and Auditory Nerve model

- Matlab Auditory Pathway (MAP) re-implemented on SpiNNaker
- 18,000 SpiNNaker cores
 - many running DSP code
 - lack of hardware floating-point on SpiNNaker1 hurts here!
- 1,500 SpiNNaker chips

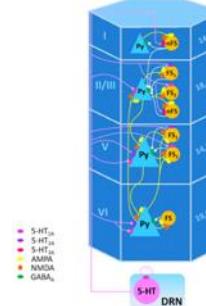


Beyond LIF models on SpiNNaker1 & 2

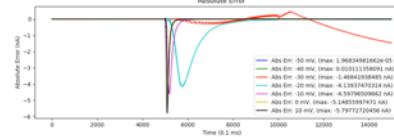
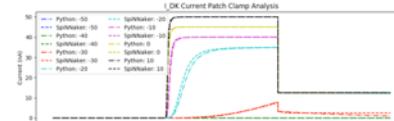
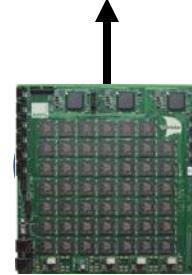
- single compartment Hodgkin-Huxley neuron
- multi-compartment neuron incorporating dendritic computation



SpINNaker applications



Computational
Neuroscience

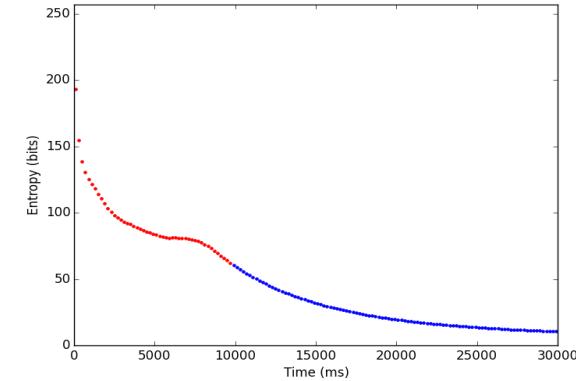
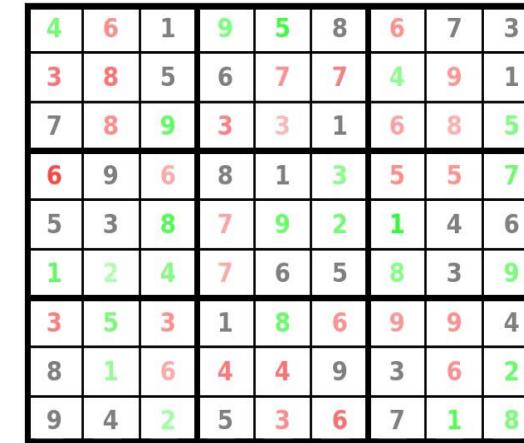


Theoretical
Neuroscience

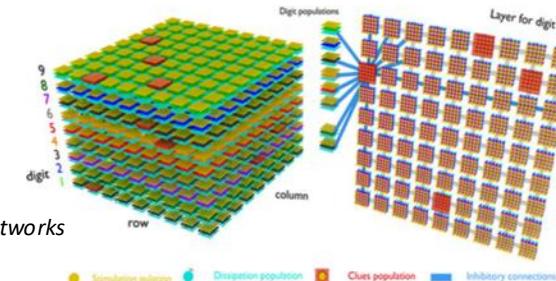
Constraint satisfaction problems

Stochastic spiking neural network:

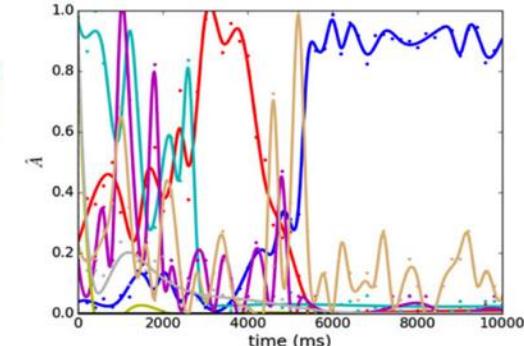
- Solves CSPs, e.g. Sudoku
 - 37k neurons
 - 86M synapses
- also
 - map colouring
 - Ising spin systems



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.

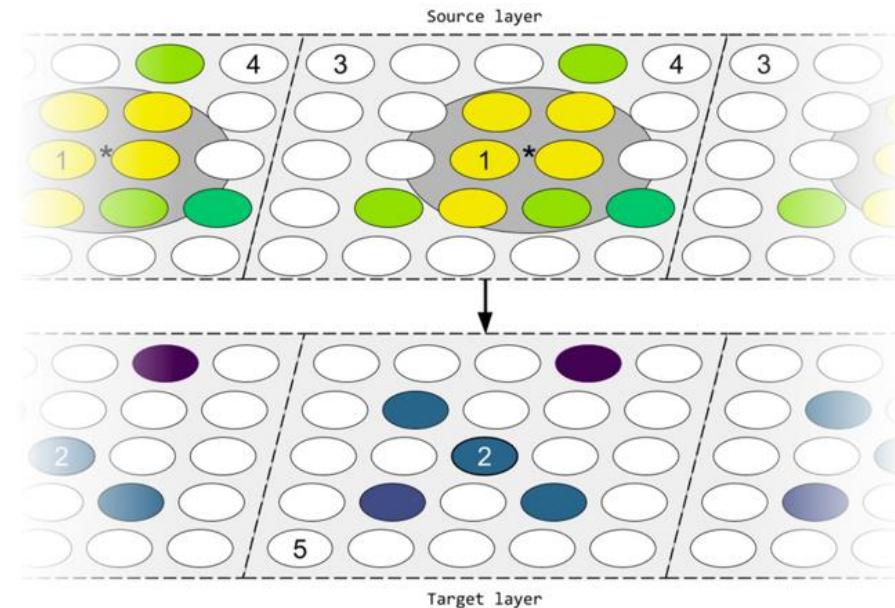


S. Habenschuss, Z. Jonke, and W. Maass, "Stochastic computations in cortical microcircuit models", PLOS Computational Biology, 9(11):e1003311, 2013.

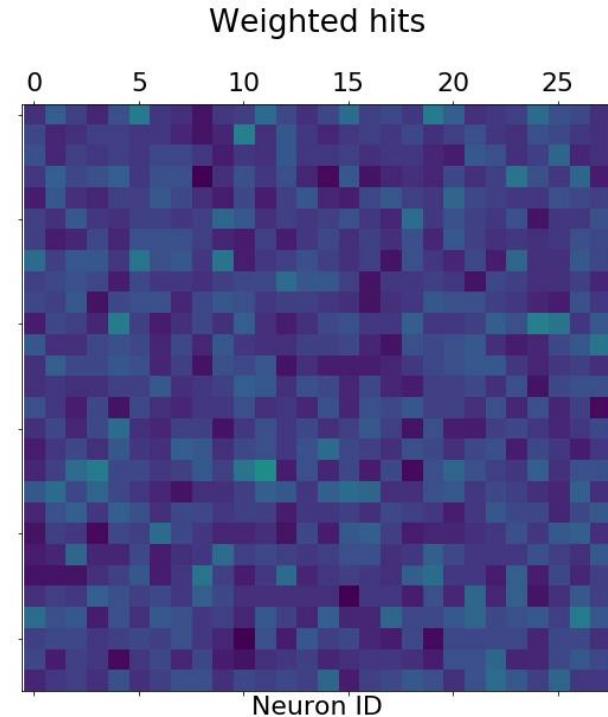
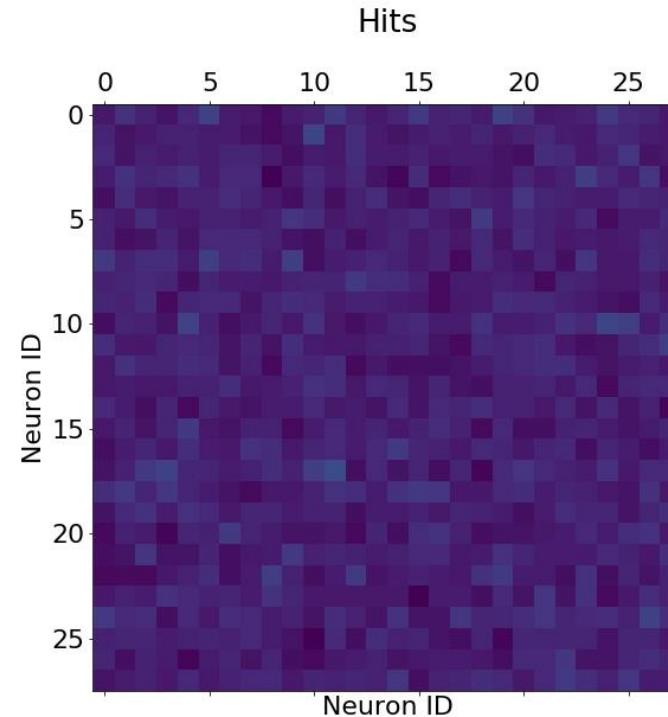
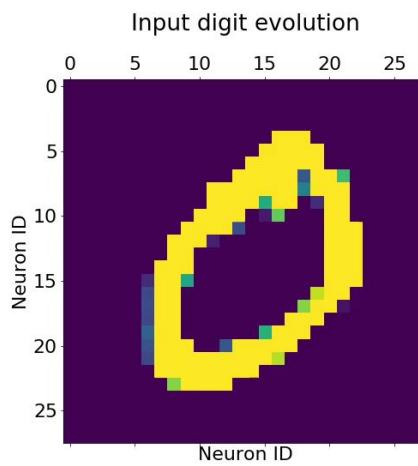
Theoretical Neuroscience

Structural plasticity

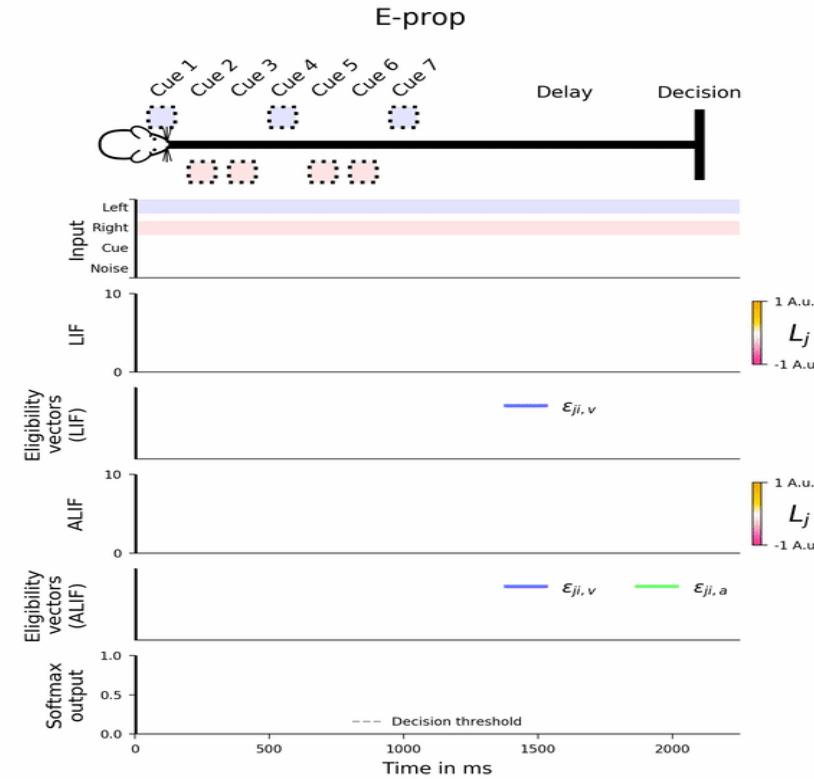
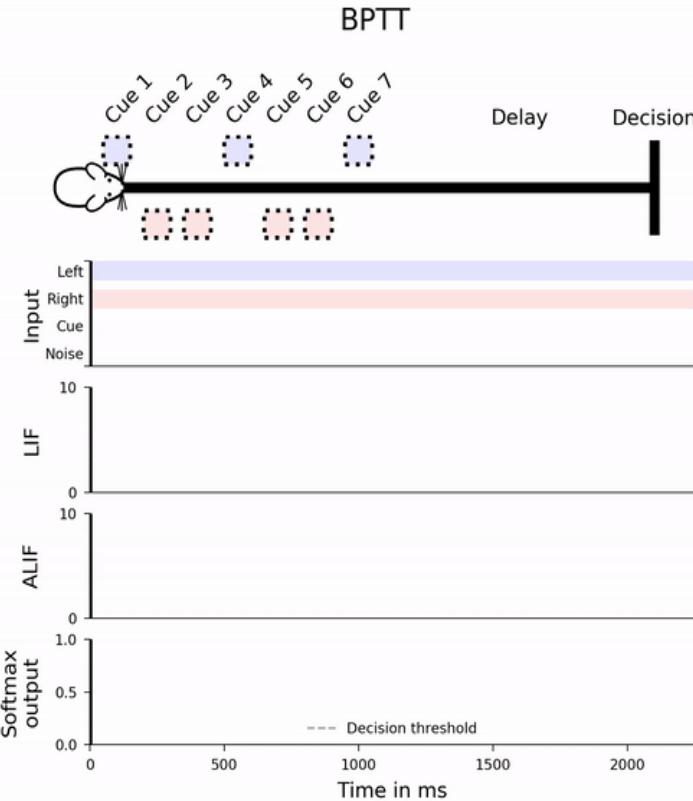
- Create/remove connections to facilitate learning/consolidation
 - feedforward and recurrent
 - distance-dependent receptive field
 - pruning of weak connections
- Computational challenge
 - update connection matrices on-the-fly
 - maintain network dynamics and computational performance



Theoretical Neuroscience



e-prop on SpiNNaker



SpiNNaker applications

Computational Neuroscience

Theoretical Neuroscience

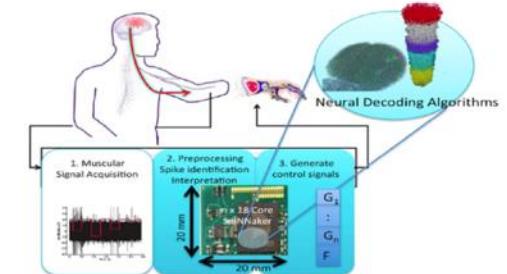
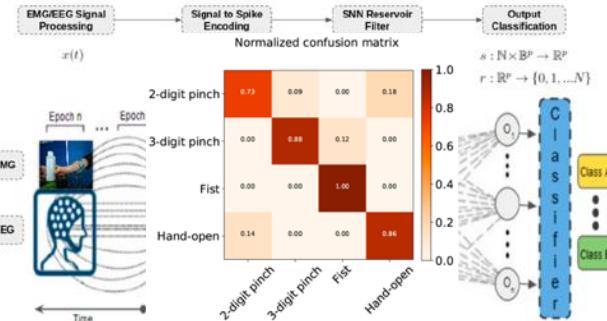
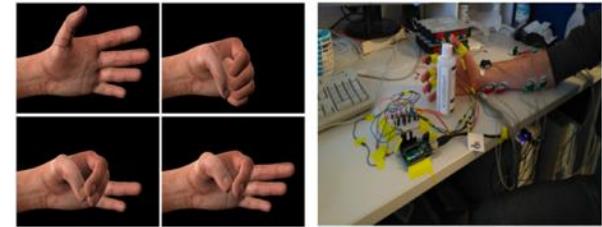
Neurorobotics

Thalamic/Cortical Input

BG Single Channel Structure

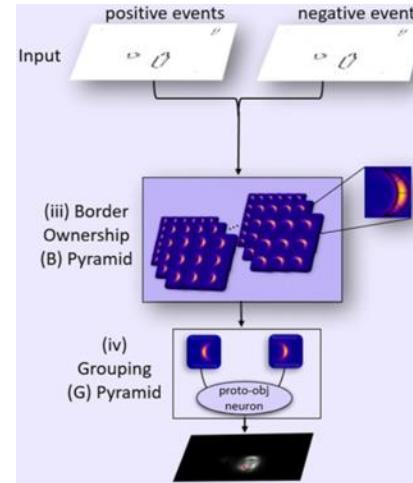
Prosthetic control

- 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



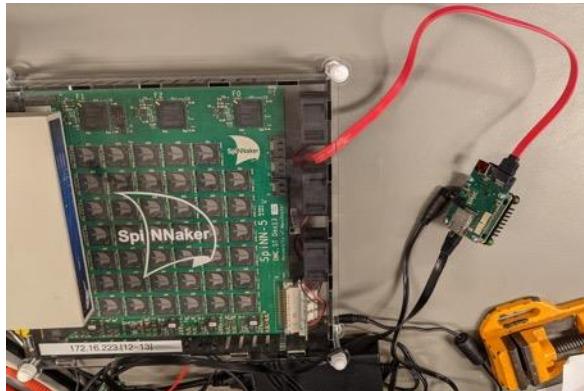
Robot vision

- Event-driven attentive system on iCub
 - SpiNNaker as neural substrate
- Event-driven vision sensors
- Produces saliency maps
- Average 16ms latency to produce a usable saliency map



SpiNNaker-based real-time vision

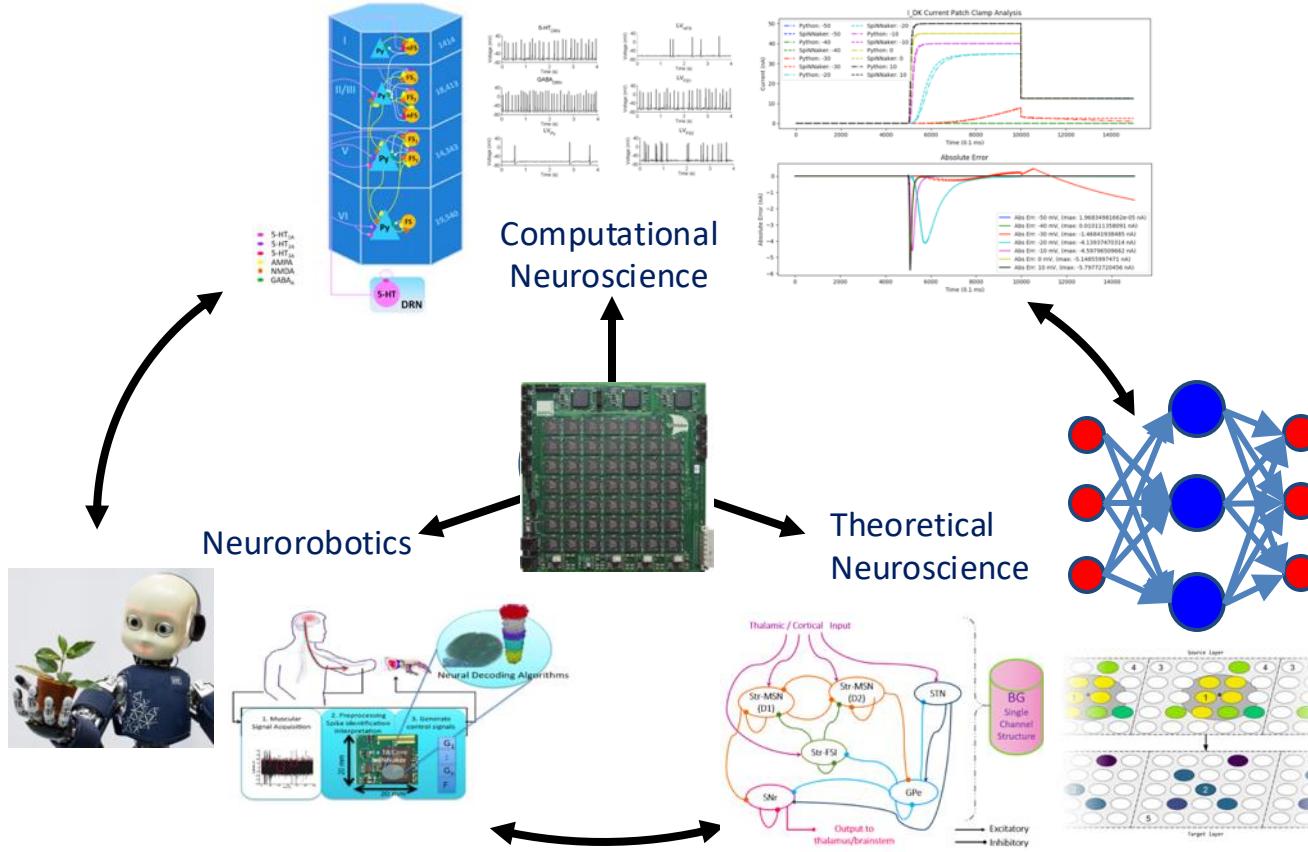
- Real-time convolutional spiking neural network on SpiNNaker
 - up to 12M events/s
 - USB or ethernet connections
 - 12 cores for input
 - 2x 12 cores for convolutions
 - FPGA-based interface
- Collaboration with Jörg Conradt, KTH



event camera (DVS)



SpiNNaker applications



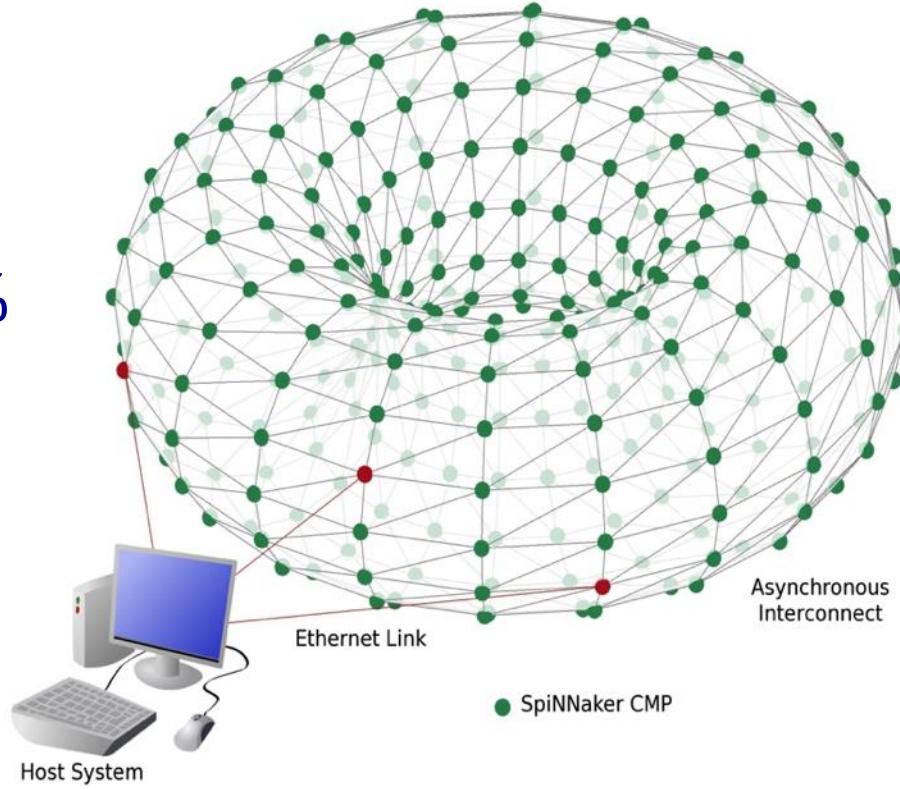
Outline



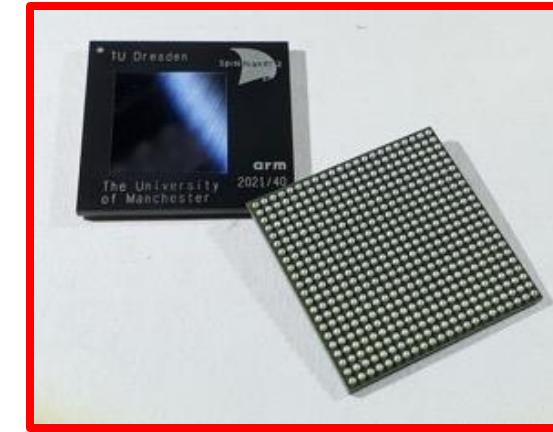
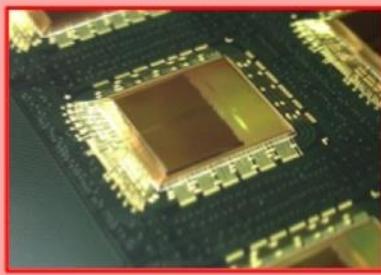
- Brains
- Neuromorphic computing
- The SpINNaker project
- SpINNaker applications
- SpINNaker2
- SpINNaker2 applications
- Conclusions

SpiNNaker concept

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



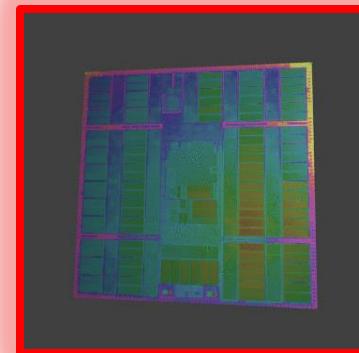
SpiNNaker chips



Multi-chip
packaging by
UNISEM



SpiNNaker1



SpiNNaker2



SpiNNaker machines

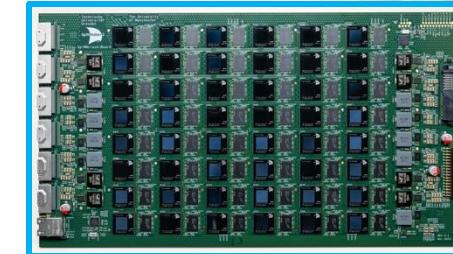
- HBP platform
 - 1M cores
 - 10 cabinets (plus server)
- Launch 30 March 2016
 - then 500k cores
 - ~450 remote users
 - 5M SpiNNaker jobs run
- SpiNNcloud platform
 - 5M cores
 - 8 cabinets (plus server)
- Launch 14 April 2025
 - then 500k cores
 - limited remote users
 - strategic apps pursued



Human Brain Project



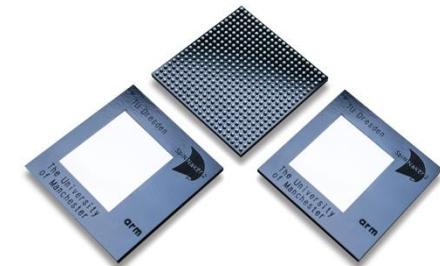
SpiNNaker1 racks
(1M ARM cores)



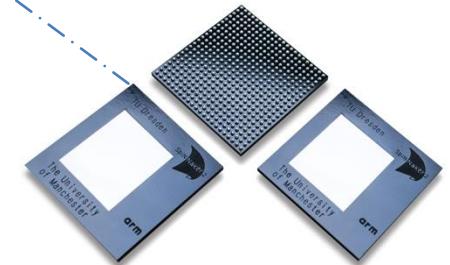
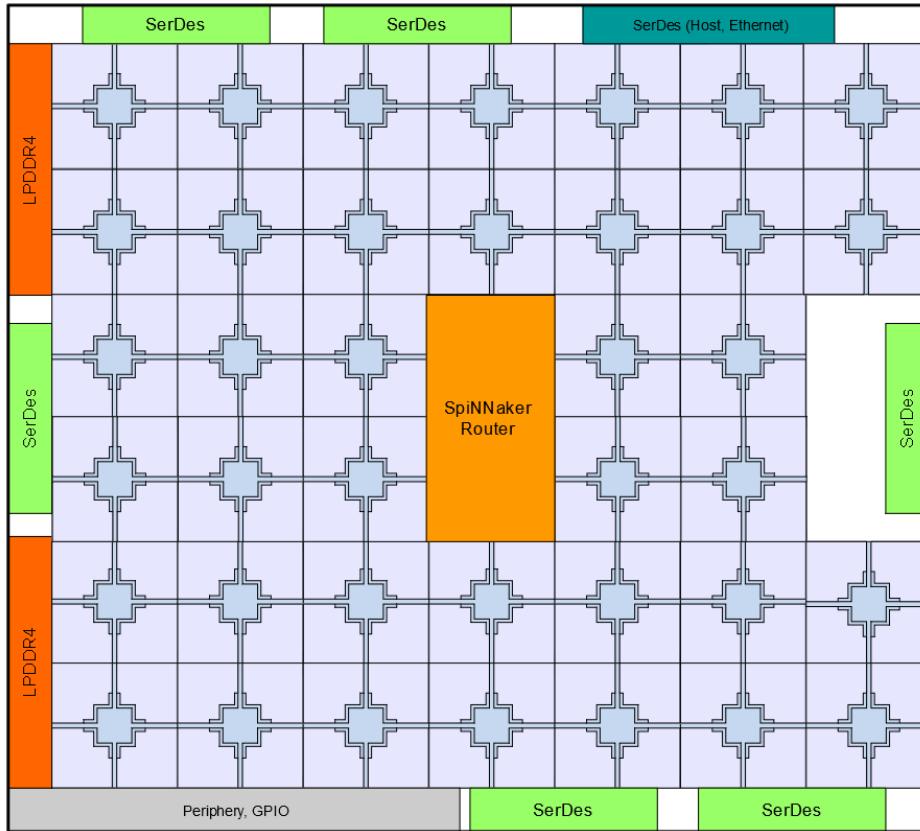
SpiNNaker2 racks
(5M ARM cores)

Spinnaker2

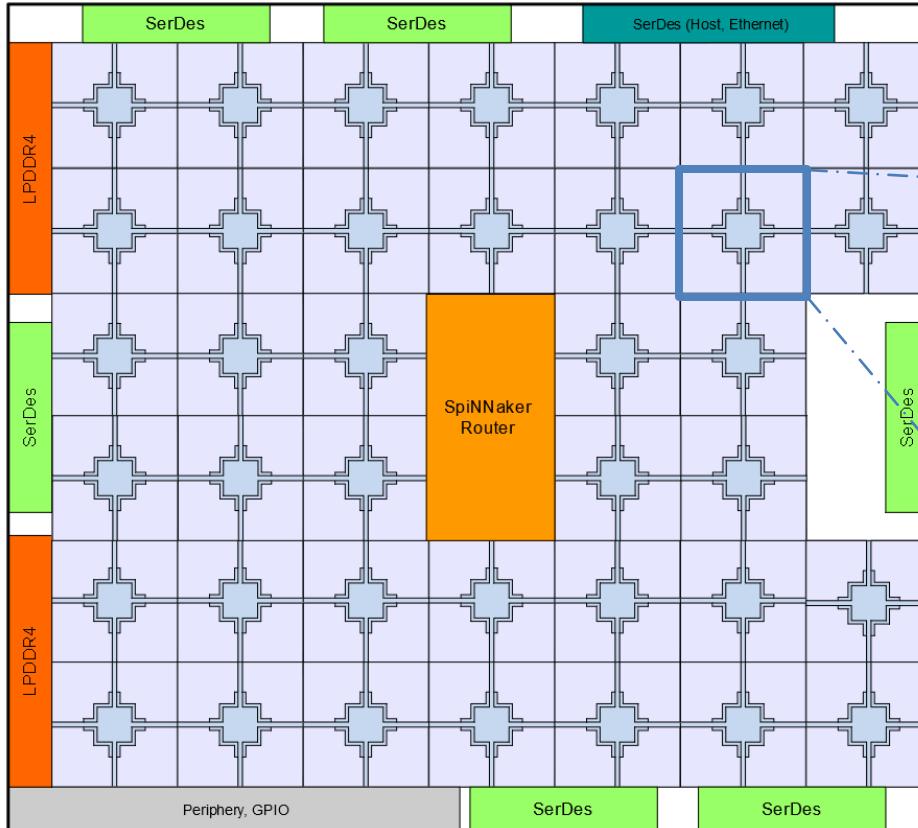
- 152 Arm-based processing elements in a low-power mesh
- High-speed infrastructure to connect chip-to-chip
- Light-weight and event-based Network-on-Chip
- Globally Asynchronous Locally Synchronous architecture
- Native accelerators for neuromorphic (Exp/Log/TRNG) and Machine Learning (Custom MACs)
- Adaptive Body Biasing (ABB) in 22FDX using reverse body bias
- Patented Dynamic Voltage Frequency Scaling at core-level



Spinnaker2



Spinnaker2



Dynamic Power Management
at a **core-level**

SpiNNaker2 Processing Element

Dynamic Power Management for enhanced energy efficiency

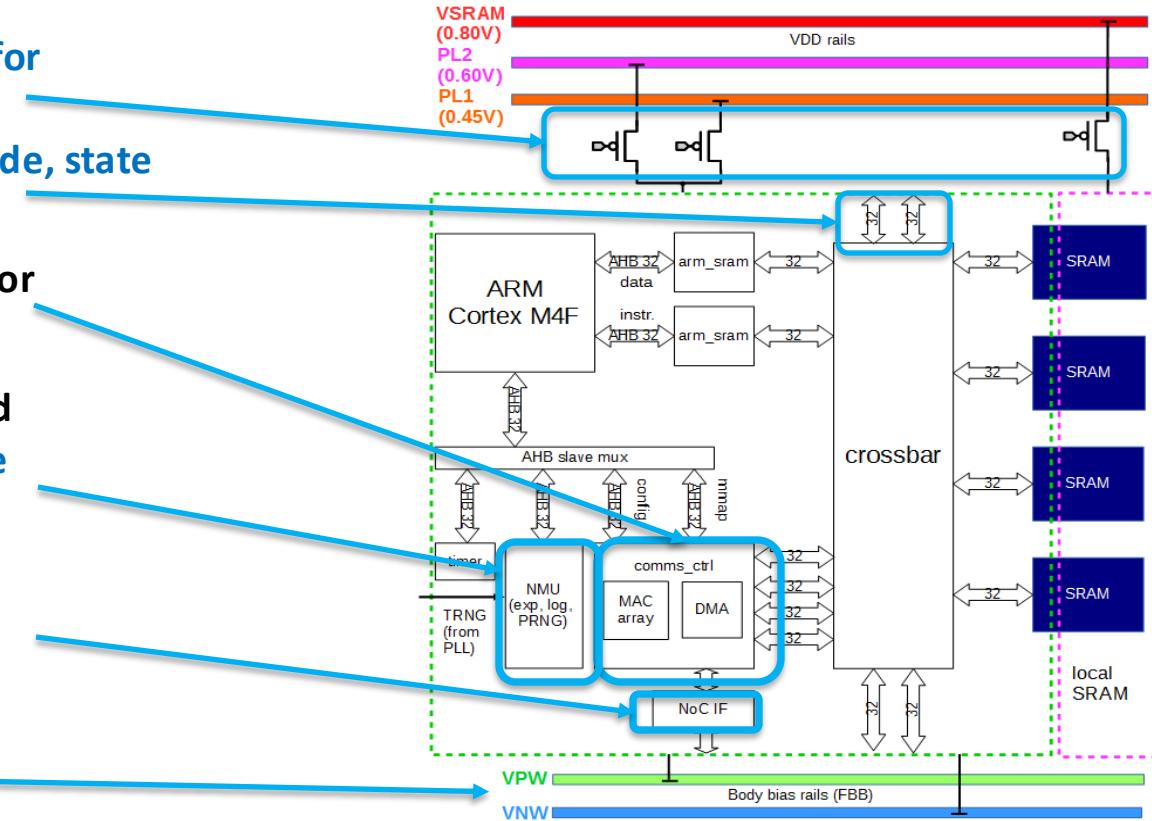
Memory sharing for flexible code, state and weight storage

Multiply-Accumulate accelerator for machine learning

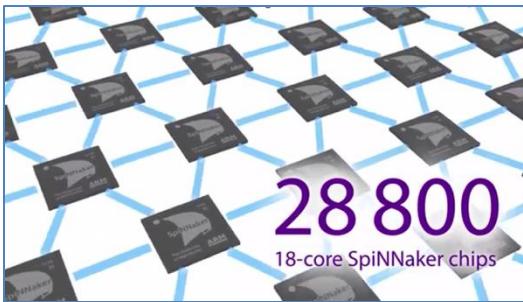
Neuromorphic accelerators and random generators for synapse and neuron computation

Network-on-Chip for efficient spike communication

Adaptive Body Biasing for energy efficient low voltage operation



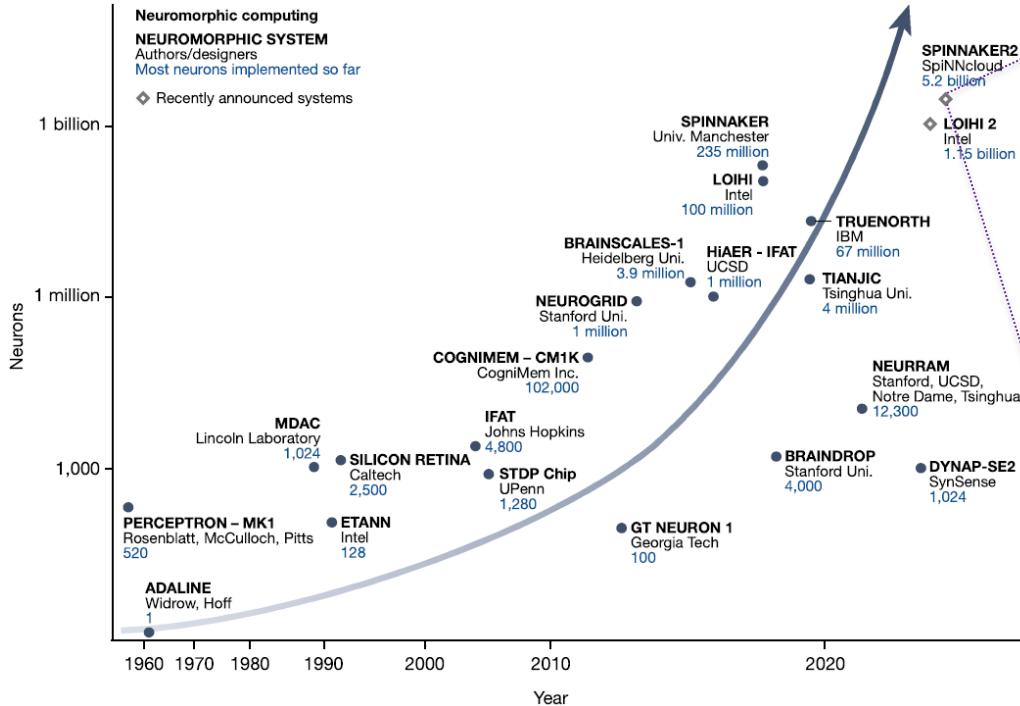
SpiNNaker concept



so the boards are
rearranged
to avoid them

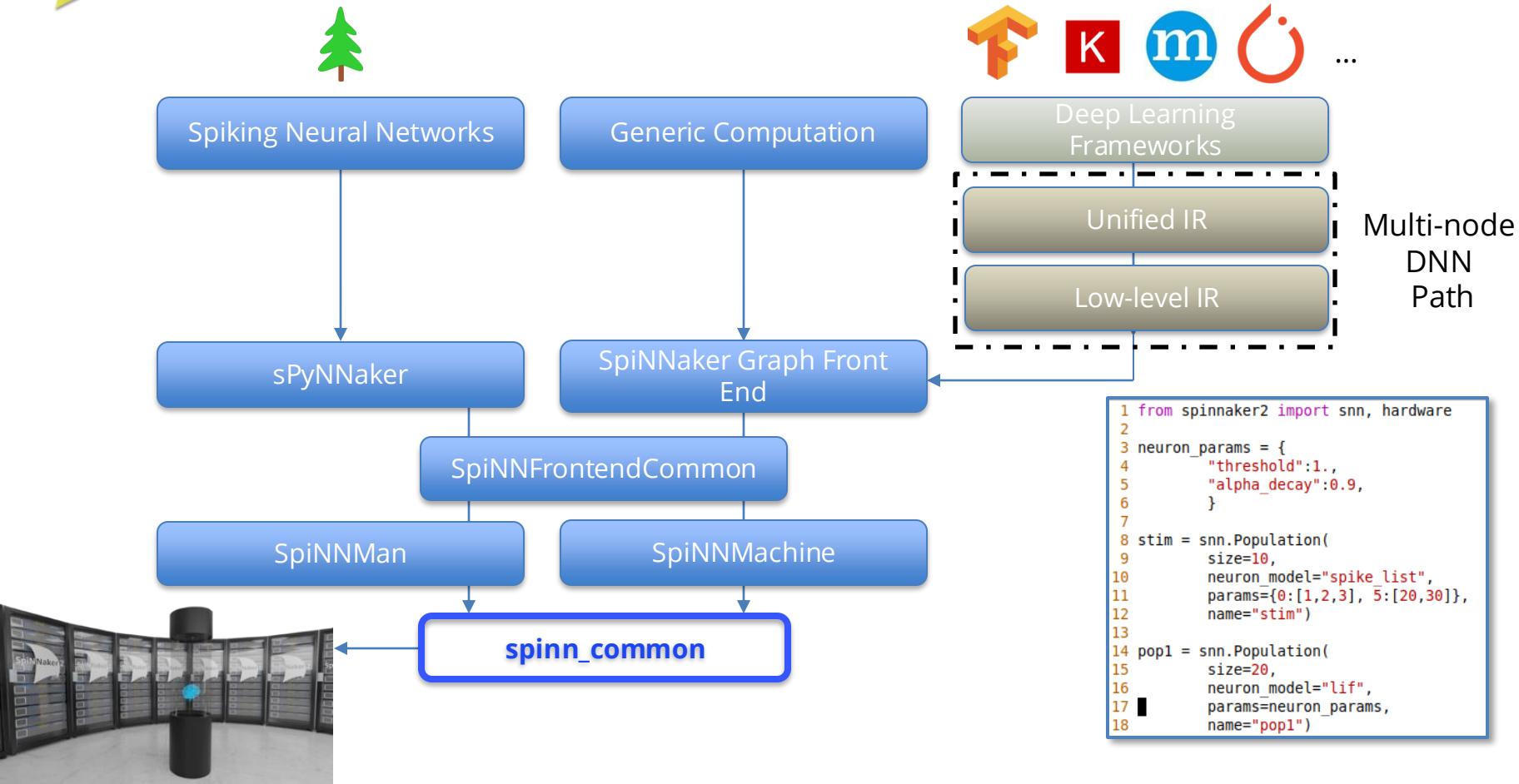


Large-scale neuromorphic systems



Source: <https://www.nature.com/articles/s41586-024-08253-8>

High-level software flow



Outline



- Brains
- Neuromorphic computing
- The SpINNaker project
- SpINNaker applications
- SpINNaker2
- SpINNaker2 applications
- Conclusions

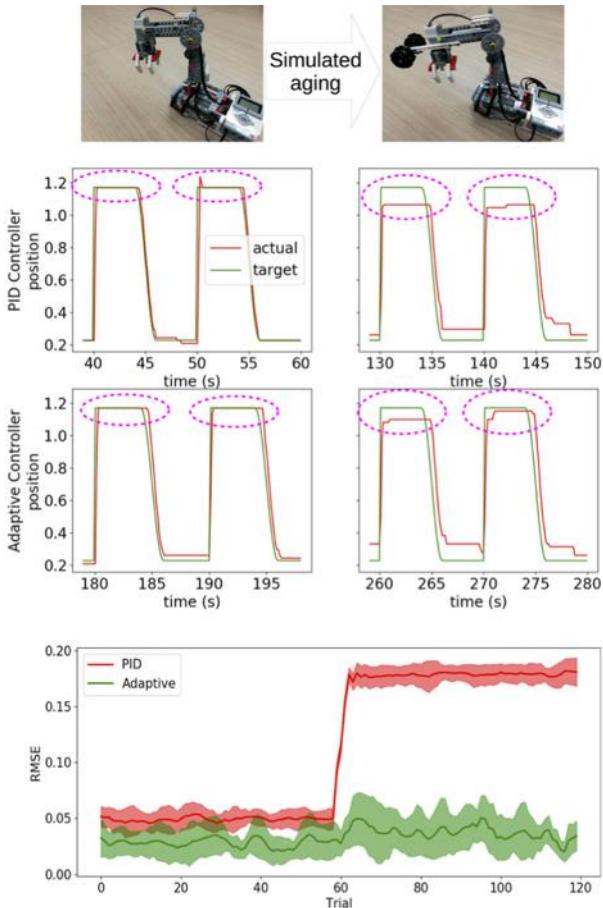
Comparing Loihi with SpiNNaker2

Table 1. Comparison of the SpiNNaker 2 prototype (SpiNN) and Loihi for the keyword spotting task.

Hardware	Inference/s	Energy/inference (μ J)
SpiNN	1000	7.1
Loihi	296	37

- Two applications compared:
 - Low-latency keyword spotting
 - Simulated aging of a robot arm
- Loihi shows better efficiency for less complicated vector-matrix multiplication
- SpiNNaker2 (prototype) shows better efficiency for high-dimensional vector-matrix multiplication

Yan, Y., Stewart, T., Choo, X., Vogginger, B., Partzsch, J., Hoppner, S., Kelber, F., Eliasmith, C., Furber, S. & Mayr, C., *Comparing Loihi with a SpiNNaker2 prototype on low-latency keyword spotting and adaptive robot control*. IOP Neuromorphic Computing and Engineering 1 (2021) 014002.

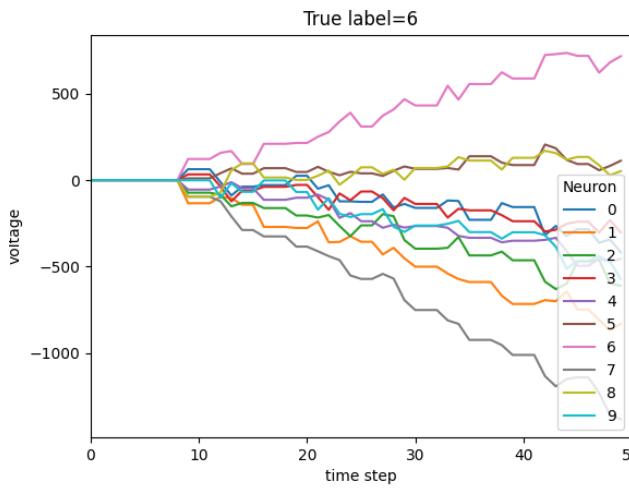


Hybrid Examples

Neural Apps on ARM core

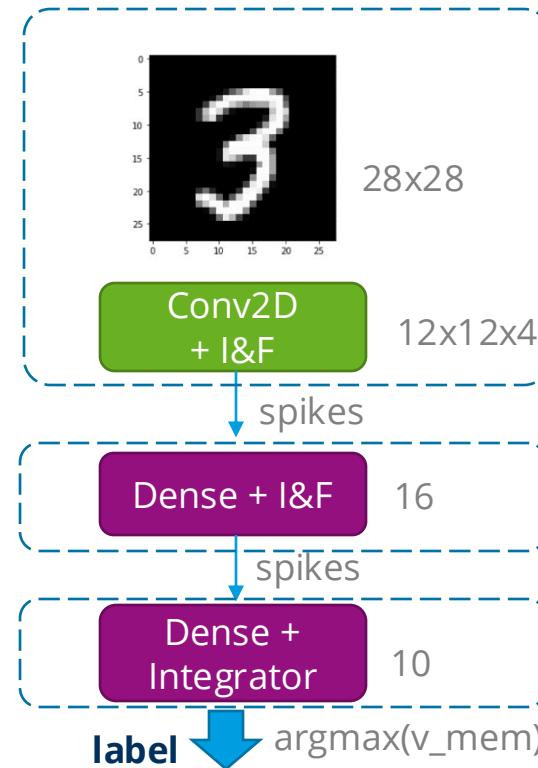
`conv2d_if_neuron_rate`

`lif_neuron`

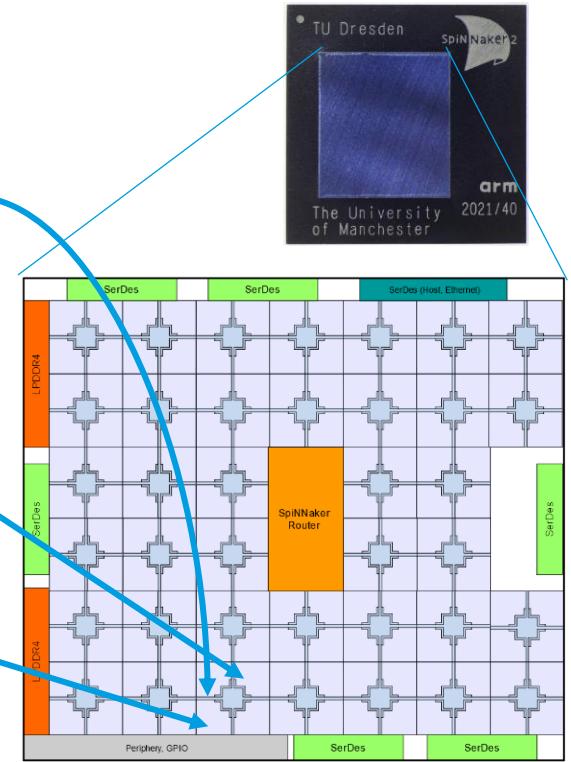


Accuracy of 97.5 % identical to DNN in Tensorflow

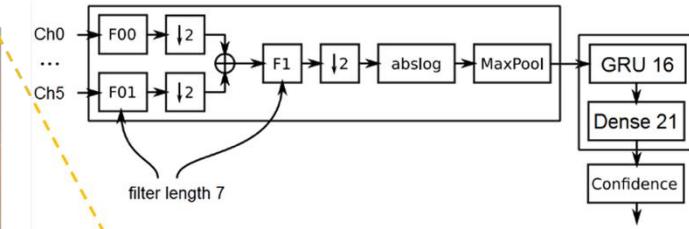
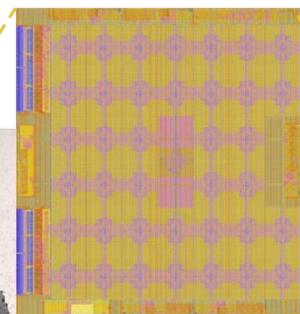
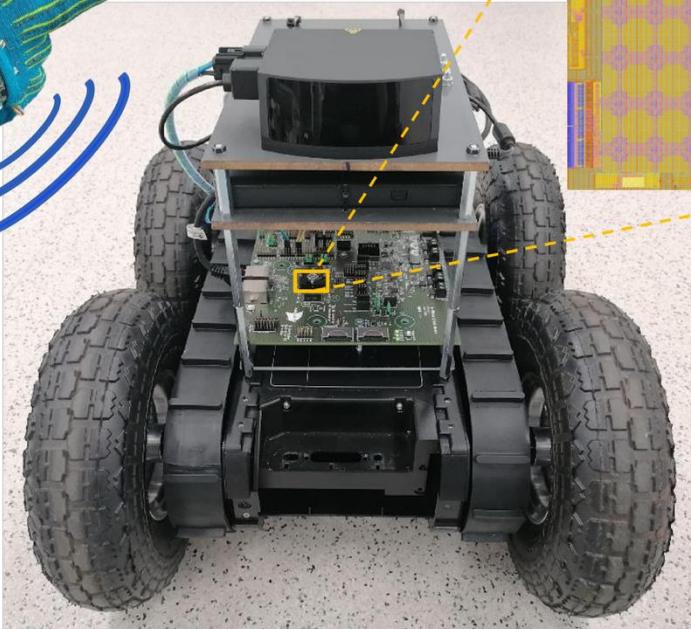
Hybrid Network



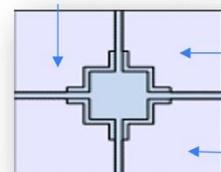
Automatic Mapping



Edge Applications

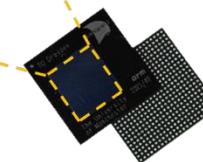


Buffering & Preprocessing



3 Cores

Sensor Parsing & Driving Commands



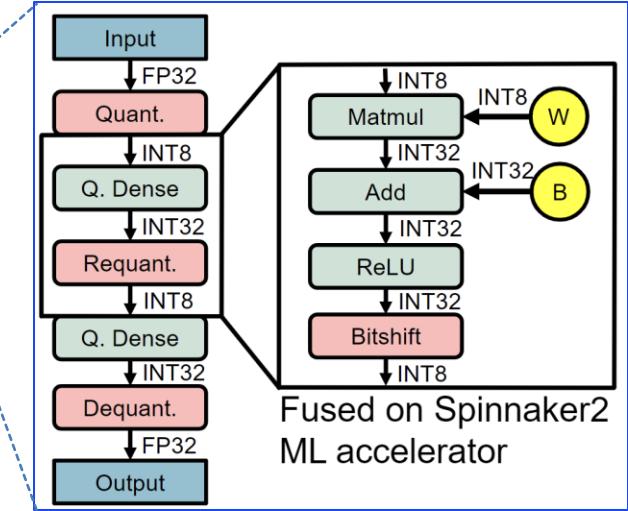
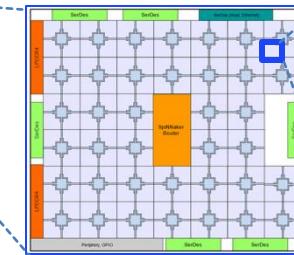
High-Level IR

Tensor Expression IR

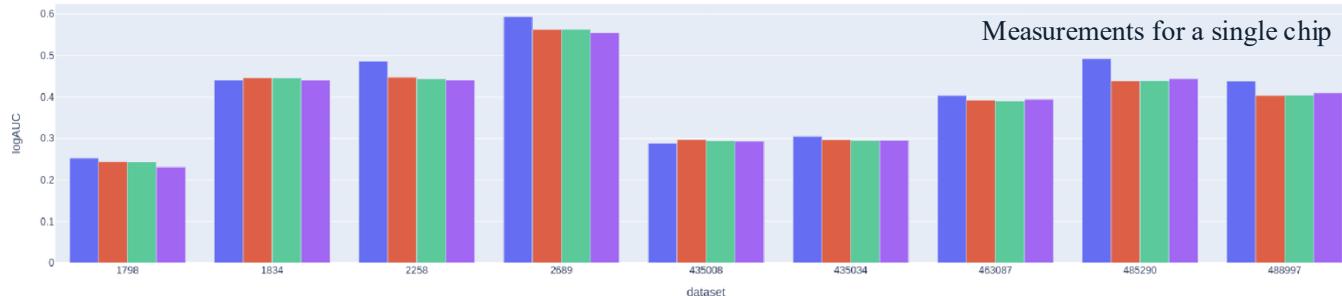
*.c, *.h Files
Ahead of Time
GRU Classifier

TensorFlowLite

Drug discovery



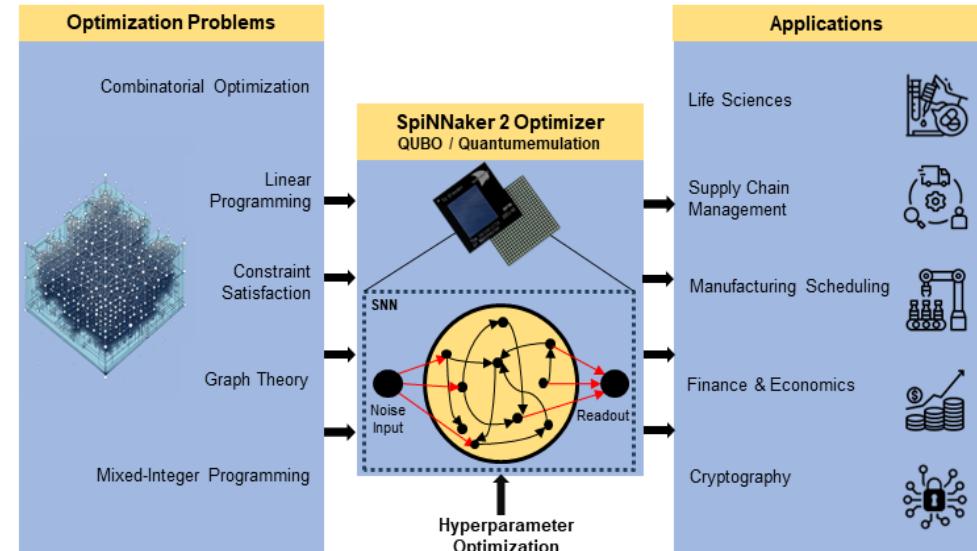
Log(Area under Curve) of Hits in a virtual High Throughput Screening (vHTS)
 A Hit is a molecule/ligand with a positive reaction with the target molecule provided as input



precision
 bcd
 full
 quantized
 bitshift

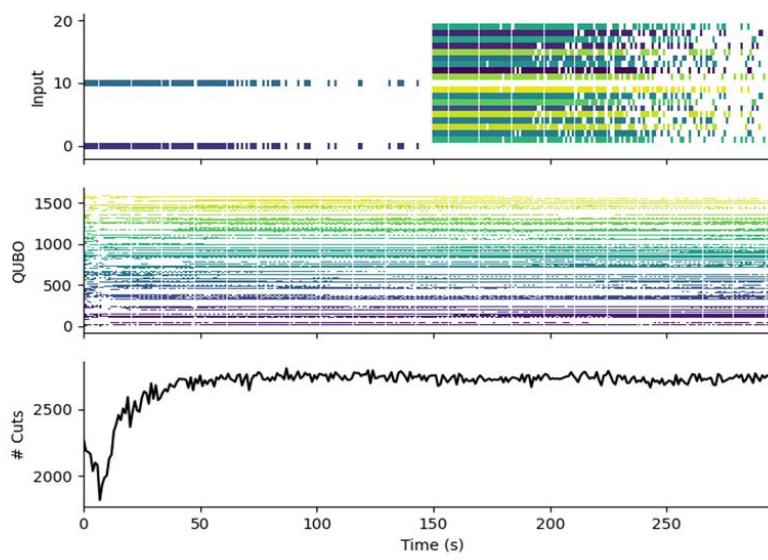
QUBO

- Finds optimal solutions on small problems
 - currently fine-tuning neuronal dynamics and benchmarking on more complex problems
- Can reach sub-optimal solutions in a few milliseconds
 - This quickly-acquired sub-optimal solution can be used by other algorithms to reach an optimal solution
 - hybrid approach



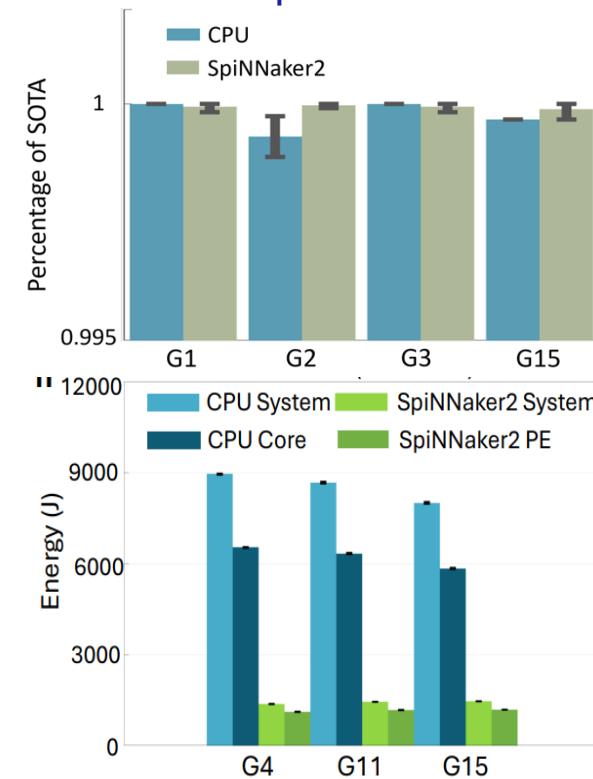
QUBO

SNN-based QUBO solvers on SpiNNaker2 achieve state-of-the-art performance at a fraction of the energy cost.



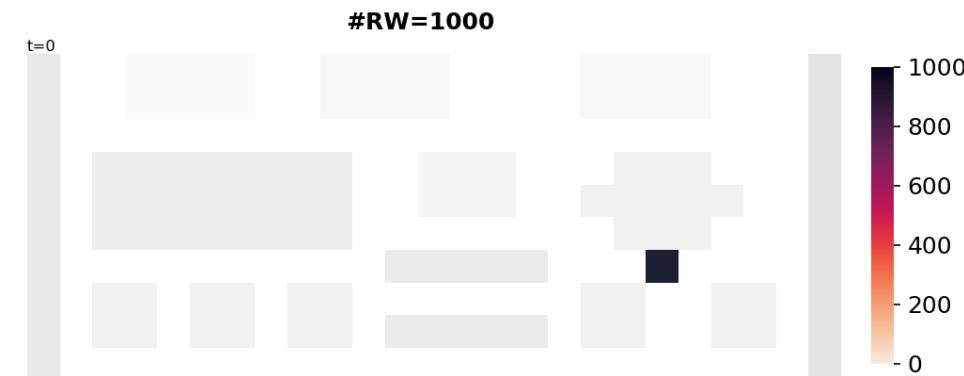
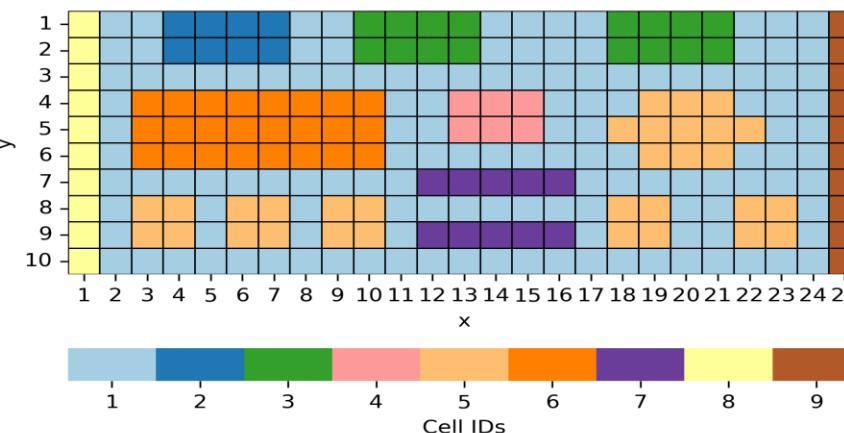
Maxcut 800-node graph SpiNNaker solution

Chen, Z., Xiao, Z., Akl, M. et al. ON-OFF neuromorphic ISING machines using Fowler-Nordheim annealers. *Nat Commun* 16, 3086 (2025). <https://doi.org/10.1038/s41467-025-58231-5>

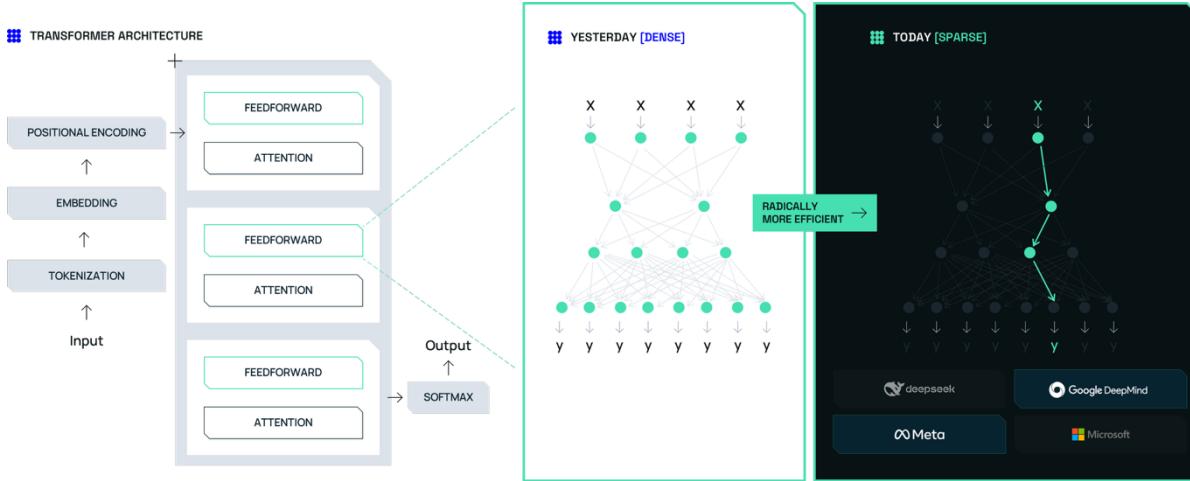


Random Walkers

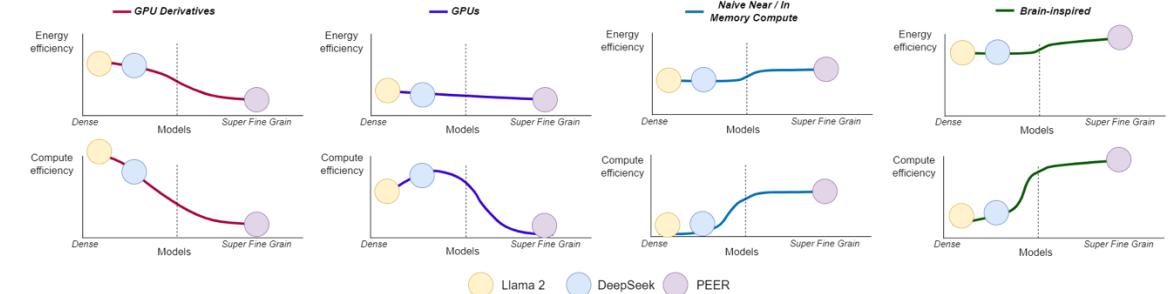
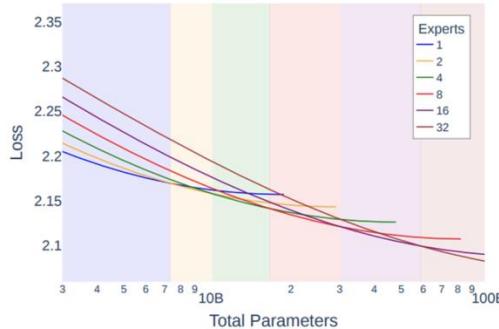
Neural Random Walkers on SpiNNaker 2 simulate the stochastic process for solving heat flow across a circuit board with a cooling fan flow



Mainstream AI trends



[source]



Outline



- Brains
- Neuromorphic computing
- The SpINNaker project
- SpINNaker applications
- SpINNaker2
- SpINNaker2 applications
- Conclusions

Conclusions

- **Industrial AI:**

- is based on dense matrix operations:
 - computes everything all the time
 - “deathly embrace” of algorithms and GPU hardware

- ***SpiNNaker and SpiNNaker2:***

- designed on brain-like principles:
 - support spatial & temporal sparsity
 - event-based software model

- **Neuromorphics and AI are converging**

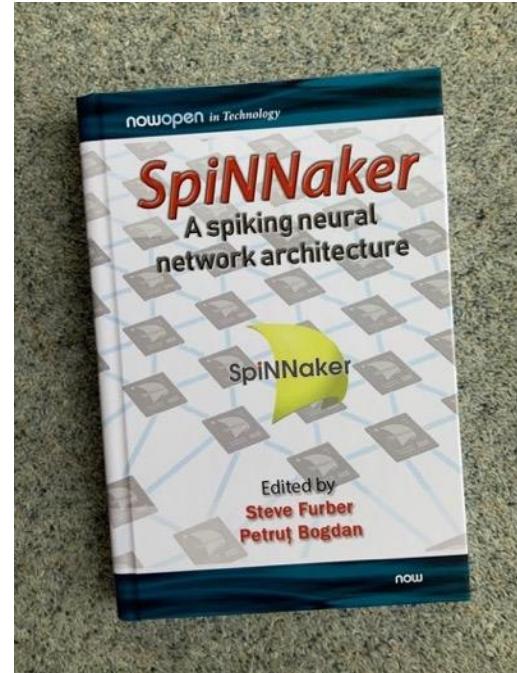
- brain-like principles will help address the unsustainable energy demands of AI



SpiNNaker book

- Open Access in PDF form (i.e. free!)
 - \$90 on paper

*20 years in conception and 15 in construction,
the SpiNNaker project has delivered the
world's largest neuromorphic computing
platform incorporating over a million ARM
mobile phone processors and capable of
modelling spiking neural networks of the scale
of a mouse brain in biological real time...*



<http://dx.doi.org/10.1561/9781680836523>