Human Brain Scale AI - coming soon?

EXPO 2020 DUBAI UAE

SLOVAK PAVILION

OFFICIAL PARTICIPANT

Steve Furber CBE FRS FREng

ICL Professor of Computer Engineering
The University of Manchester

MANCHESTER 1824 The University of Manchester | EPSRC | SpiNNaker | Human Brain Project | ARM

AI history

1939
Tin man
Wizard of Oz

1950
Turing –
“Can machines think?”
Turing test

1956
Dartmouth Summe
Research Project on AI
(McCarthy&Minsky)

1980s
Connectionism
(Hopfield, Rumelhart)
& Expert Systems
(Feigenbaum)

1997
Kasparov defeated
by Deep Blue
(chess)

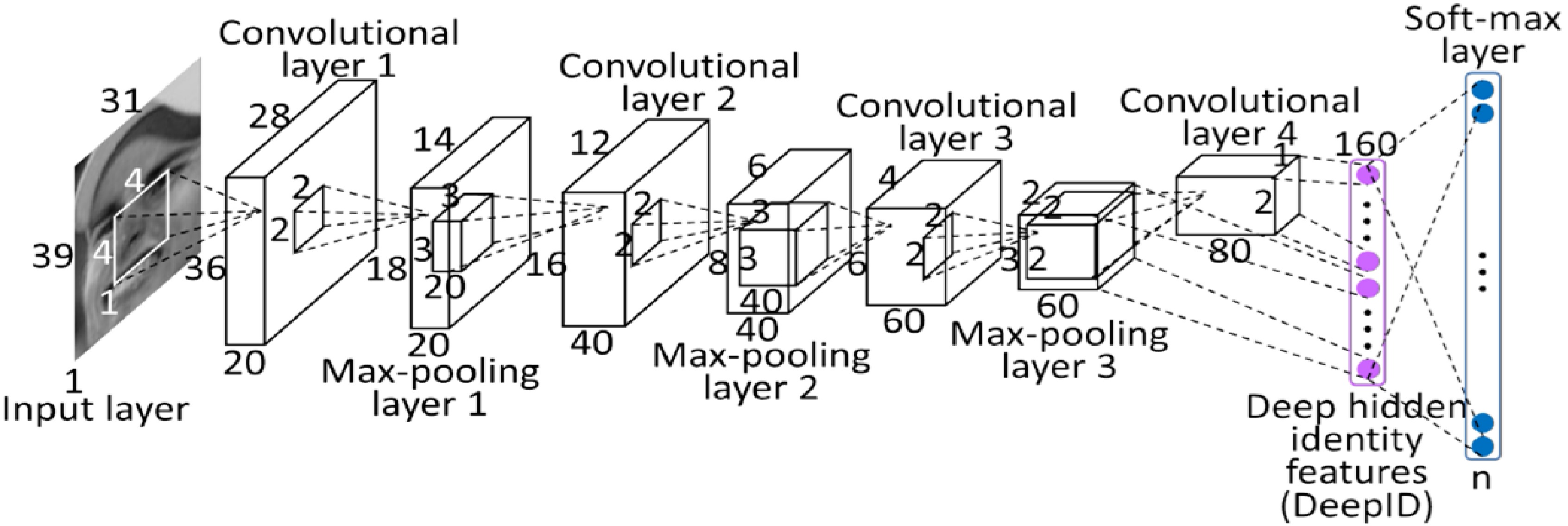
2006
Hinton & others,
deep belief nets

2009
IBM cat cortex model (Modha)
- 109 neurons, 1012 synapses
- IBM BlueGene/P
- 147,456 cores, 144 TB memory

2016
Lee Sedol defeated
by AlphaGo (Go)

AI today

- Deep learning
  - Massive computational power, big data, huge energy
  - E.g. recognize objects, translate speech in real time
- Explainable AI (e.g. Kyndi)
  - AI decisions can be queried and reasons given
- General AI (e.g. GoodAI)
  - “Develop safe general AI – as fast as possible – to help humanity and understand the universe”
- Brain-inspired AI – neuromorphic computing
  - Spikes – event-based models
- AI is evolving rapidly in many directions, requiring general-purpose (rather than specialized) computing platforms

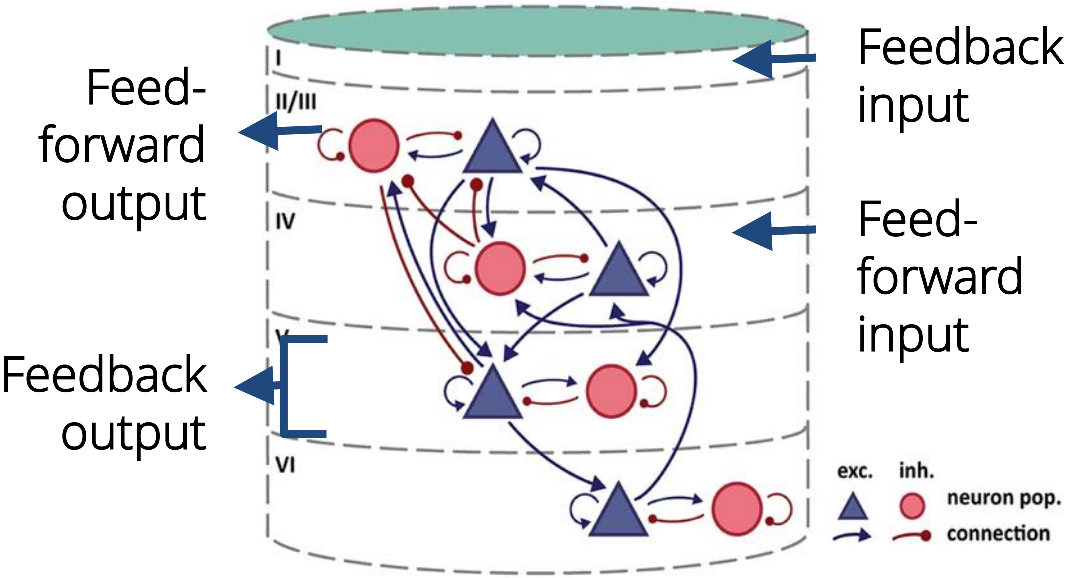


ConvNets - structure

- 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%



Neuromorphic Computing

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

Carver Mead
California Institute of Technology (1980s)

Why focus on the brain?

Human Brain Project

1. Understanding the brain (Unifying Science Goal)

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

2. 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.

3. 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 system
1 Million ARM cores
Real-time simulator

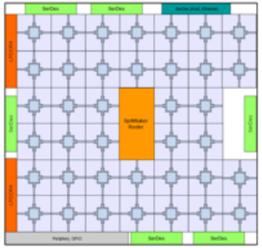
1st generation BrainScaleS-1 Machine

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

Many-core Architecture SIMULATION

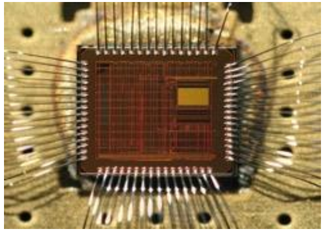
Physical mode EMULATION

Towards 2nd generation SpiNNaker-2



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

Towards 2nd generation BrainScaleS-2

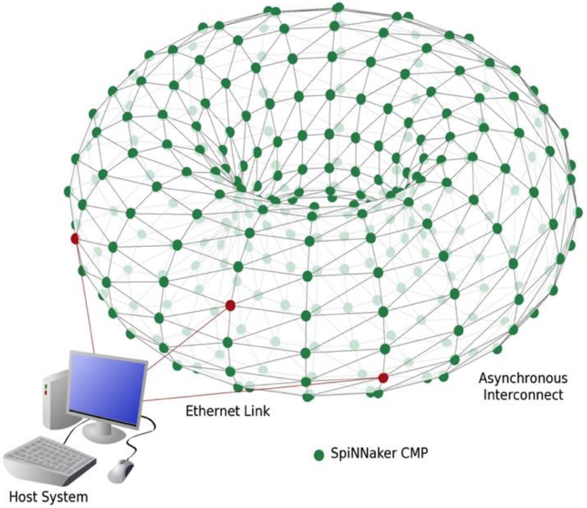


On-chip plasticity processors
Flexible hybrid plasticity
Active dendrites

Designed and built from the transistor up !

Common software ecosystem, remote access, open user facility Co-designed with (theoretical) neuroscience

SpiNNaker project

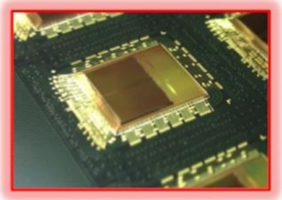


1 million
mobile phone processors in one computer

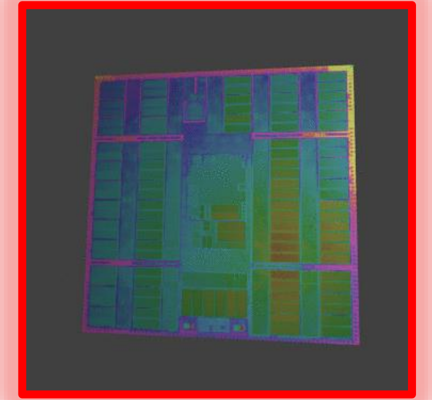
1%
of the human brain

or

10
mice



Multi-chip packaging by UNISEM



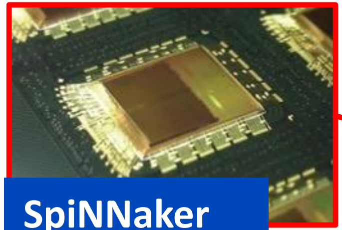
SpiNNaker machines

HBP platform

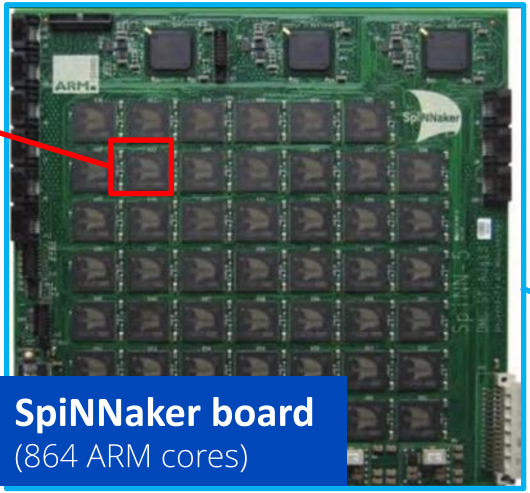
- 1M cores
- 11 cabinets (including server)

Launch 30 March 2016

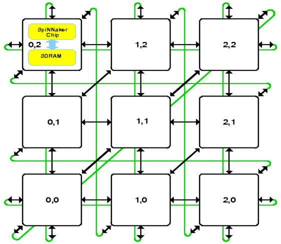
- then 500k cores
- ~450 remote users
- 5M SpiNNaker jobs run

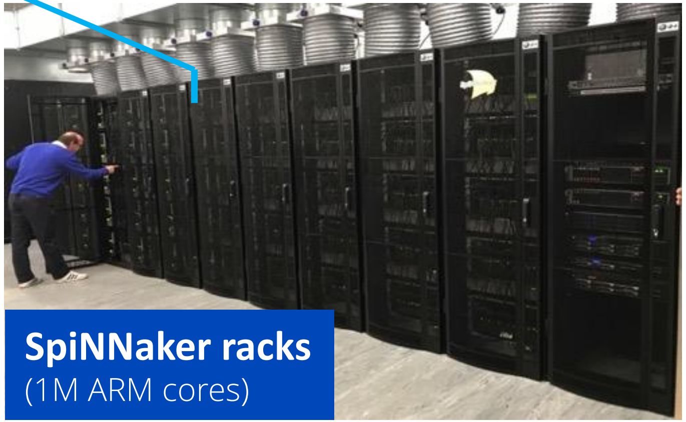


SpiNNaker chip
(18 ARM cores)



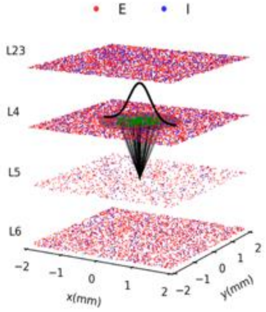
SpiNNaker board
(864 ARM cores)

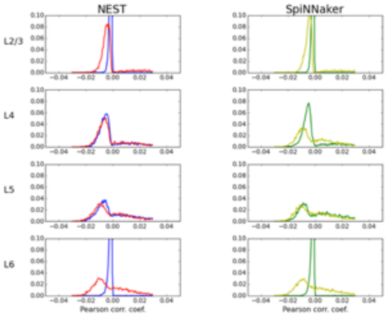




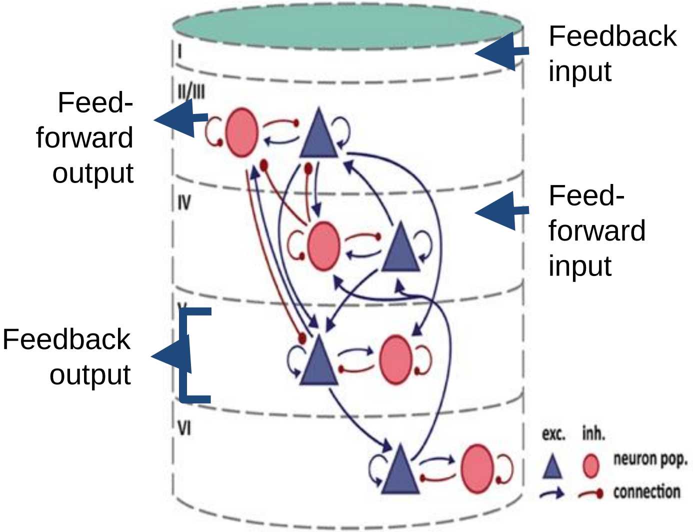
SpiNNaker racks
(1M ARM cores)

Cortical microcircuit





- ***Realtime execution of cortical model***
  - 1mm^2 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 100mm^2 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.

Human brain scale AI

- 10^{11} neurons
  firing at < 10 Hz
- 10^{14} synapses
  petabytes of memory
- 10^{15} connections/sec
  exascale compute
- 100x to 1,000x SpiNNaker

Slovakian Super Computer to reach the Human brain scale AI

1 AI ExaFLOPs of Training and Inferencing per rack
64 AI ExaFLOPs
>500 DP PetaFLOPs

Ceph-based storage rack
6.6 petabyte per rack of usable storage
16 – 32 Storage Racks
100 – 200 PB

Conclusions

Human brain scale AI

- Requires huge computational resources
  - Petabytes of memory
  - Exascale compute
  - 100x to 1,000x SpiNNaker

- Within reach of the AI supercomputer Slovakia plans to build by the end of 2022