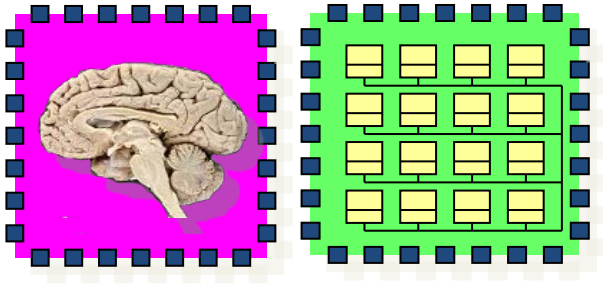
SpiNNaker
Biologically Inspired Massively Parallel Architectures



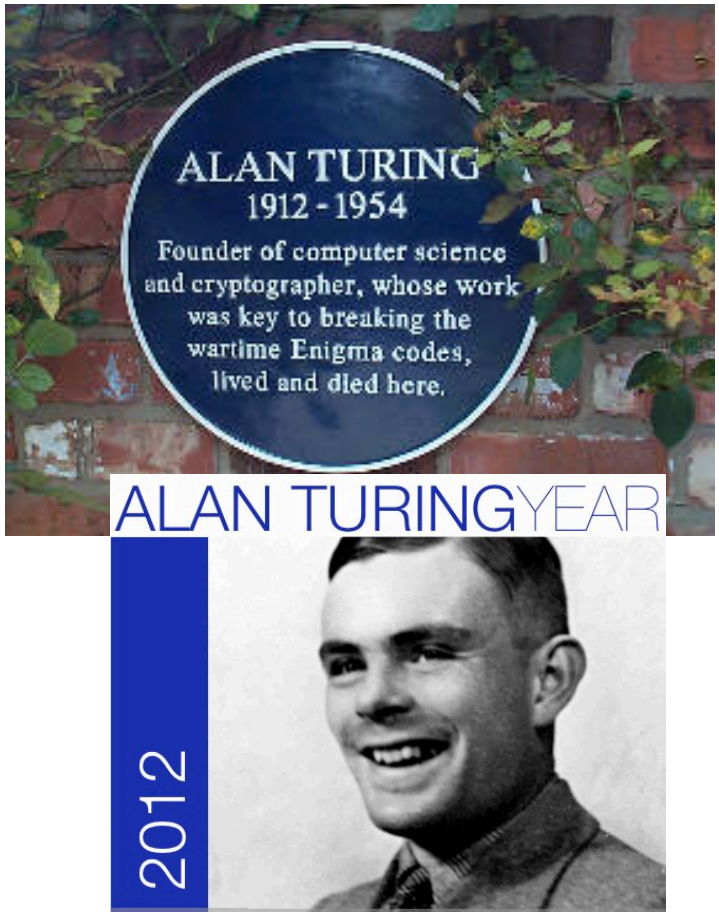
Biologically-Inspired Massively-Parallel Computation

Steve Furber
The University of Manchester
steve.furber@manchester.ac.uk

EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Turing Centenary



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Turing in Manchester

Computing Machinery and Intelligence

A. M. Turing

1950

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

EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Outline

- 63 years of progress
- Many cores make light work
- Building brains
- The *SpiNNaker* project
- The networking challenge
- A generic neural modelling platform
- Plans & conclusions

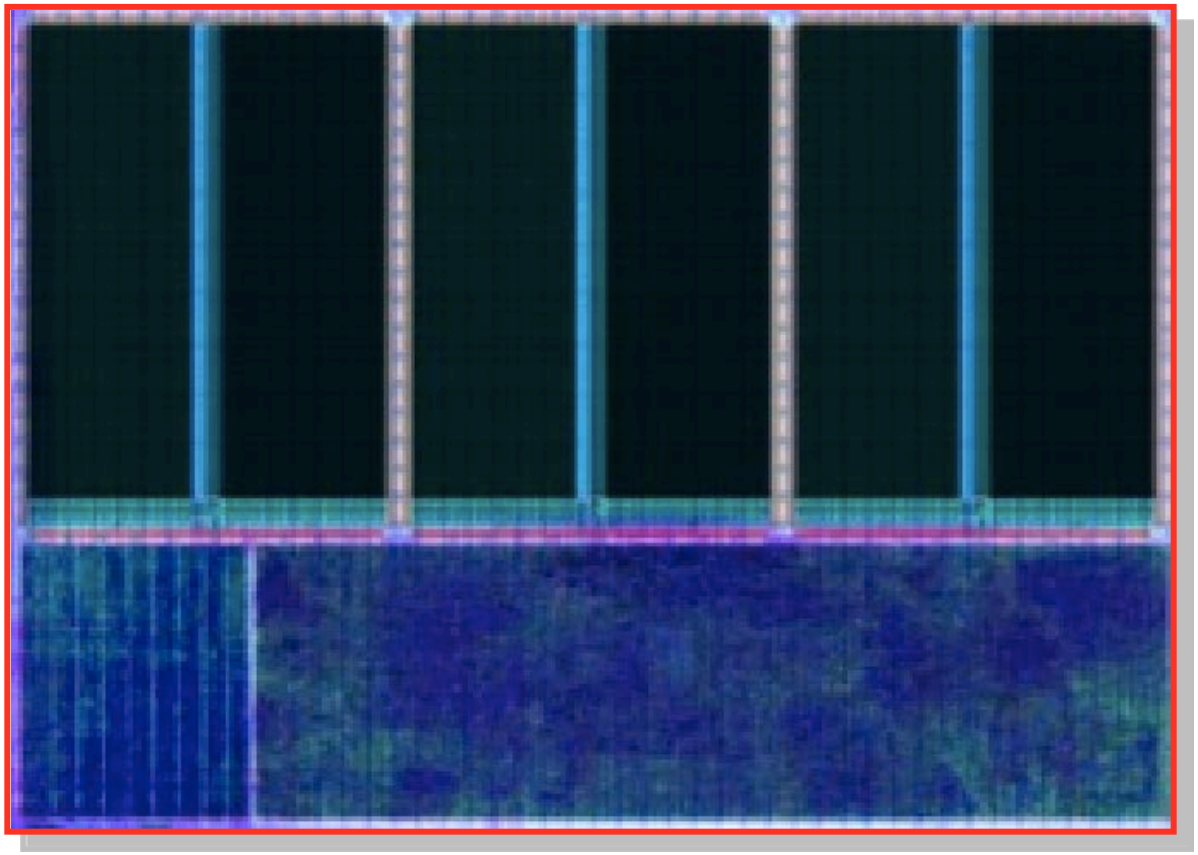
EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Manchester Baby (1948)

SpiNNaker
Biologically Inspired Massively Parallel Architectures

SpiNNaker CPU (2011)

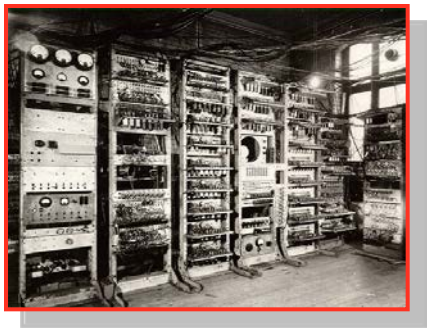


EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

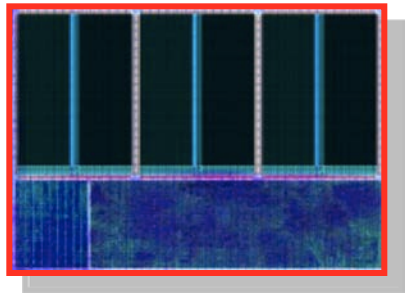
SpiNNaker
Biologically Inspired Massively Parallel Architectures

63 years of progress

- ***Baby:***
  - filled a medium-sized room
  - used 3.5 kW of electrical power
  - executed 700 instructions per second



- ***SpiNNaker ARM968 CPU node:***
  - fills ~3.5mm^2 of silicon (130nm)
  - uses 40 mW of electrical power
  - executes 200,000,000 instructions per second



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Energy efficiency

- Baby:
  - 5 Joules per instruction
- SpiNNaker ARM968:
  - 0.000 000 000 2 Joules per instruction

25,000,000,000 times better than Baby!

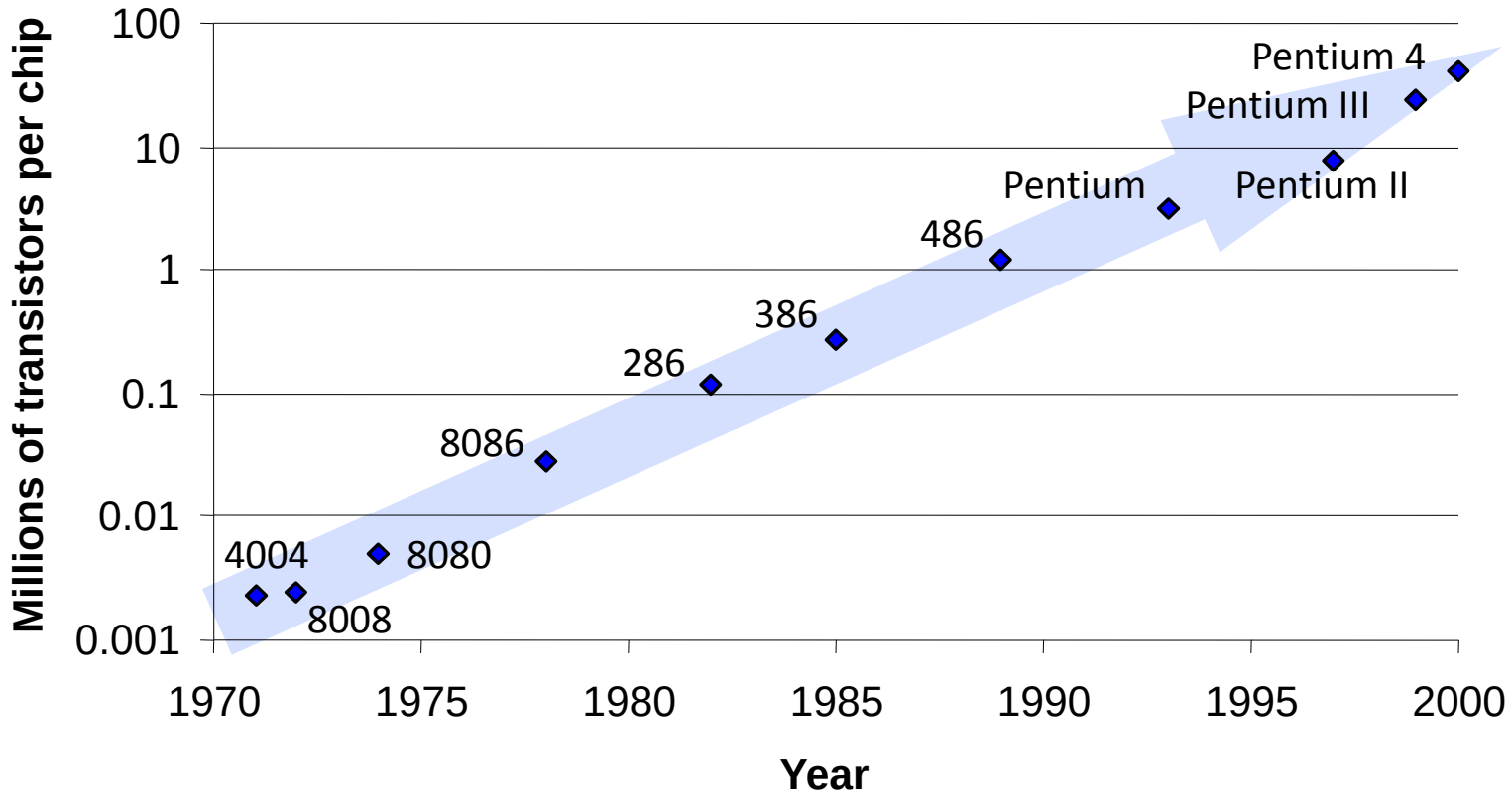
(James Prescott Joule born Salford, 1818)

EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Moore's Law

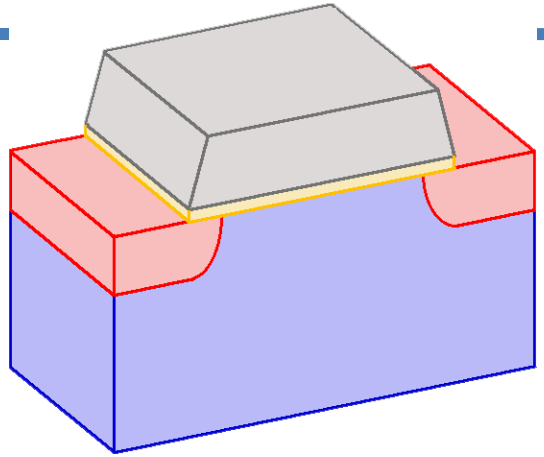
Transistors per Intel chip

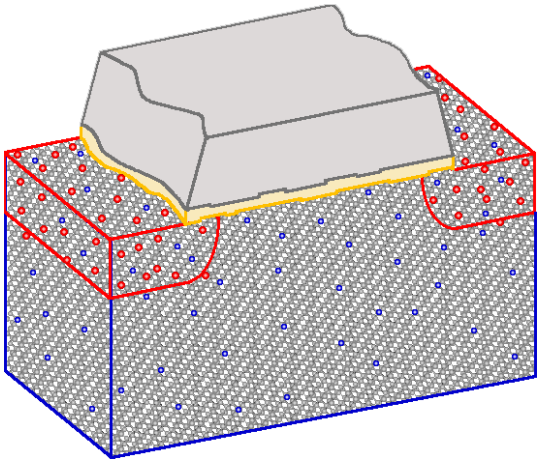


EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

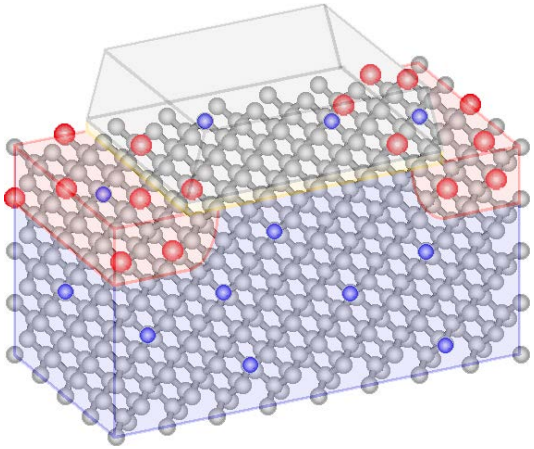
...the Bad News





DEVICE MODELLING GROUP

UNIVERSITY of GLASGOW



- atomic scales
  - less predictable
  - less reliable

EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Outline

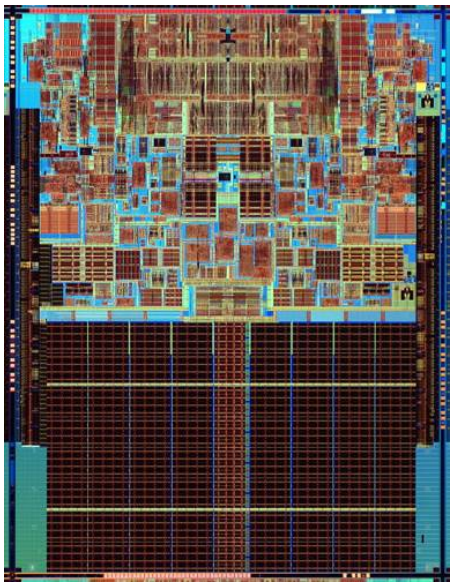
- 63 years of progress
- Many cores make light work
- Building brains
- The *SpiNNaker* project
- The networking challenge
- A generic neural modelling platform
- Plans & conclusions

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Multi-core CPUs

- High-end uniprocessors
  - diminishing returns from complexity
  - wire vs transistor delays
- Multi-core processors
  - cut-and-paste
  - *simple* way to deliver more MIPS
- Moore's Law
  - more transistors
  - more cores

… but what about the software?



SpiNNaker
Biologically Inspired Massively Parallel Architectures

Back to the future

- Imagine…
  - a limitless supply of (free) processors
  - load-balancing is irrelevant
  - all that matters is:
    - the energy used to perform a computation
    - formulating the problem to avoid synchronisation
    - abandoning determinism
- How might such systems work?

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Outline

- 63 years of progress
- Many cores make light work
- Building brains
- The *SpiNNaker* project
- The networking challenge
- A generic neural modelling platform
- Plans & conclusions

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Bio-inspiration

- How can massively parallel computing resources accelerate our understanding of brain function?
- How can our growing understanding of brain function point the way to more efficient parallel, fault-tolerant computation?

SpiNNaker
Biologically Inspired Massively Parallel Architectures

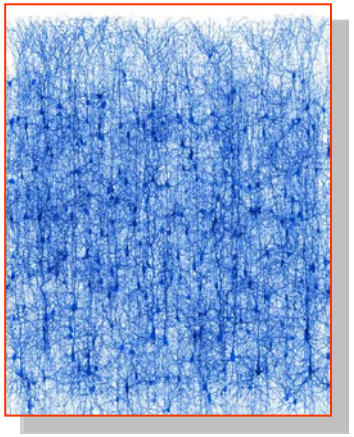
Building brains

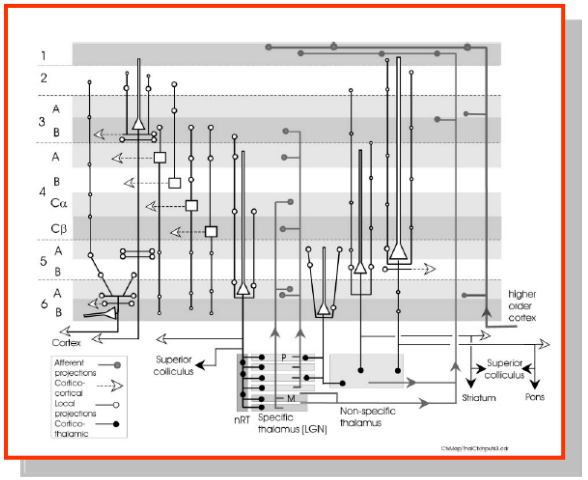
- Brains demonstrate
  - massive parallelism (10^{11} neurons)
  - massive connectivity (10^{15} synapses)
  - excellent power-efficiency
    - much better than today's microchips
  - low-performance components (~ 100 Hz)
  - low-speed communication (~ metres/sec)
  - adaptivity – tolerant of component failure
  - autonomous learning

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Building brains

- Neurons
  - multiple inputs, single output (c.f. logic gate)
  - useful across multiple scales (10^2 to 10^{11})
- Brain structure
  - regularity
  - e.g. 6-layer cortical 'microarchitecture'



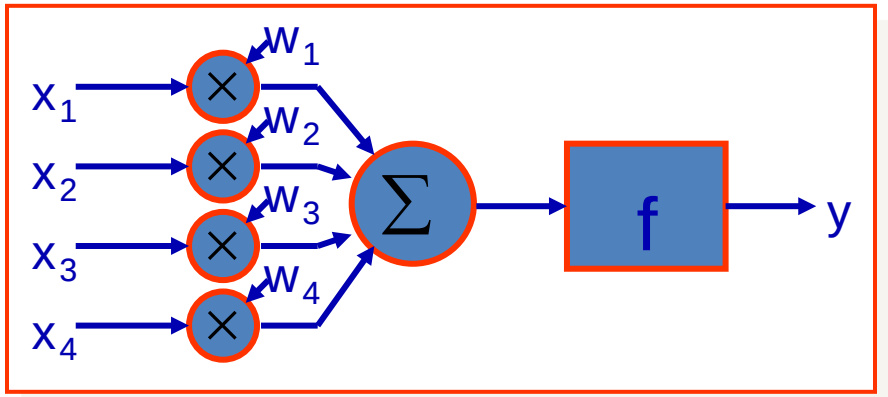


EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

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



SpiNNaker
Biologically Inspired Massively Parallel Architectures

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, …?

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

$$I = \sum_i w_i x_i$$

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

$$if\ A > \vartheta_A\ fire$$

$$\&\ set\ A = 0$$

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Processing

- Izhikevich model
  - two variables, one fast, one slow:

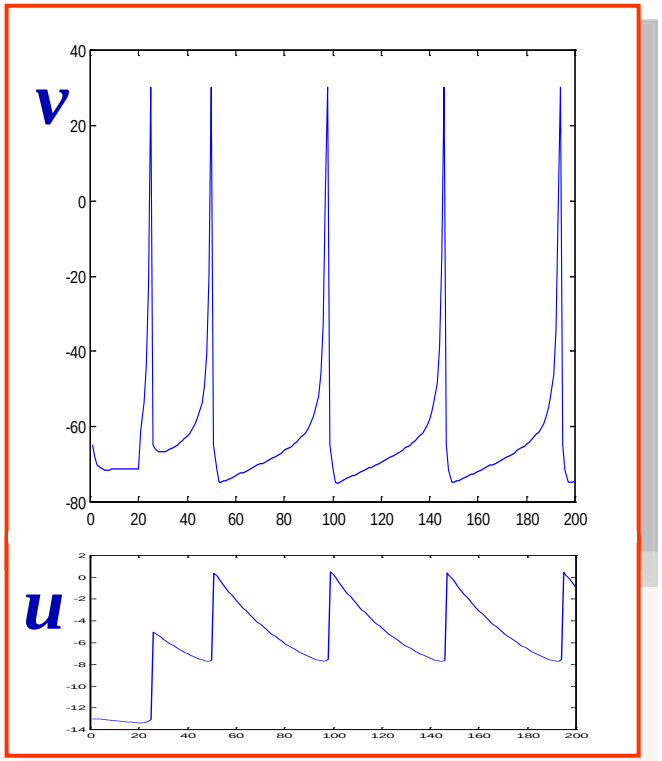
$$\dot{v} = 0.04v^2 + 5v + 140 - u + I$$
$$\dot{u} = a \cdot (bv - u)$$

  - neuron fires when

    $v > 30$; then:

$$v = c$$
$$u = u + d$$

  - a, b, c & d select behaviour

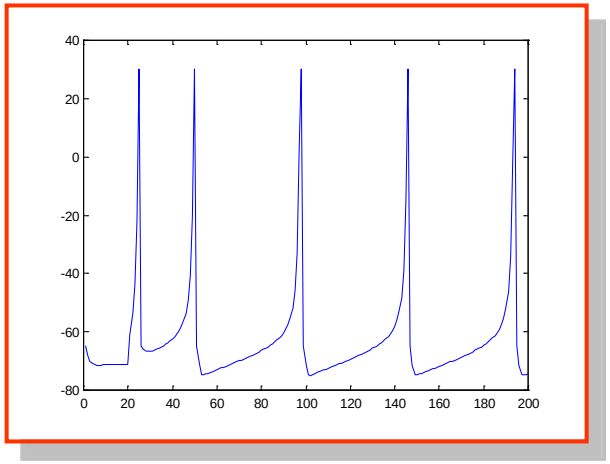


(www.izhikevich.com)

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Communication

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



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

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

SpiNNaker
Biologically Inspired Massively Parallel Architectures

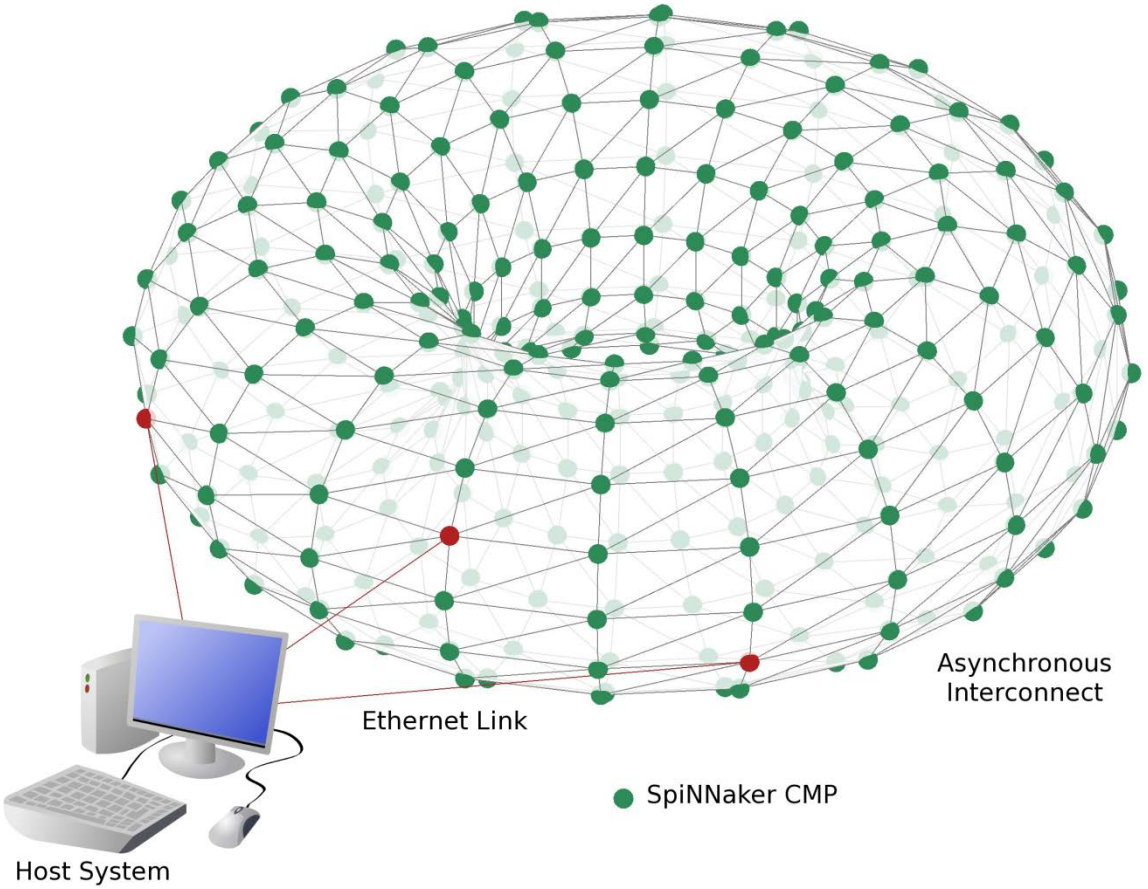
Outline

- 63 years of progress
- Many cores make light work
- Building brains
- The *SpiNNaker* project
- The networking challenge
- A generic neural modelling platform
- Plans & conclusions

SpiNNaker
Biologically
Inspired
Massively
Parallel
Architectures

SpiNNaker project

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



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

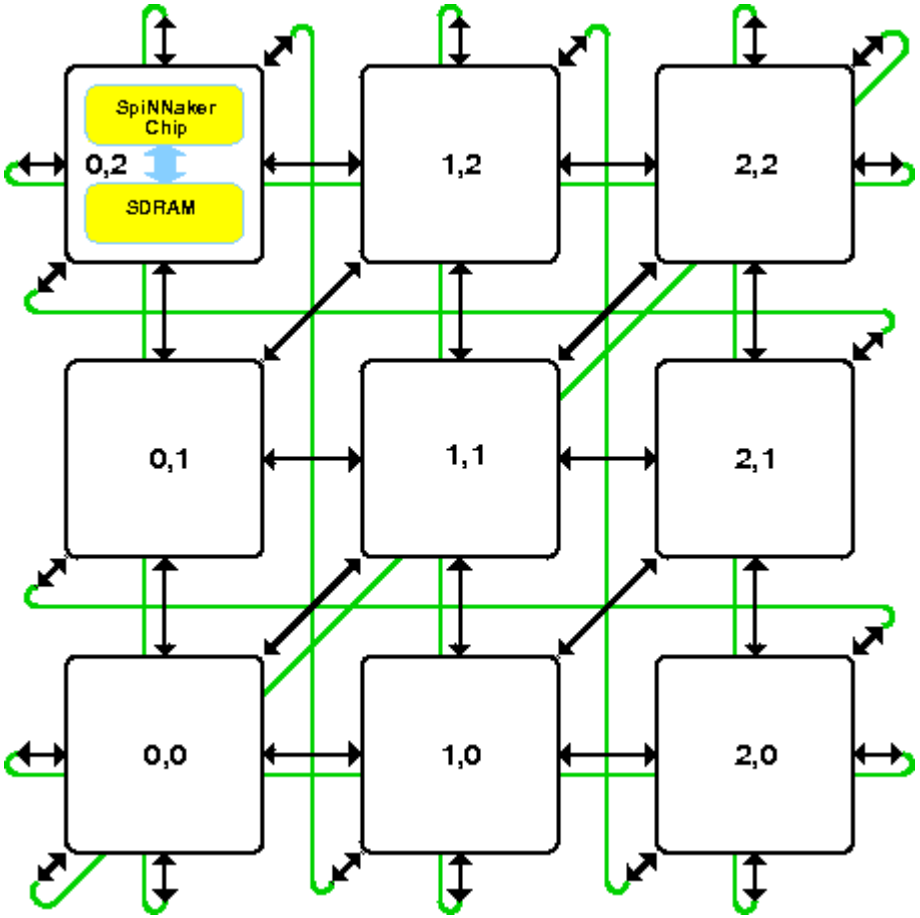
SpiNNaker
Biologically Inspired Massively Parallel Architectures

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
Biologically Inspired Massively Parallel Architectures

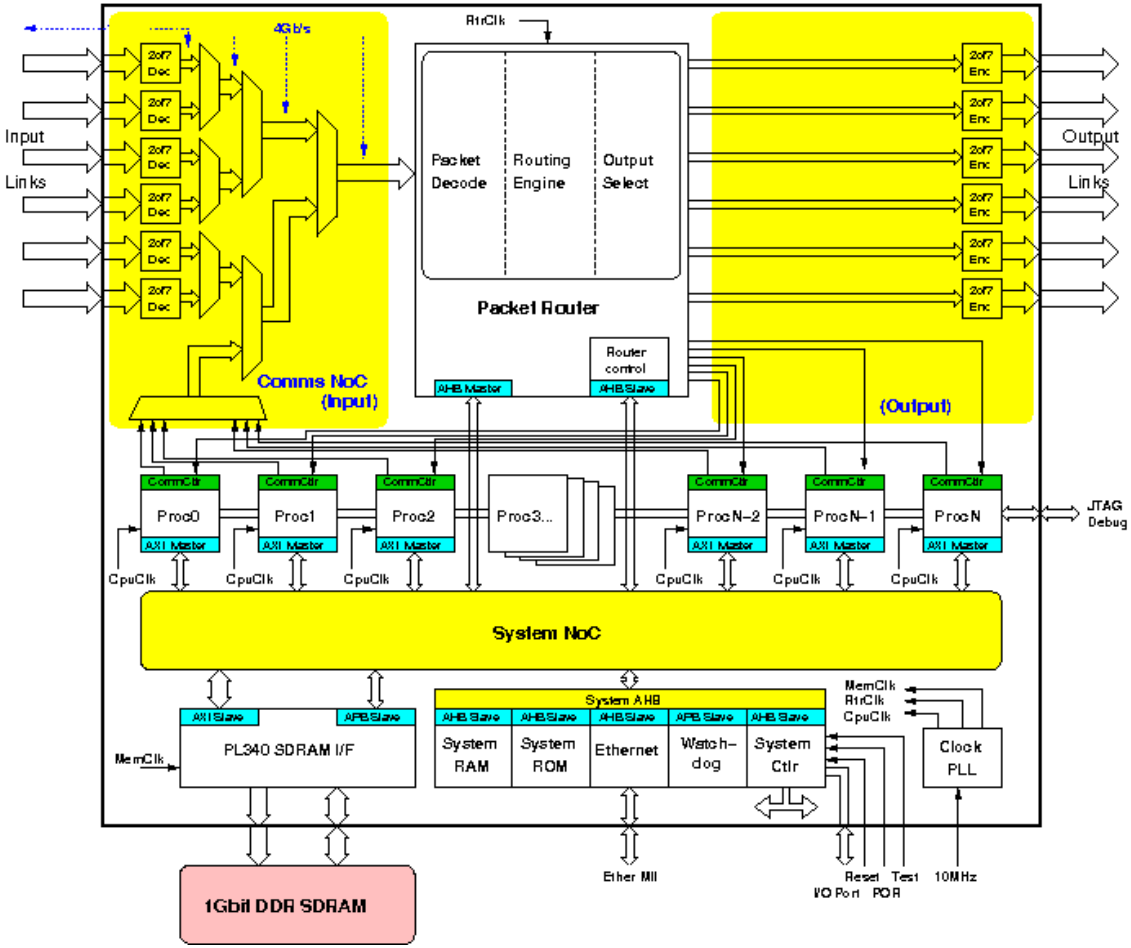
SpiNNaker system



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

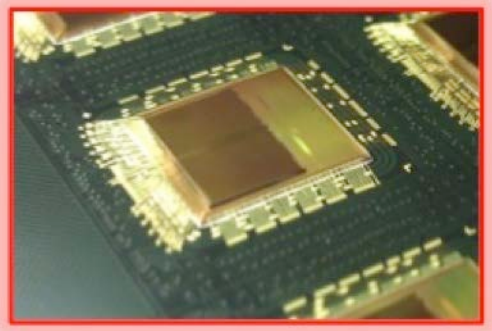
SpiNNaker node



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

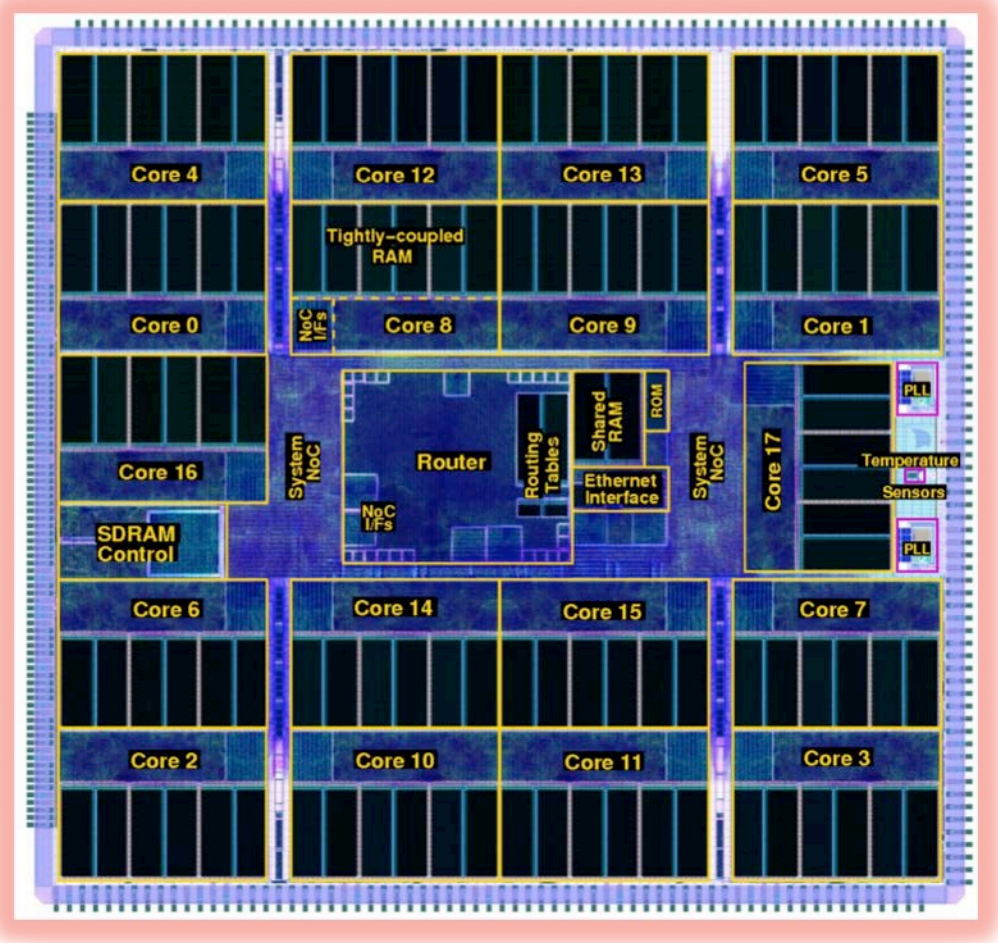
SpiNNaker chip



Mobile DDR SDRAM interface

Multi-chip packaging by UNISEM Europe

unisem



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Outline

- 63 years of progress
- Many cores make light work
- Building brains
- The *SpiNNaker* project
- The networking challenge
- A generic neural modelling platform
- Plans & conclusions

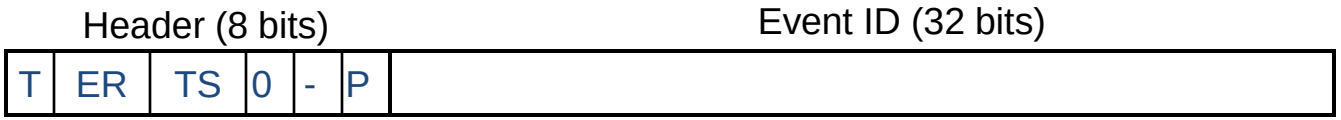
SpiNNaker
Biologically Inspired Massively Parallel Architectures

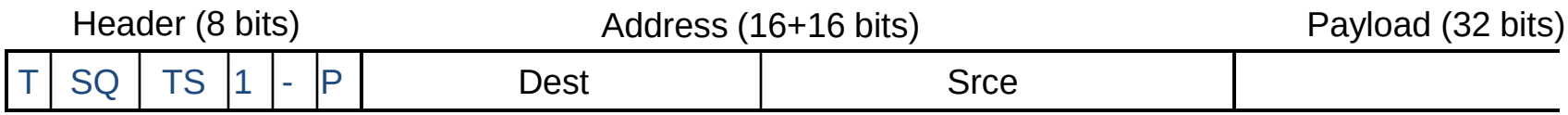
The networking challenge

- Emulate the very high connectivity of real neurons
- A spike generated by a neuron firing must be conveyed efficiently to >1,000 inputs
- On-chip and inter-chip spike communication should use the same delivery mechanism

Network – packets

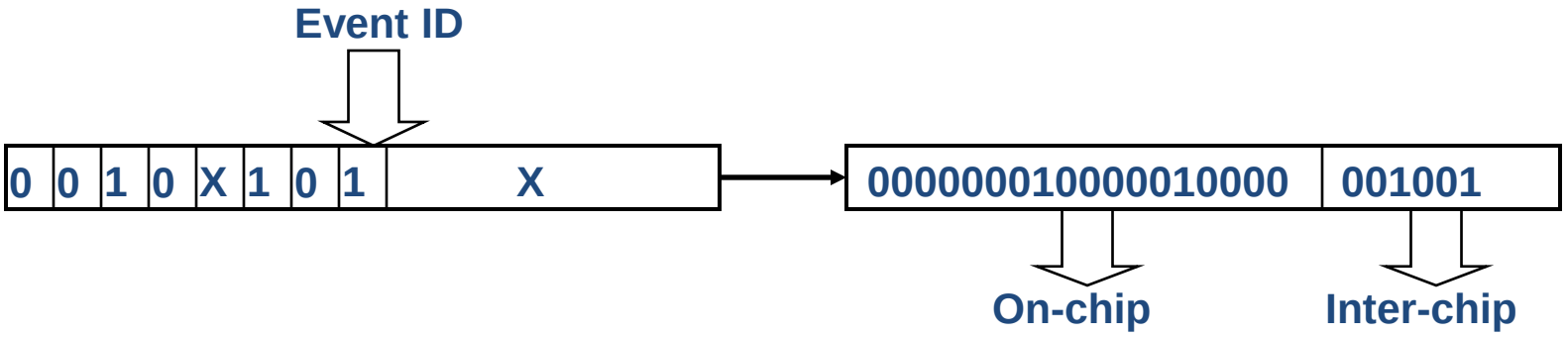
- Four packet types
  - MC (multicast): source routed; carry events (spikes)
  - P2P (point-to-point): used for bootstrap, debug, monitoring, etc
  - NN (nearest neighbour): build address map, flood-fill code
  - FR (fixed route): carry 64-bit debug data to host
- Timestamp mechanism removes errant packets
  - which could otherwise circulate forever





Network – MC Router

- All MC spike event packets are sent to a router
- Ternary CAM keeps router size manageable at 1024 entries (but careful network mapping also essential)
- CAM 'hit' yields a set of destinations for this spike event
  - automatic multicasting
- CAM 'miss' routes event to a 'default' output link

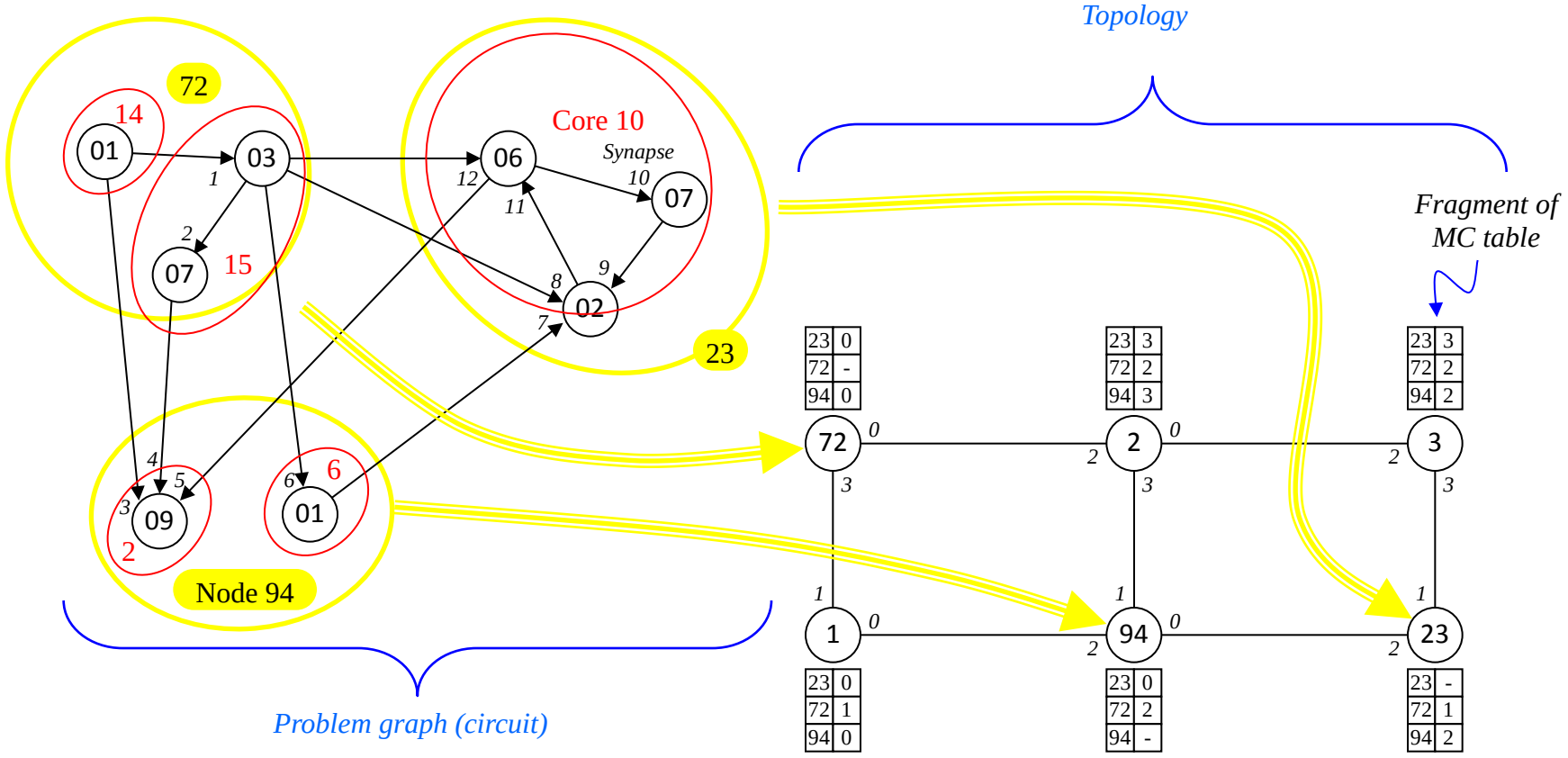


SpiNNaker
Biologically Inspired Massively Parallel Architectures

Outline

- 63 years of progress
- Many cores make light work
- Building brains
- The *SpiNNaker* project
- The networking challenge
- A generic neural modelling platform
- Plans & conclusions

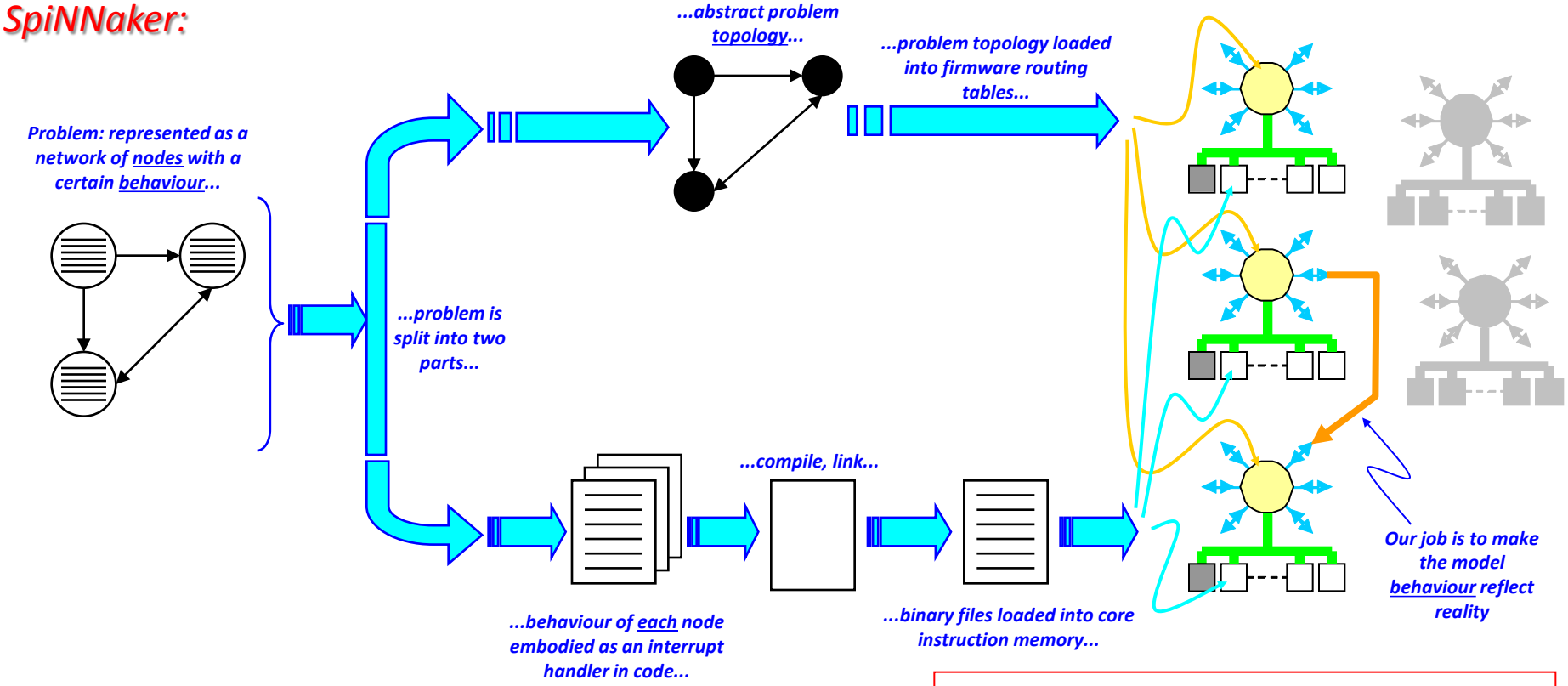
Topology mapping



SpiNNaker
Biologically Inspired Massively Parallel Architectures

Problem mapping

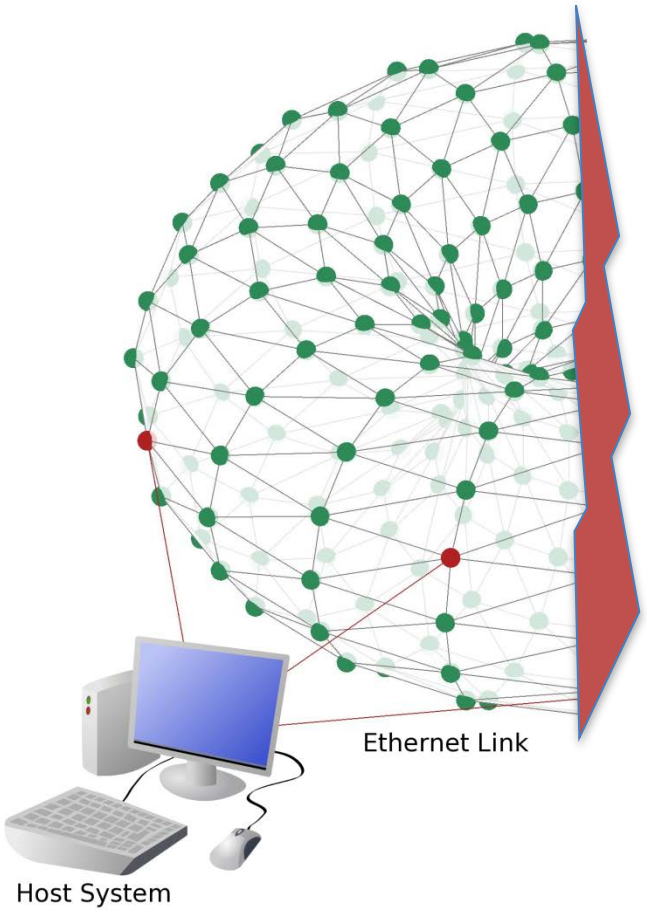
SpiNNaker:



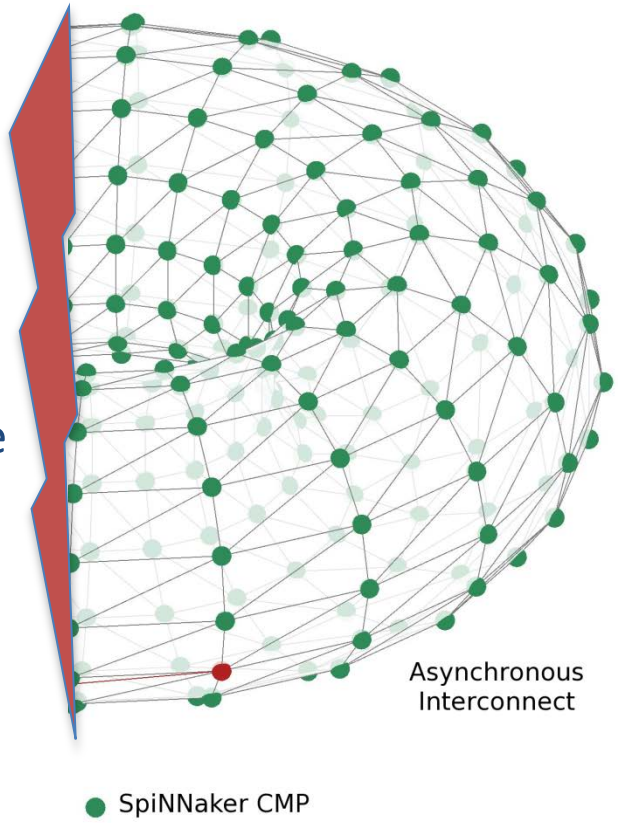
The code says "send message" but has no control where the output message goes

EPSRC · University of Cambridge · Manchester 1824 · The University of Sheffield · University of Southampton · ARM · Silistix · THALES

Bisection performance

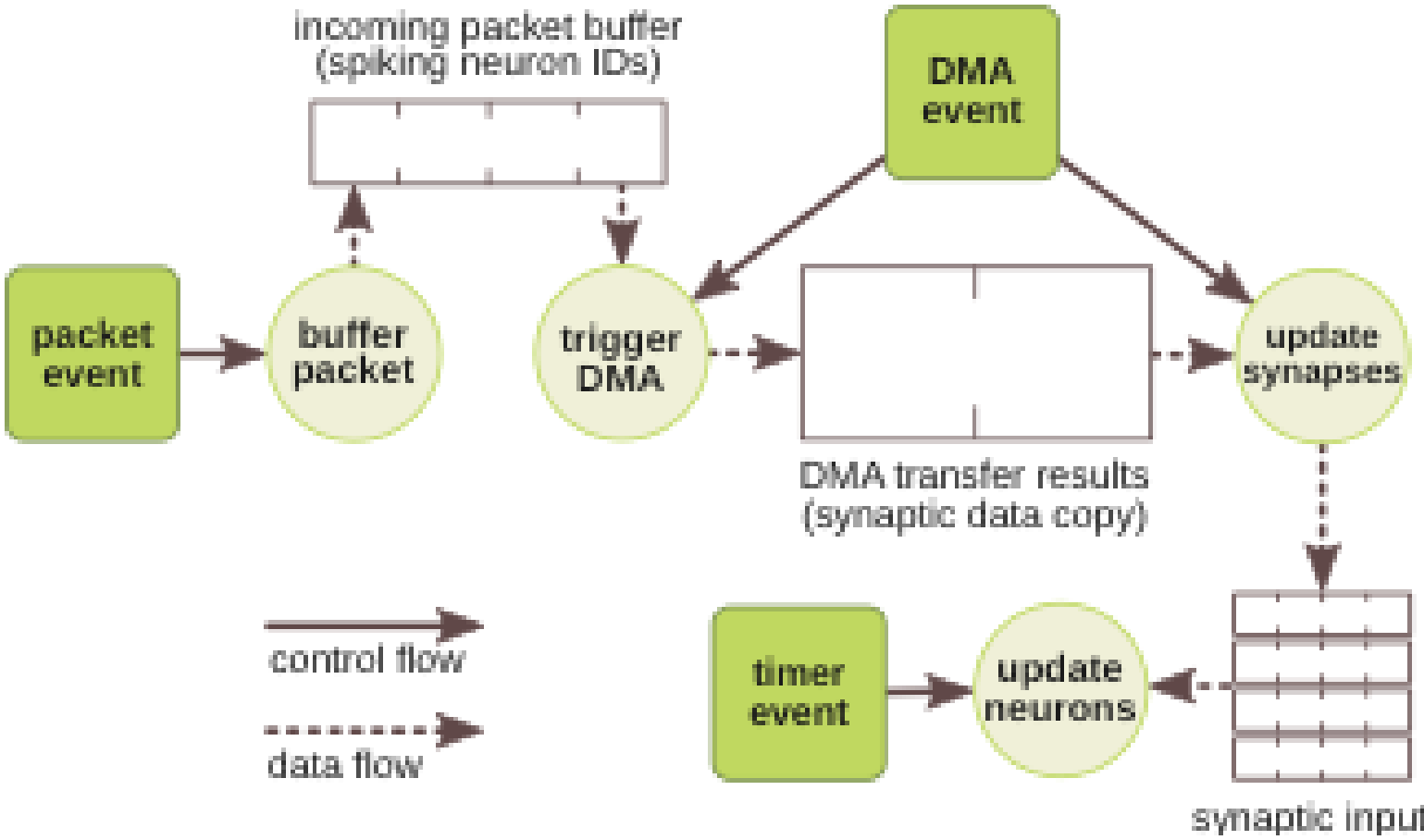


- 1,024 links
  - in each direction
- ~10 billion packets/s
- 10Hz mean firing rate
- 250 Gbps bisection bandwidth



SpiNNaker
Biologically Inspired Massively Parallel Architectures

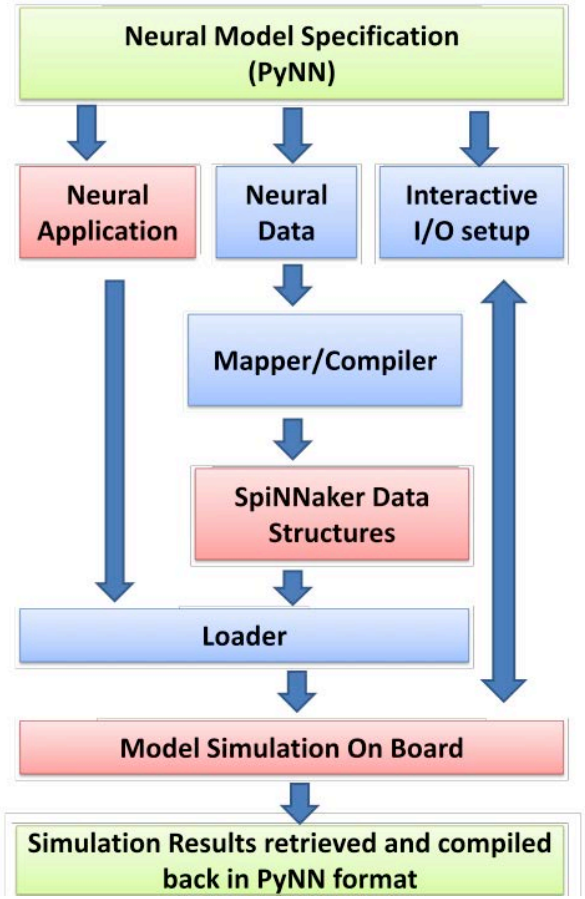
Event-driven software model

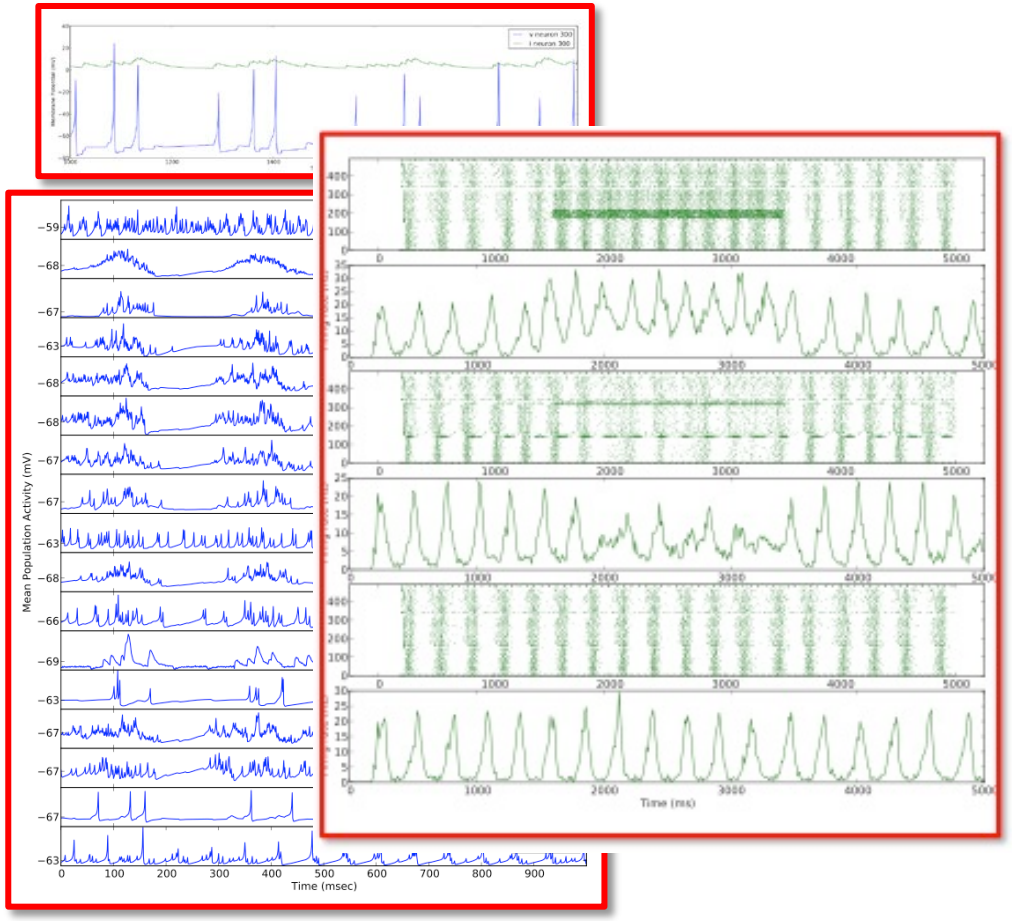


EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

PyNN design flow



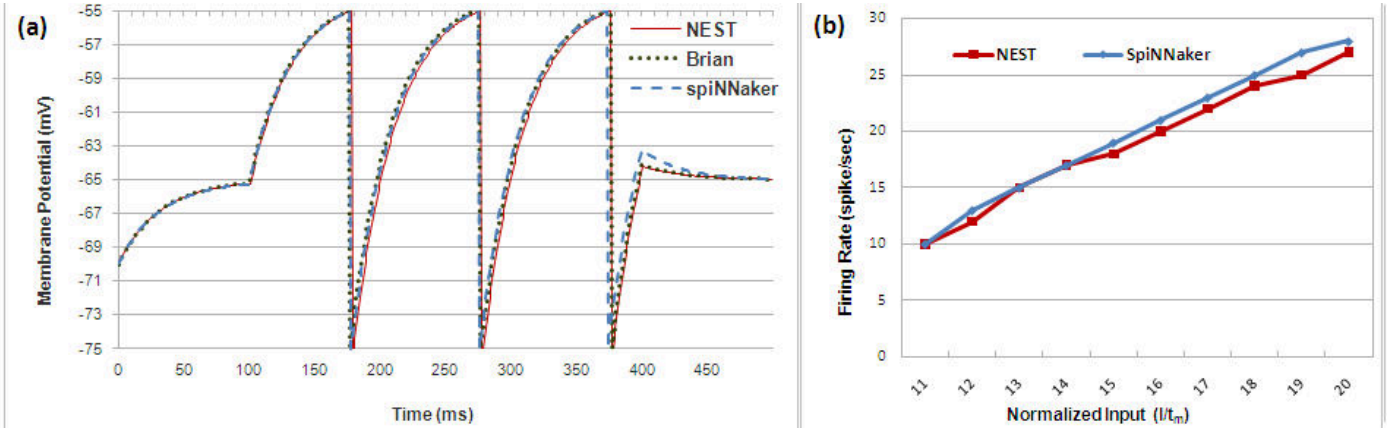


EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

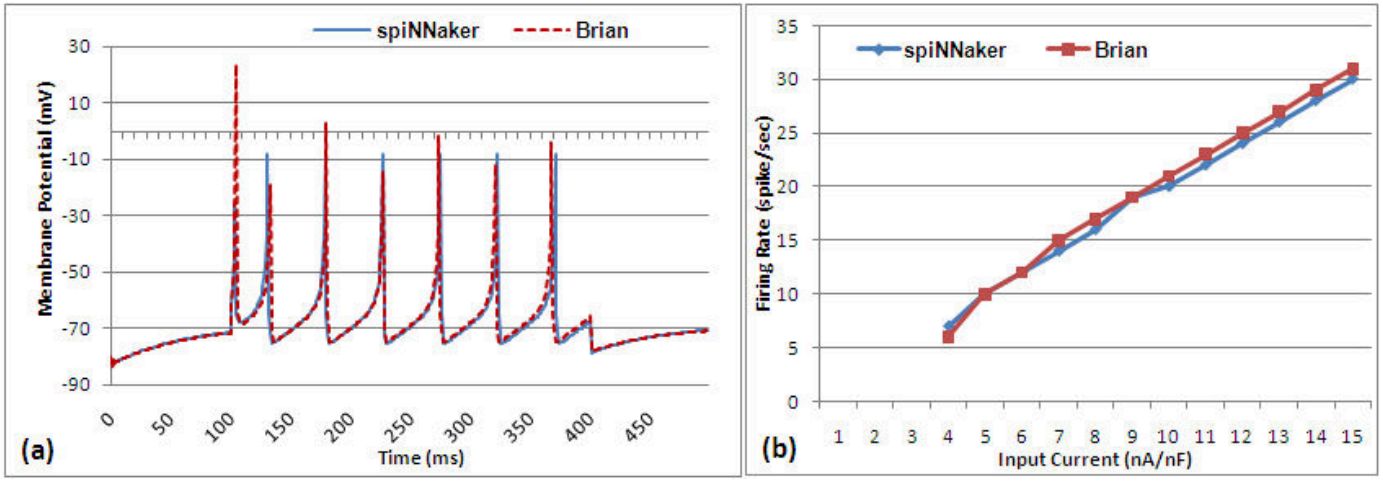
SpiNNaker
Biologically Inspired Massively Parallel Architectures

PyNN integration

- LIF



- Izhikevich

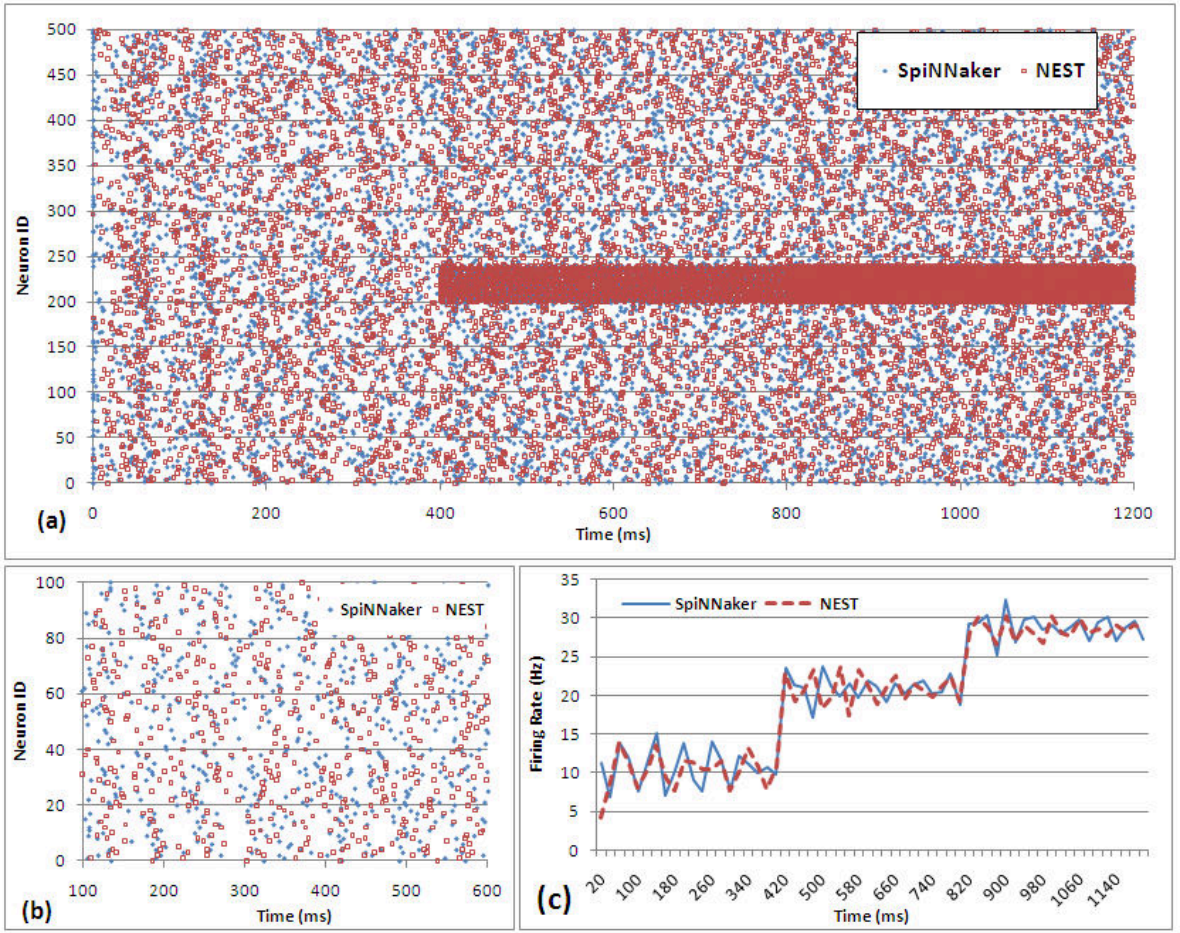


EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

PyNN integration

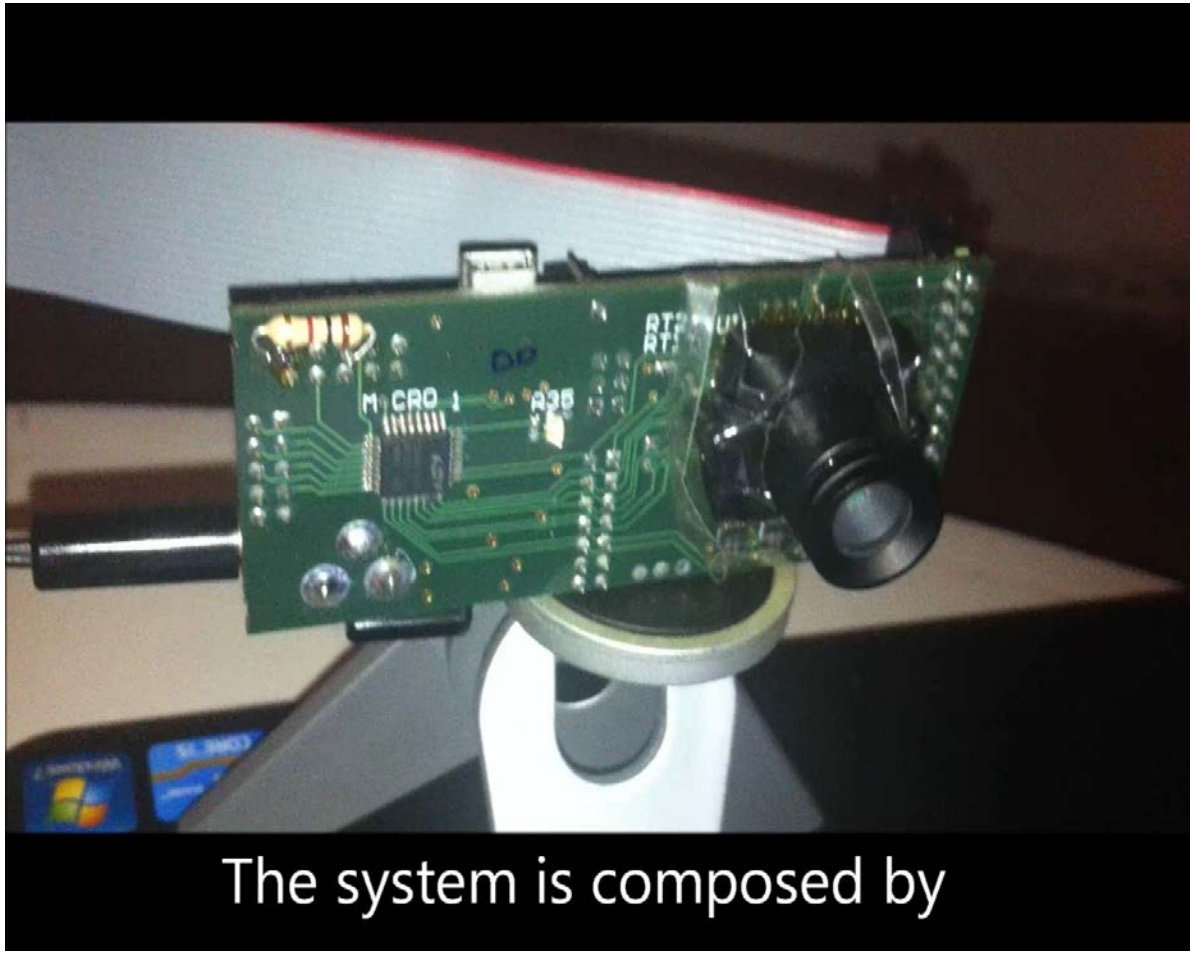
- Vogels-Abbott benchmark
  - 500 LIF neurons



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

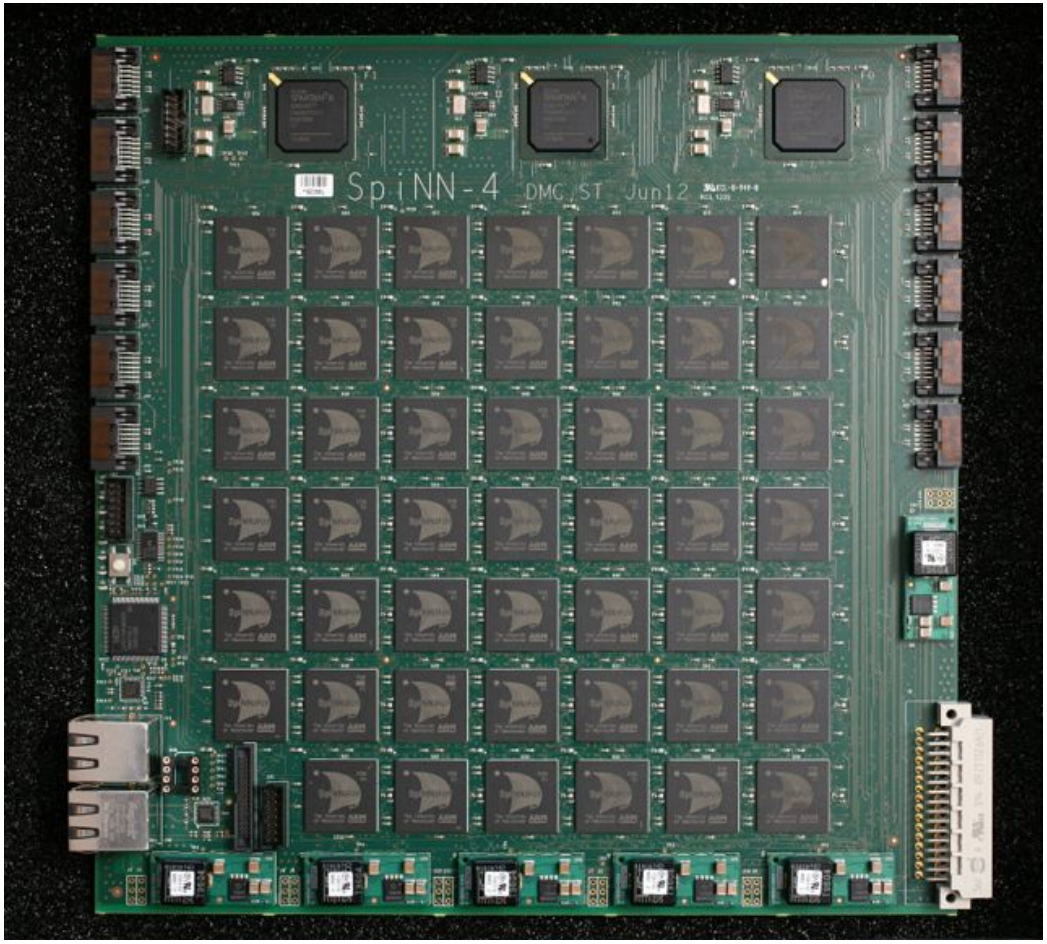
SpiNNaker vision



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

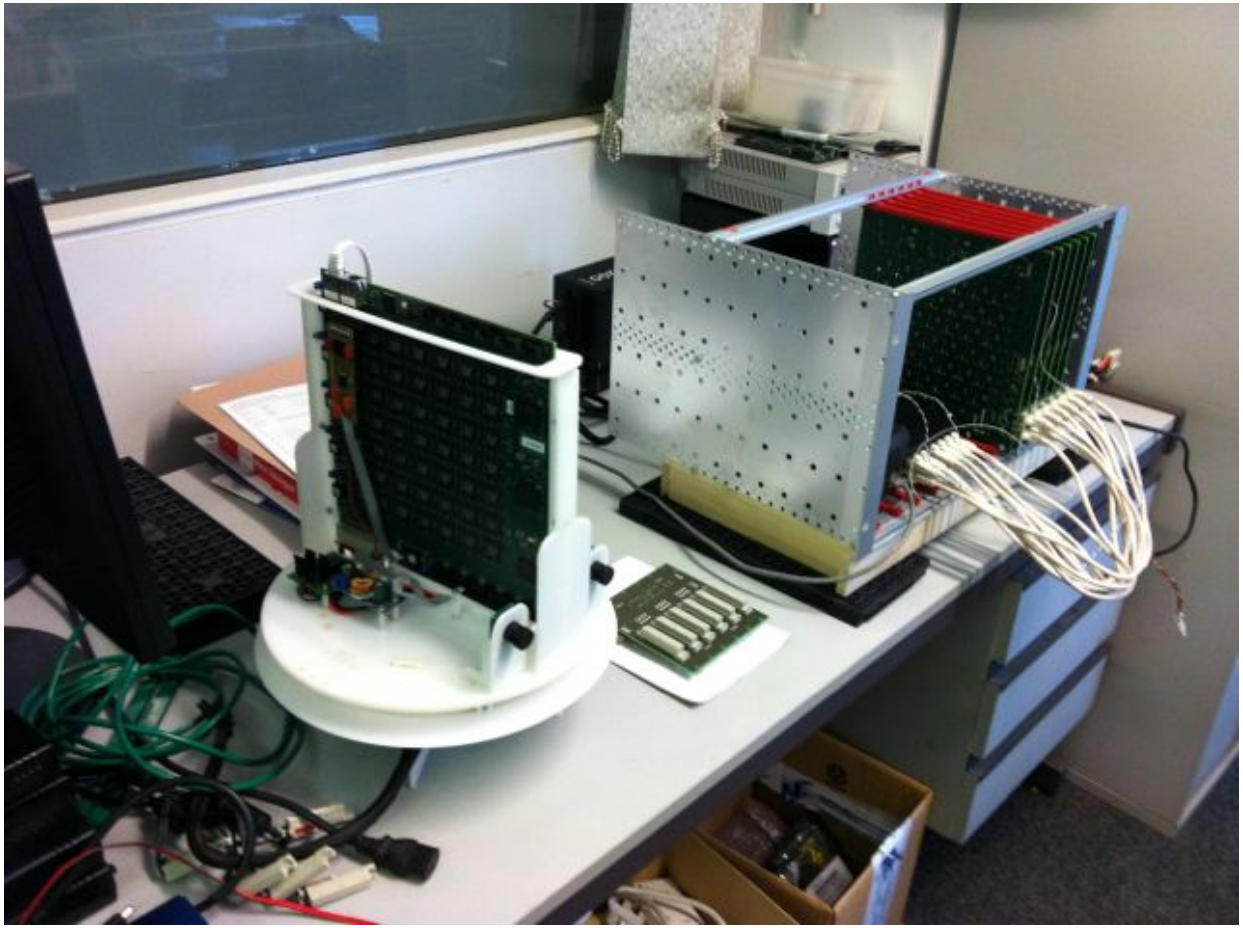
48-node PCB



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

SpiNNaker platforms



SpiNNaker
Biologically Inspired Massively Parallel Architectures

Outline

- 63 years of progress
- Many cores make light work
- Building brains
- The *SpiNNaker* project
- The networking challenge
- A generic neural modelling platform
- Plans & conclusions

SpiNNaker
Biologically Inspired Massively Parallel Architectures

SpiNNaker machines

103 machine: 864 cores, 1 PCB, 75W

104 machine:10,368 cores, 1 rack, 900W
(NB 12 PCBs for operation without aircon)

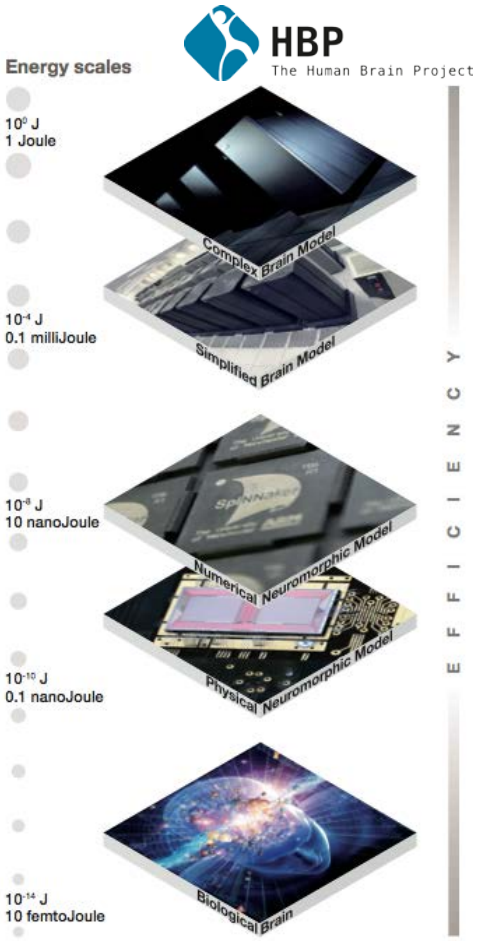
105 machine: 103,680 cores, 1 cabinet, 9kW

106 machine: 1M cores, 10 cabinets, 90kW

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Conclusions

- Brains represent a significant computational challenge
  - now coming within range?
- *SpiNNaker* is driven by the brain modelling objective
  - virtualised topology, bounded asynchrony, energy frugality
- The major architectural innovation is the multicast communications infrastructure
- We have working hardware
  - 48-node 864-ARM PCBs now
  - first multi-PCB systems now working



EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

SpiNNaker
Biologically Inspired Massively Parallel Architectures

SpiNNaker team

Manchester

Southampton

EPSRC · University of Cambridge · Manchester 1824 · The University of Sheffield · University of Southampton · ARM · Silistix · THALES