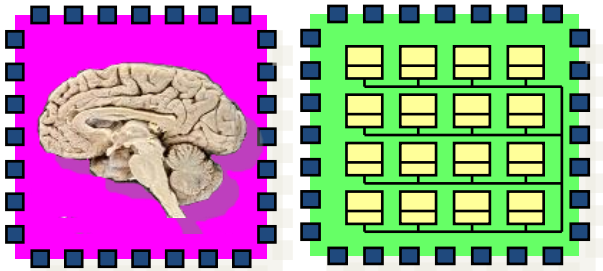
SpiNNaker
Biologically Inspired Massively Parallel Architectures



Biologically-Inspired Massively-Parallel Architectures

- computing beyond a million processors

Dave Lester
The University of Manchester
d.lester@manchester.ac.uk

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

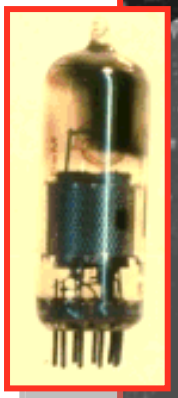
SpiNNaker
Biologically Inspired Massively Parallel Architectures

Outline

- 60 years of progress
- Computer Architecture Perspective
- Building Brains
- The ***SpiNNaker*** system
- A generic neural modelling platform
- Conclusions

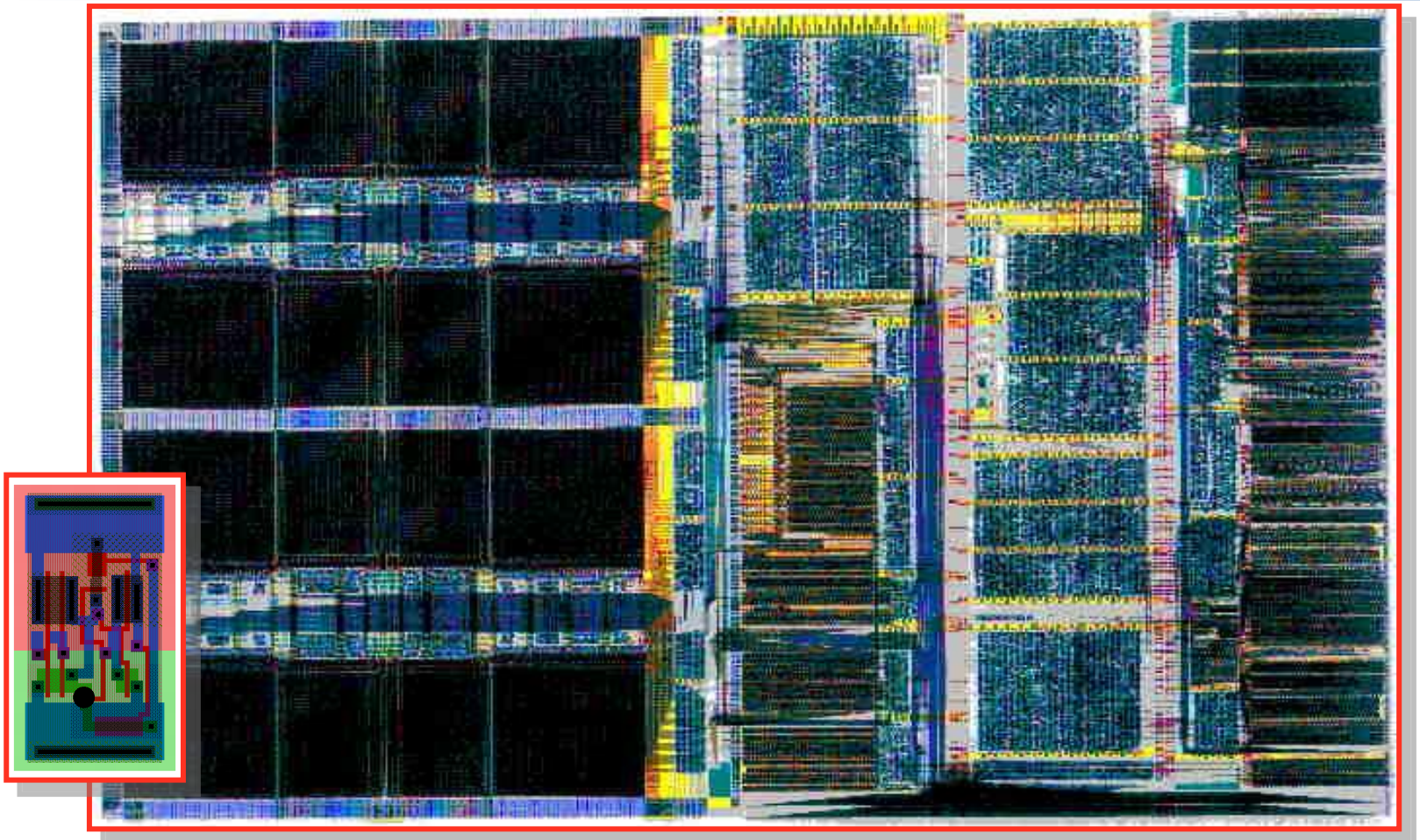
SpiNNaker
Biologically Inspired Massively Parallel Architectures

Baby (1948)



SpiNNaker
Biologically Inspired Massively Parallel Architectures

ARM9 (2005)

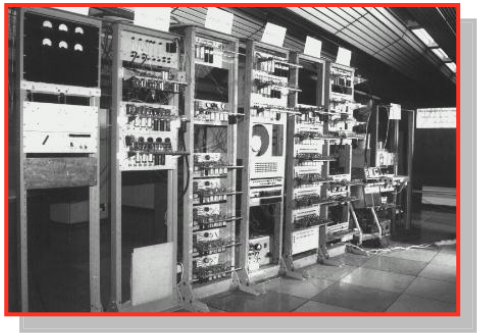


EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

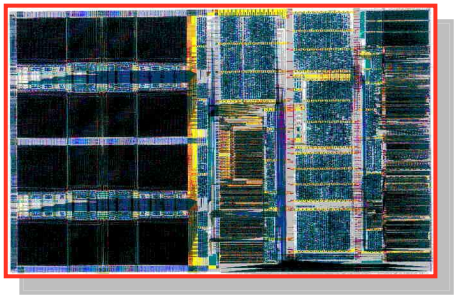
SpiNNaker
Biologically Inspired Massively Parallel Architectures

60 years of progress

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



- ***ARM968:***
  - fills 0.4mm^2 of silicon (130nm)
  - uses 20 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

Energy efficiency

- Baby:
  - 5 Joules per instruction
- ARM968:
  - 0.000 000 000 1 Joules per instruction

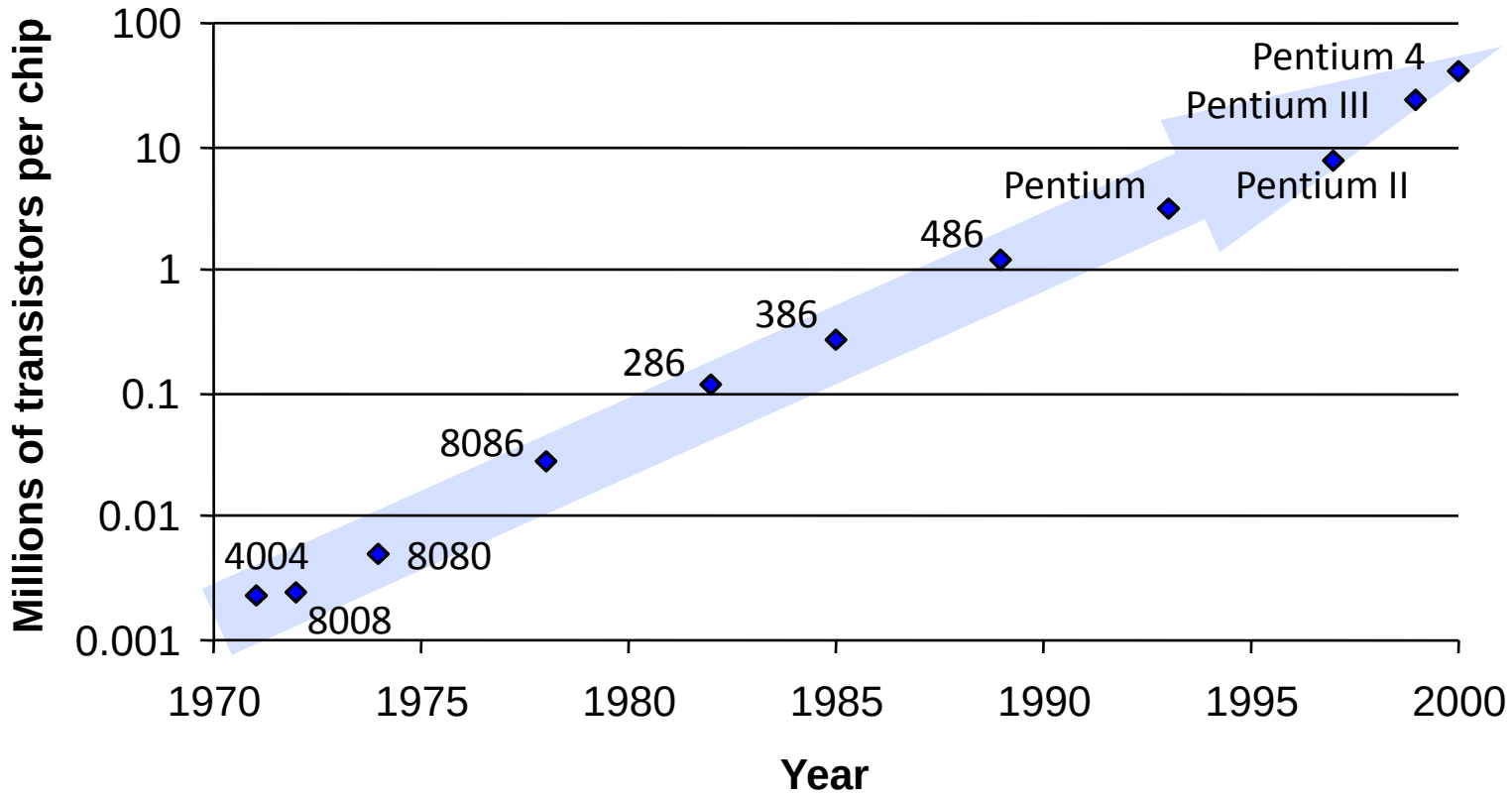
50,000,000,000 times better than Baby!

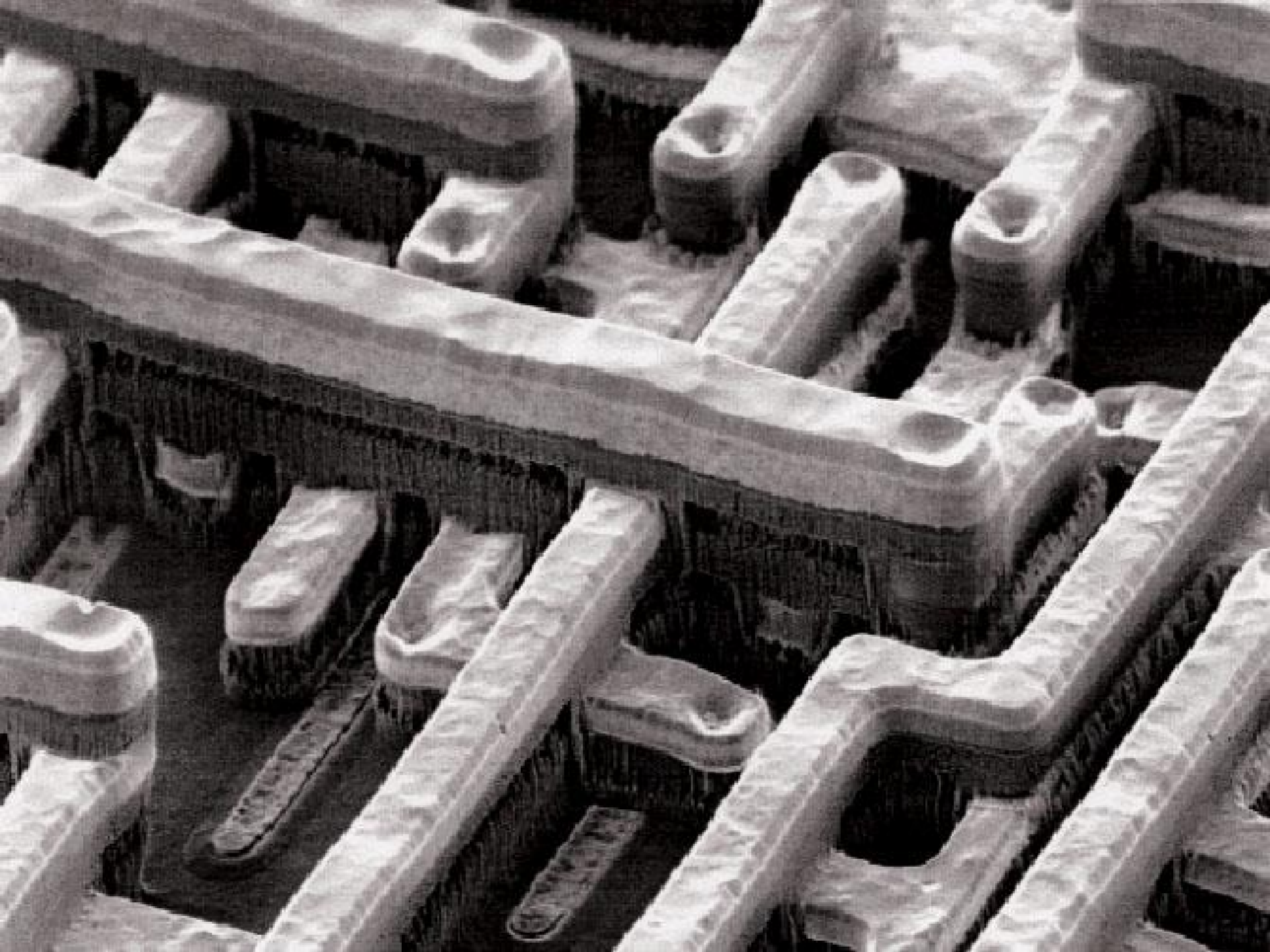
(James Prescott Joule born Salford, 1818)

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Moore's Law

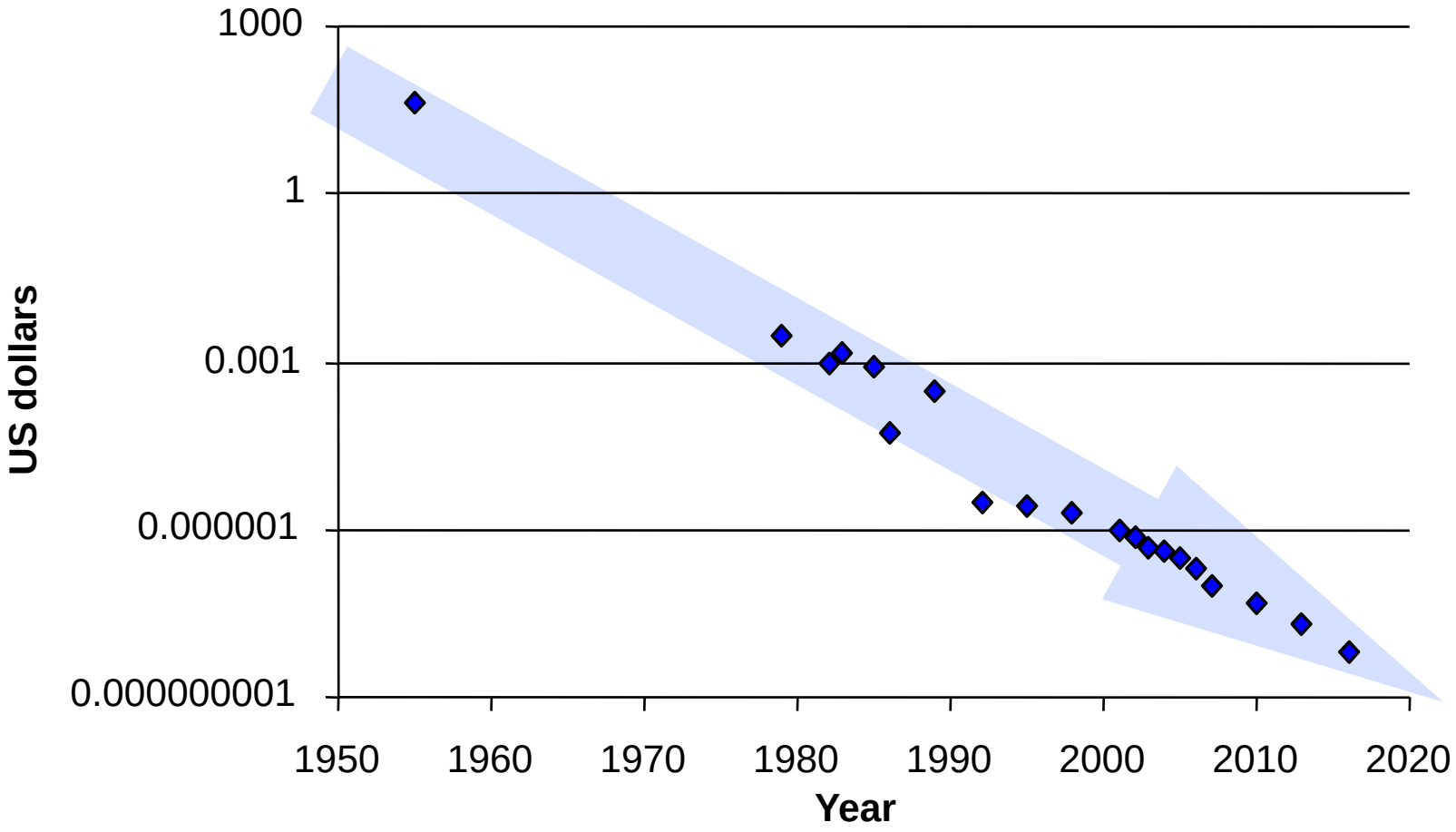
Transistors per Intel chip





SpiNNaker
Biologically Inspired Massively Parallel Architectures

Cost of a Transistor



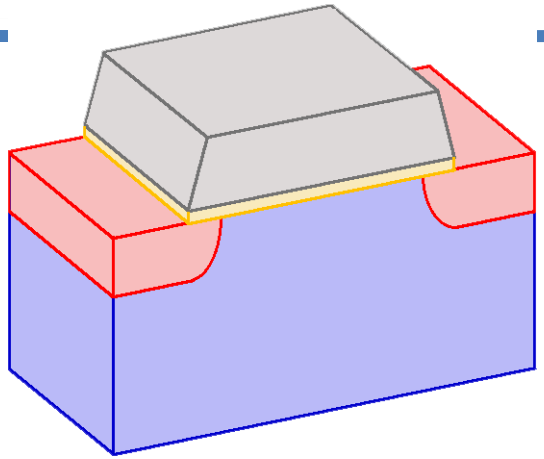
EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

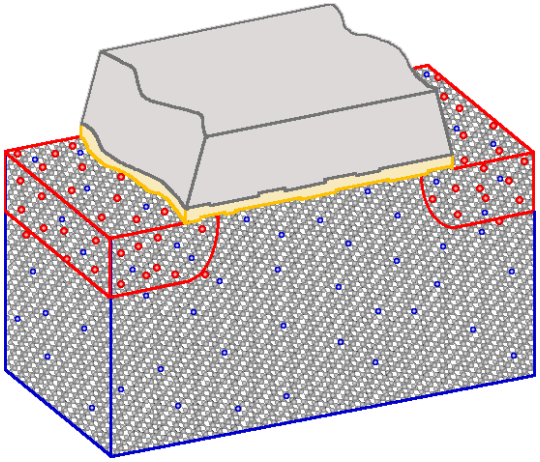
Moore's Law

- SanDisk 12GB microSD
  - 50 billion transistors
    - for £20!

SpiNNaker
Biologically Inspired Massively Parallel Architectures

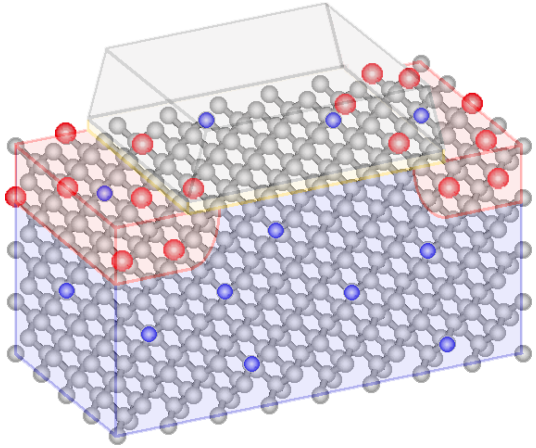
...the Bad News





DEVICE MODELLING GROUP

UNIVERSITY of GLASGOW

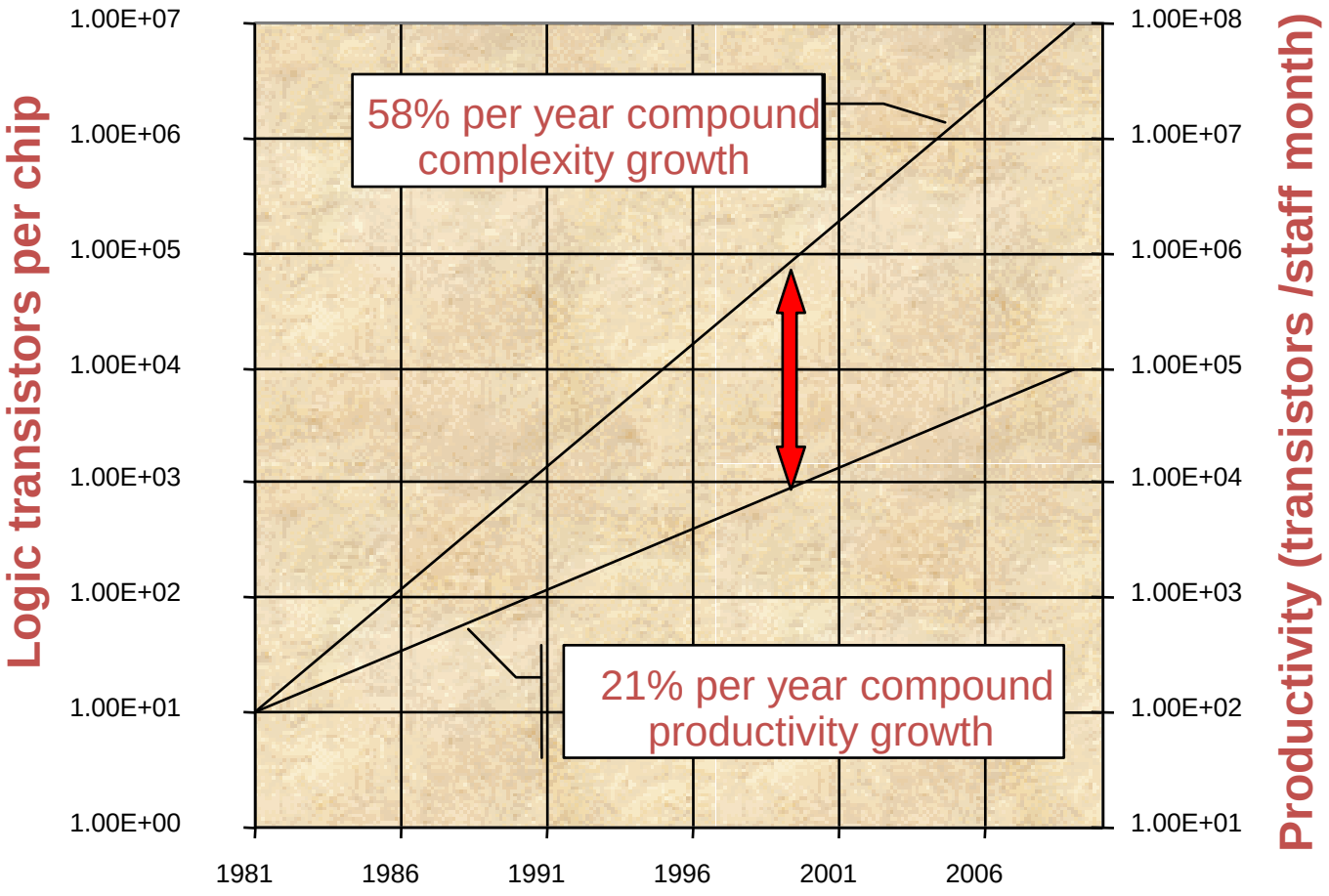


- atomic scales
  - less predictable
  - less reliable

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

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Cost of design



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

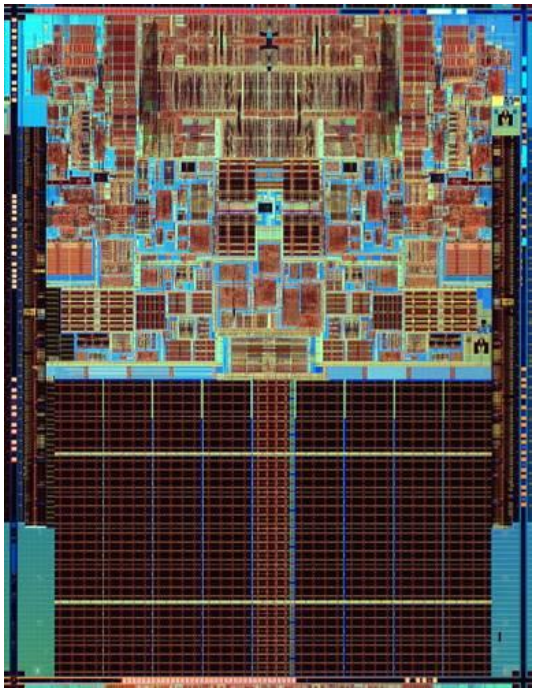
SpiNNaker
Biologically Inspired Massively Parallel Architectures

Outline

- 60 years of progress
- Computer Architecture Perspective
- Building Brains
- The ***SpiNNaker*** system
- A generic neural modelling platform
- Conclusions

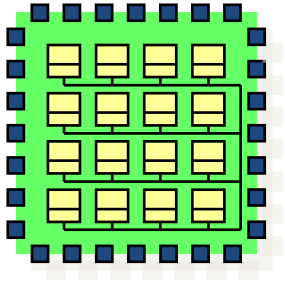
Multi-core CPUs

- 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?

Multi-core CPUS



- General-purpose parallelization
  - an unsolved problem
  - the ‘Holy Grail’ of computer science for half a century?
  - but imperative in the many-core world
- Once solved
  - few complex cores, or many simple cores?
  - simple cores win hands-down on power-efficiency!

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

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

Outline

- 60 years of progress
- Computer Architecture Perspective
- Building Brains
- The ***SpiNNaker*** system
- A generic neural modelling platform
- Conclusions

SpiNNaker
Biologically Inspired Massively Parallel Architectures

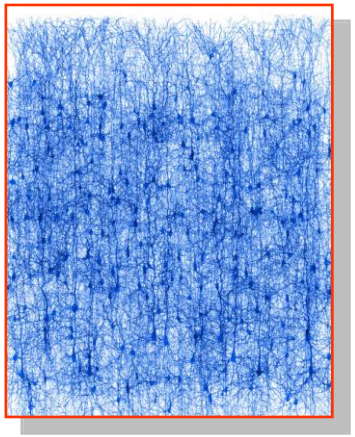
Building brains

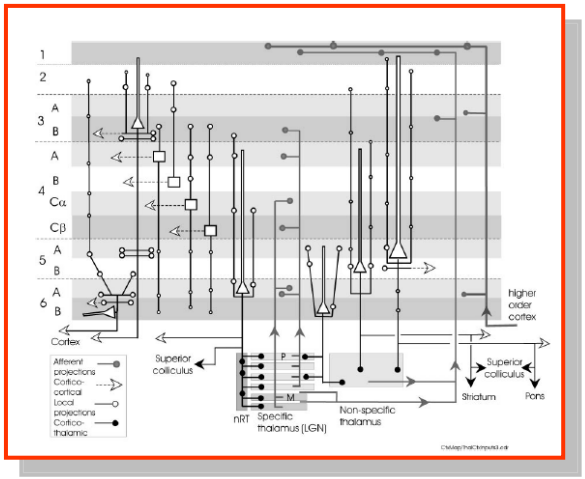
- Brains demonstrate
  - massive parallelism (10^{11} neurons)
  - massive connectivity (10^{15} synapses)
  - excellent power-efficiency
    - much better than today's microchips
  - low-performance components (~ 100 Hz)
  - low-speed communication (~ metres/sec)
  - adaptivity – tolerant of component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

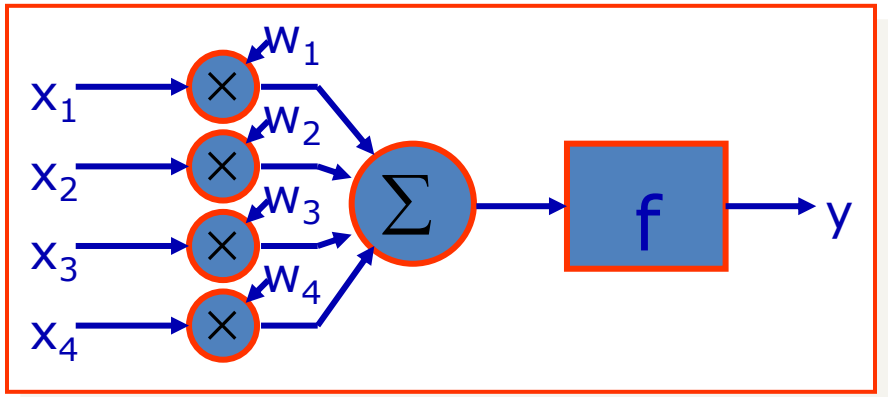




EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

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:

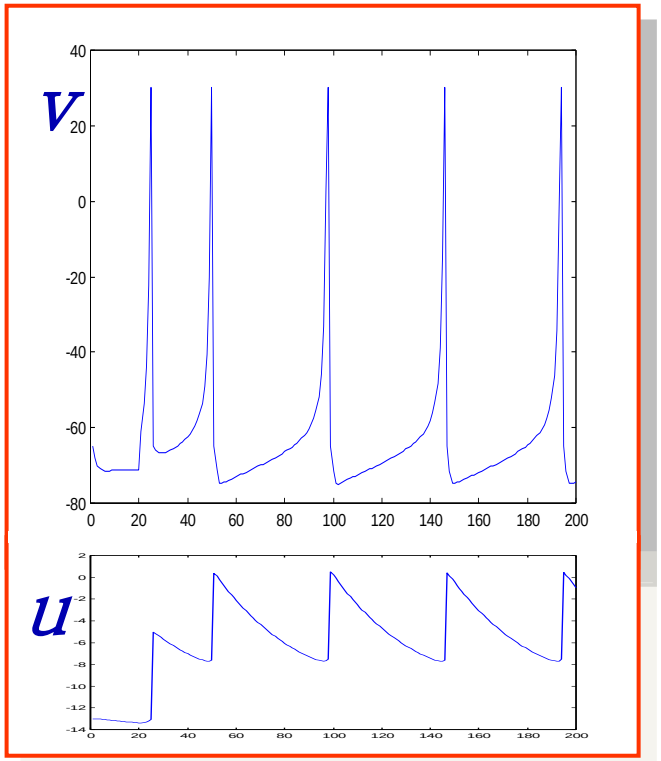
$$\dot{v} = 0.04v^2 + 5v + 140 - u + I$$
$$\dot{u} = a \cdot (bv - u)$$

  - neuron fires when

$V > 30$; then:

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

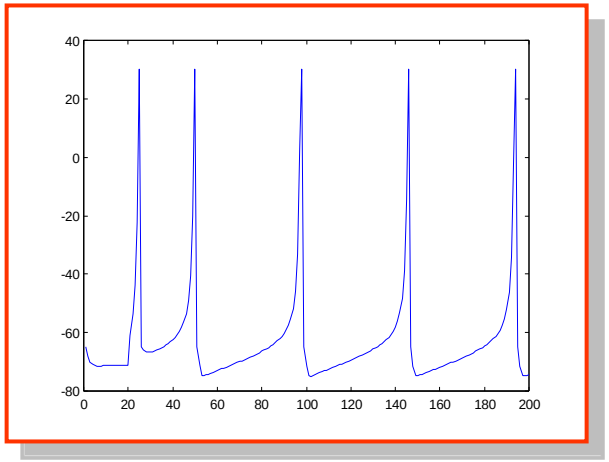
  - a, b, c & d select behaviour



(www.izhikevich.com)

Communication

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



SpiNNaker
Biologically Inspired Massively Parallel Architectures

Storage

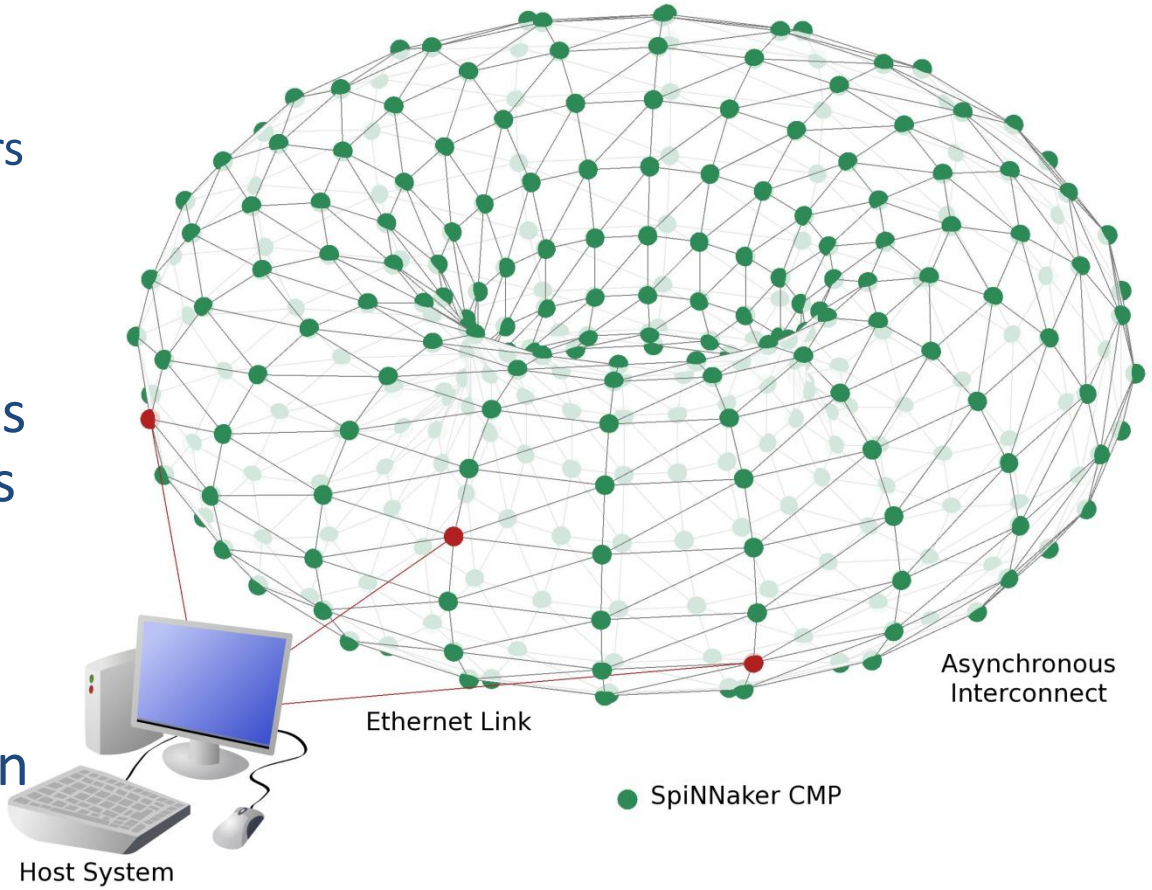
- Synaptic weights
  - stable over long periods of time
    - with diverse decay properties?
  - adaptive, with diverse rules
    - Hebbian, anti-Hebbian, LTP, LTD, …
- Axon ‘delay lines’
- Neuron dynamics
  - multiple time constants
- Dynamic network states

Outline

- 60 years of progress
- Computer Architecture Perspective
- Building Brains
- The *SpiNNaker* system
- A generic neural modelling platform
- Conclusions

SpiNNaker project

- Multi-core CPU node
  - 18 ARM968 processors
  - to model large-scale systems of spiking neurons
- Scalable up to systems with 10,000s of nodes
  - over a million processors
  - >10^8 MIPS total
- Power ~ 25µw/neuron

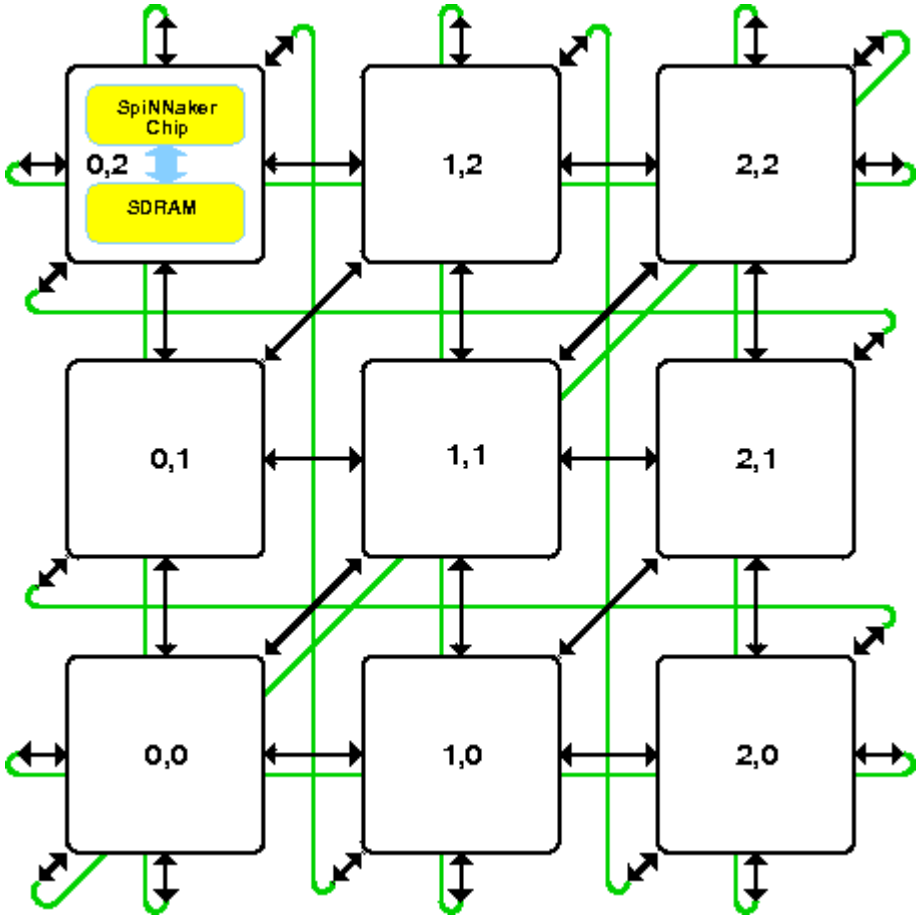


SpiNNaker
Biologically Inspired Massively Parallel Architectures

Design principles

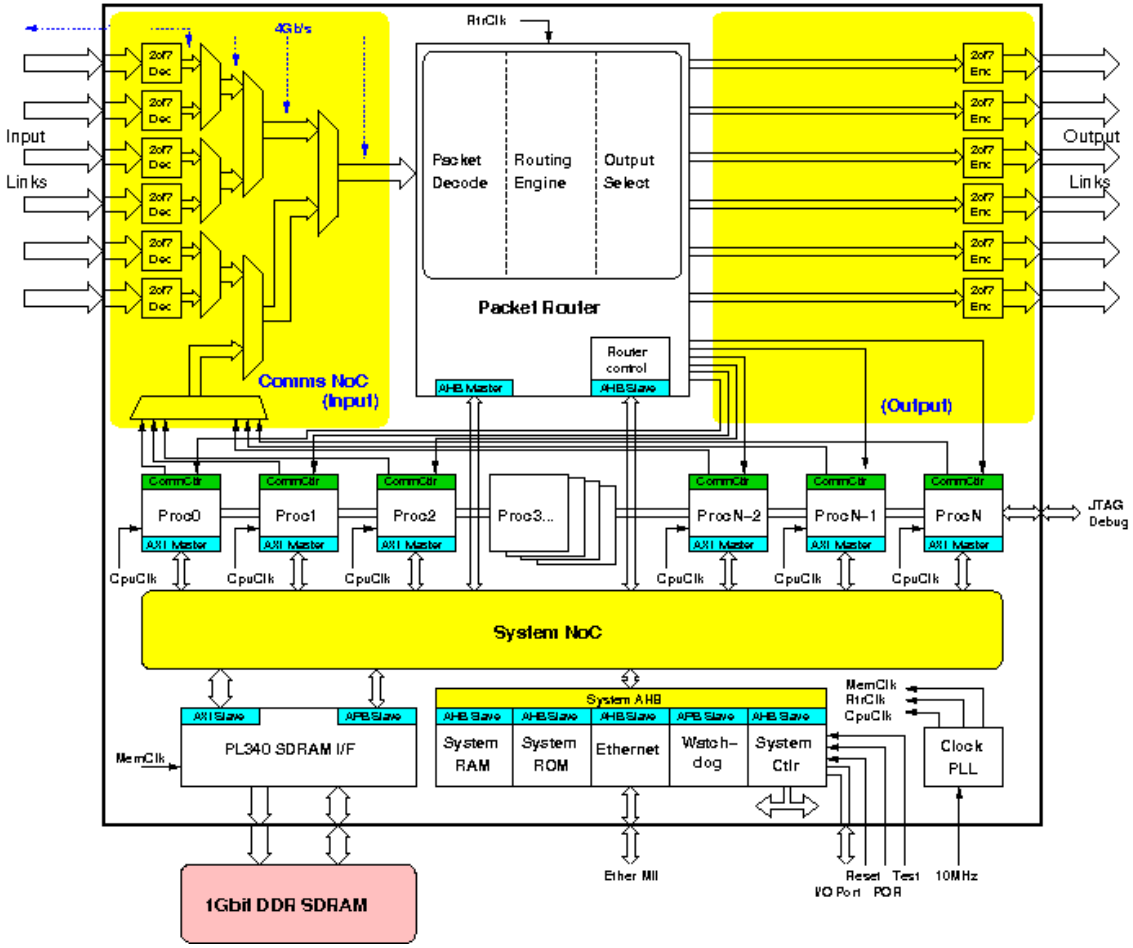
- *Virtualised topology*
  - physical and logical connectivity are decoupled
- *Bounded asynchrony*
  - time models itself
- *Energy frugality*
  - processors are free
  - the real cost of computation is energy

SpiNNaker system



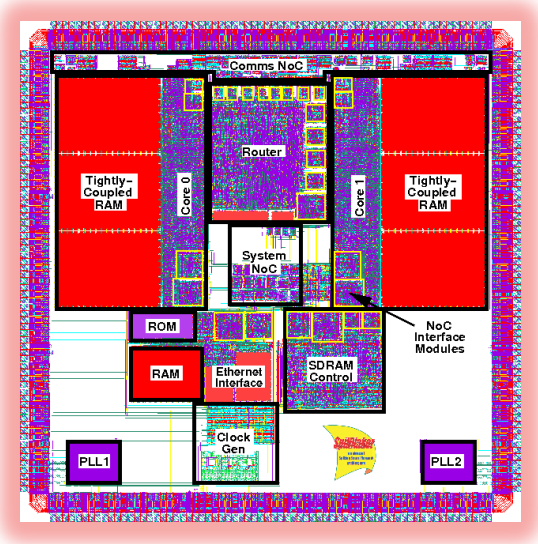
SpiNNaker
Biologically Inspired Massively Parallel Architectures

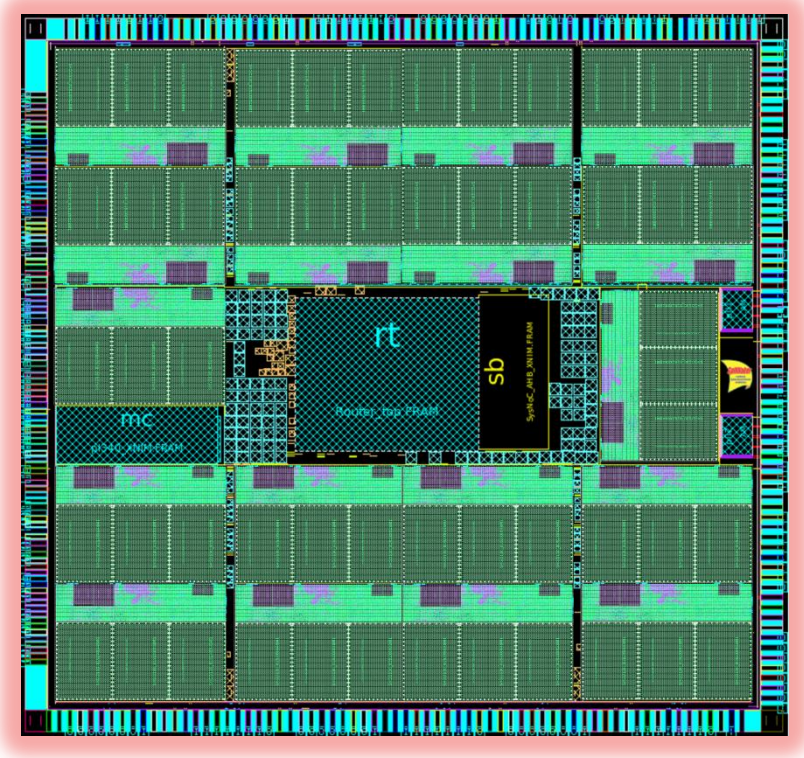
CMP node



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

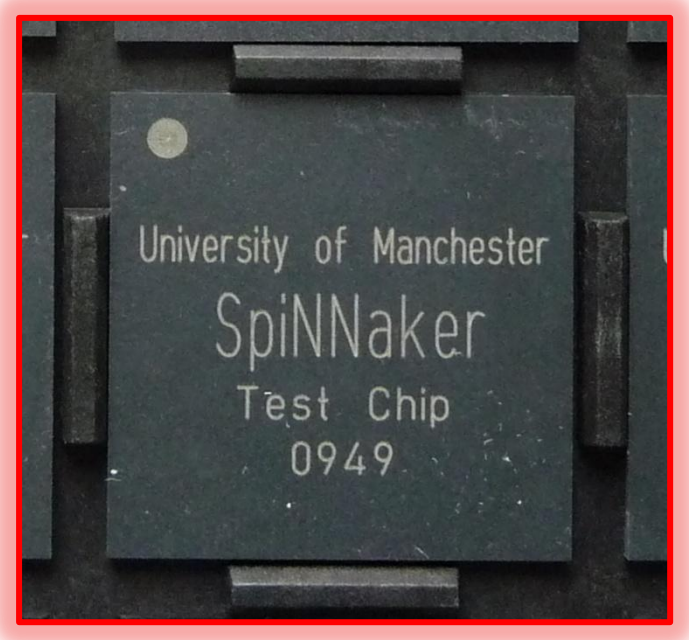
SpiNNaker chips

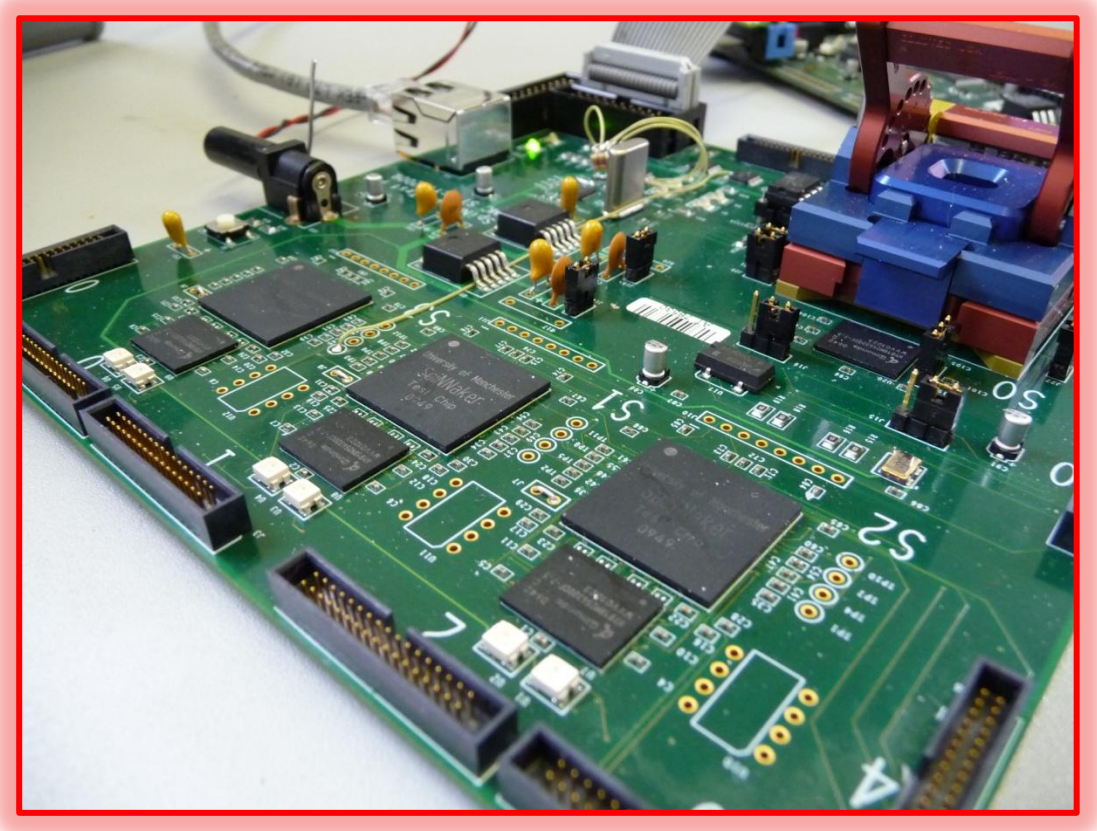




SpiNNaker
Biologically Inspired Massively Parallel Architectures

SpiNNaker test chip



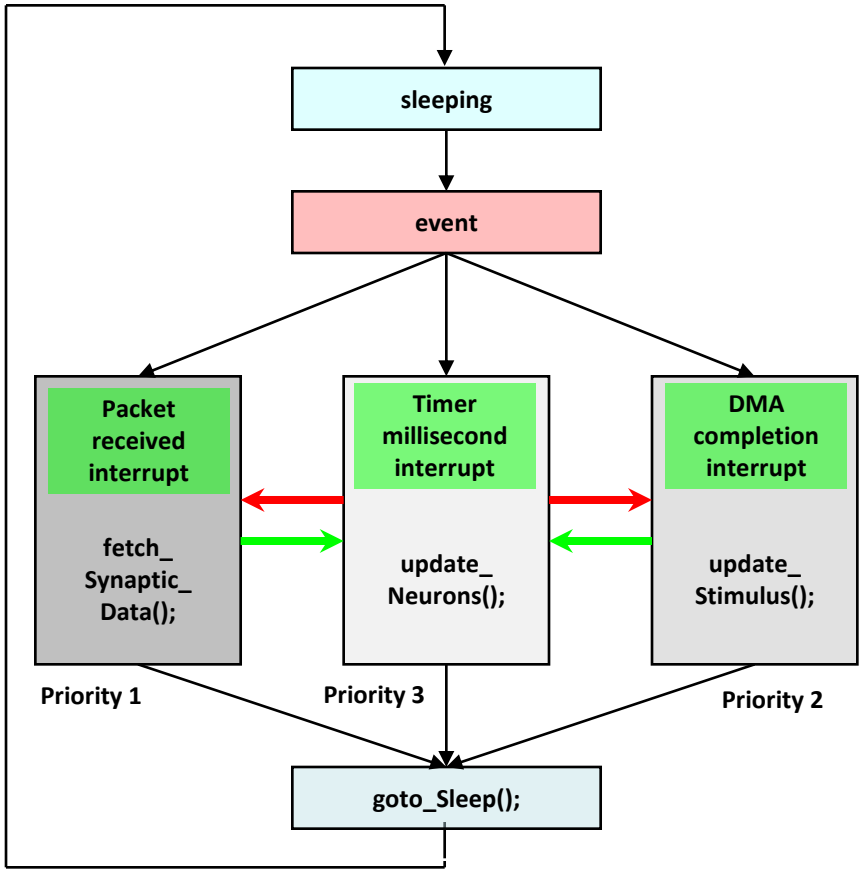


Outline

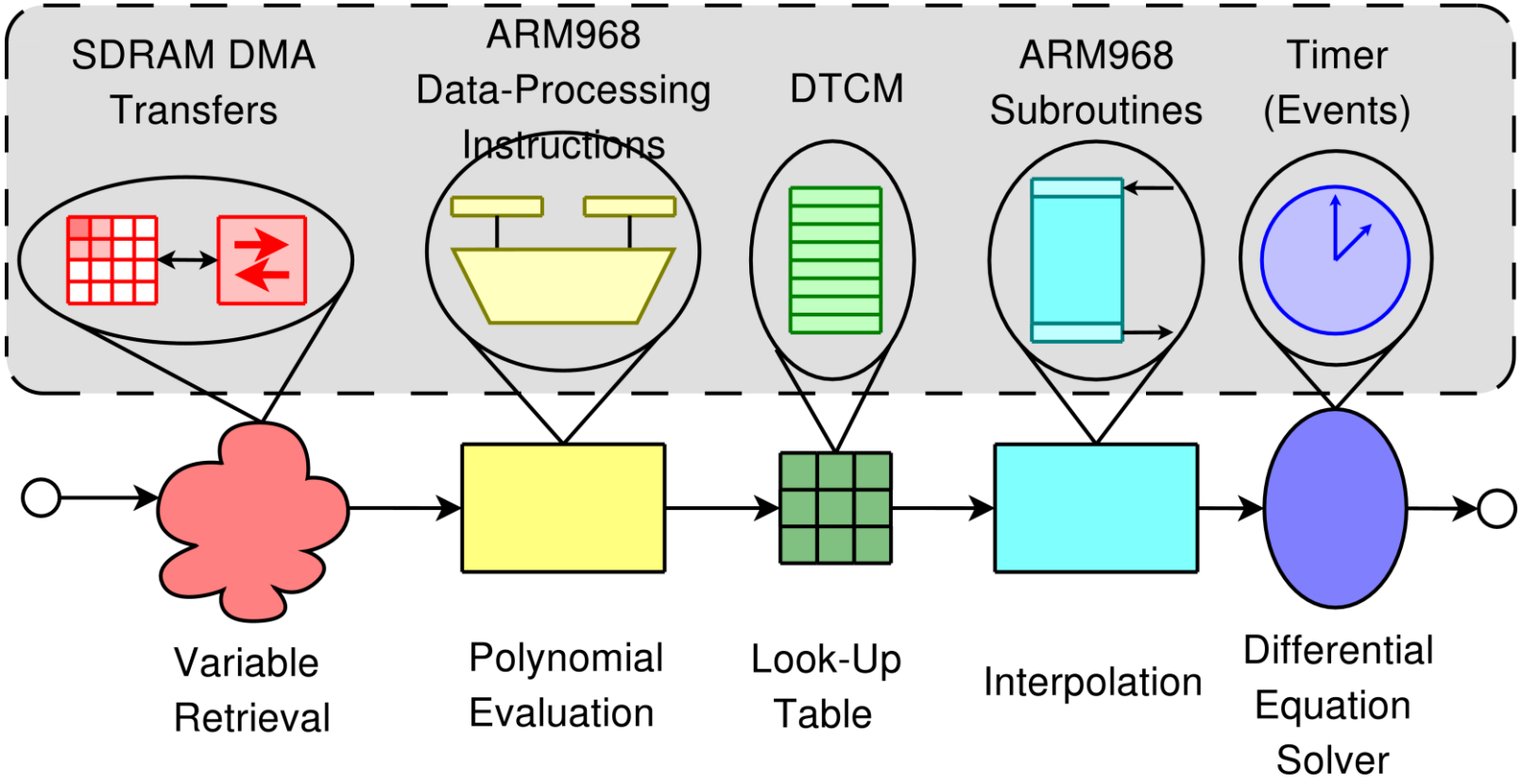
- 60 years of progress
- Computer Architecture Perspective
- Building Brains
- The ***SpiNNaker*** system
- A generic neural modelling platform
- Conclusions

Event-driven software model

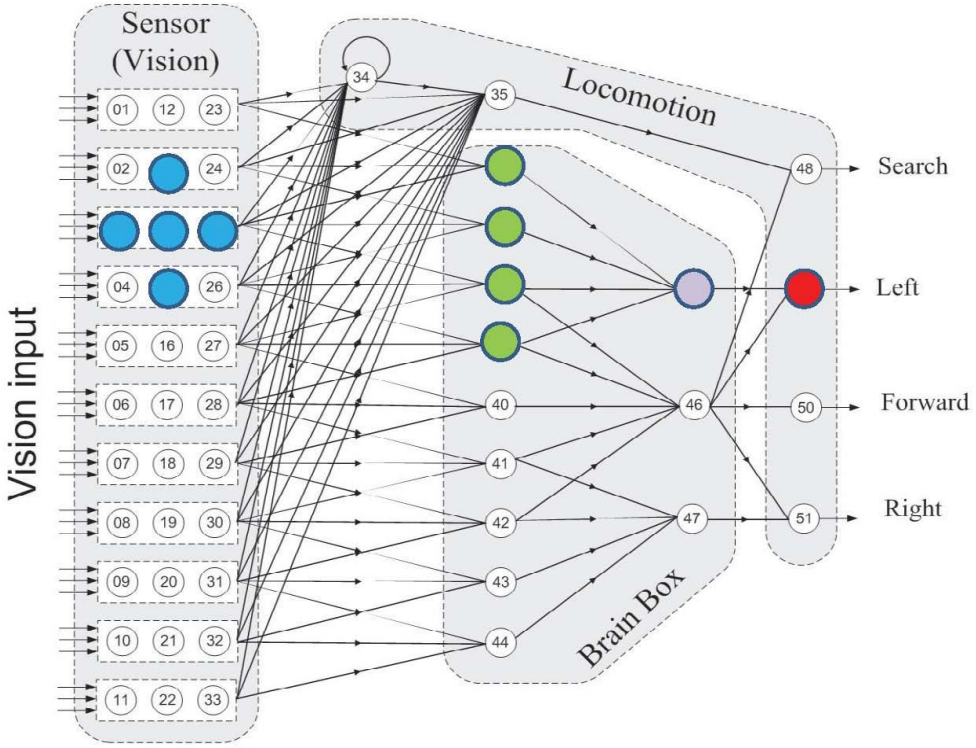
- Spike packet arrived
  - initiate DMA
- DMA of synaptic data completed
  - process inputs
  - insert axonal delay
- 1ms timer interrupt
  - differential equation solver

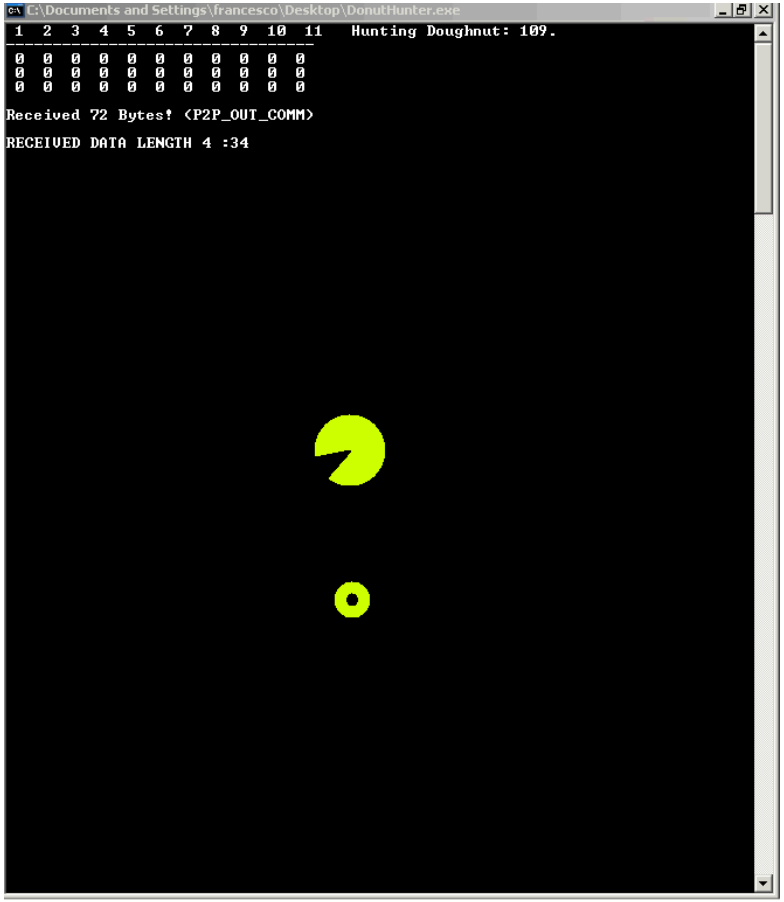


Function pipeline



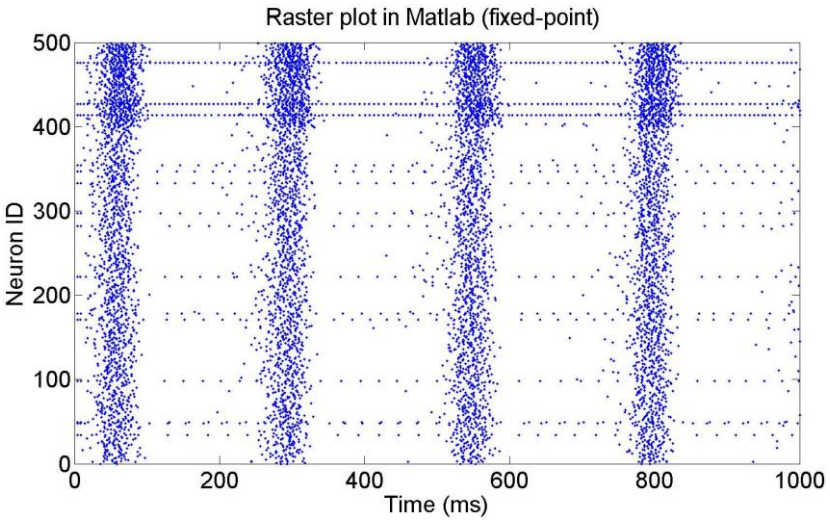
The doughnut hunter

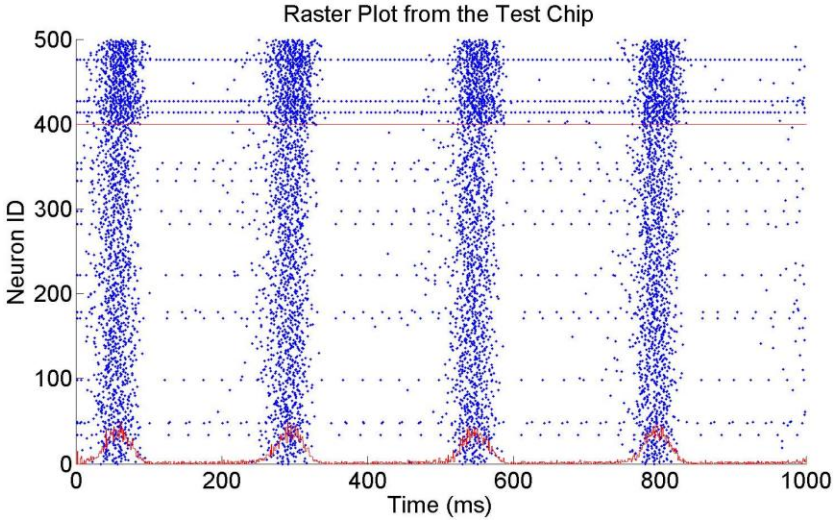


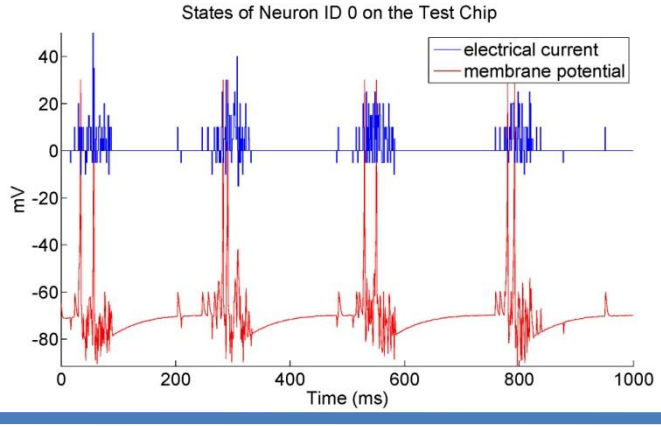


SpiNNaker
Biologically Inspired Massively Parallel Architectures

500 neuron test

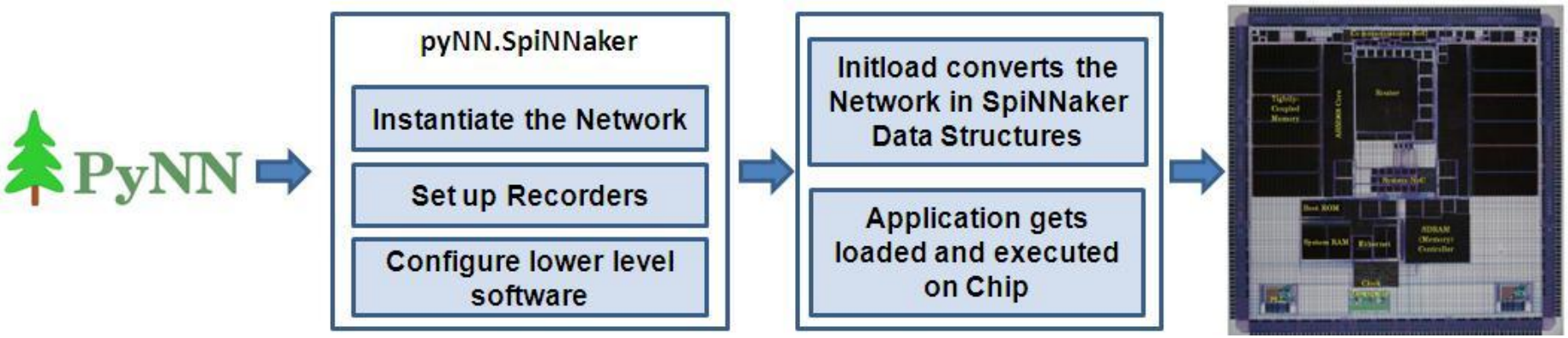






EPSRC UNIVERSITY OF CAMBRIDGE MANCHESTER 1824 The University Of Sheffield. UNIVERSITY OF Southampton ARM Silistix THALES

PyNN integration



pyNN.SpiNNaker module

- builds the network
- extracts the information regarding the network structure and parameters
- scripts the execution of low-level tools
- drives Ethernet interface to load, execute and retrieve results

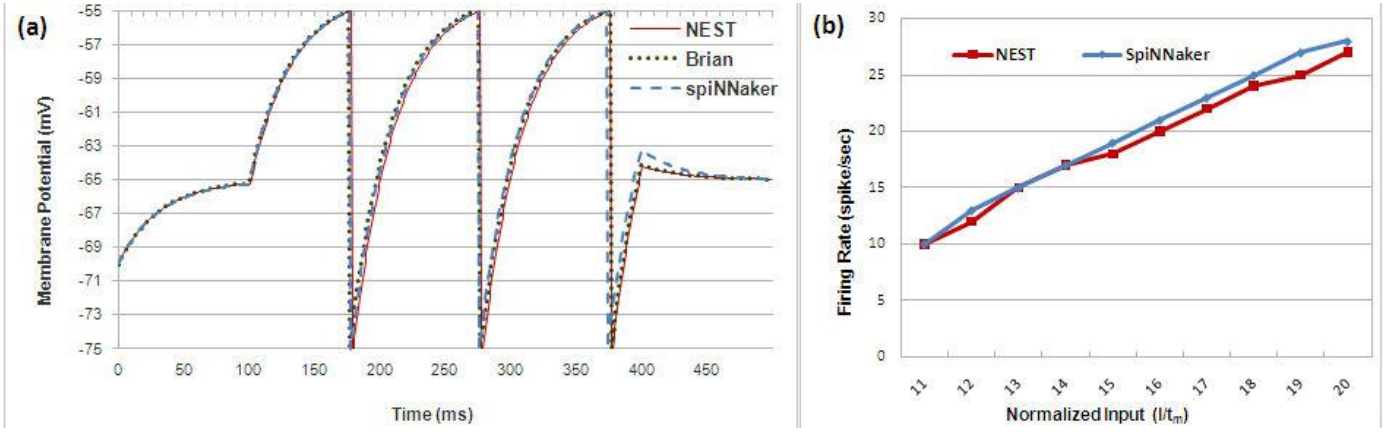
Initload

- compiles the network, mapping into SpiNNaker binary data structures

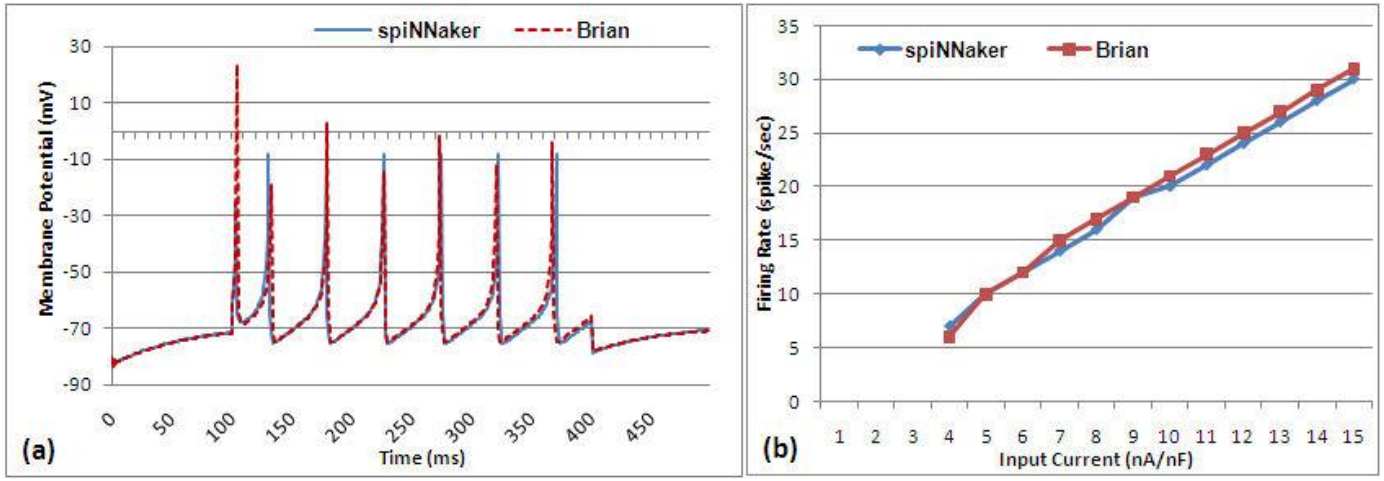
SpiNNaker
Biologically Inspired Massively Parallel Architectures

PyNN integration

- LIF



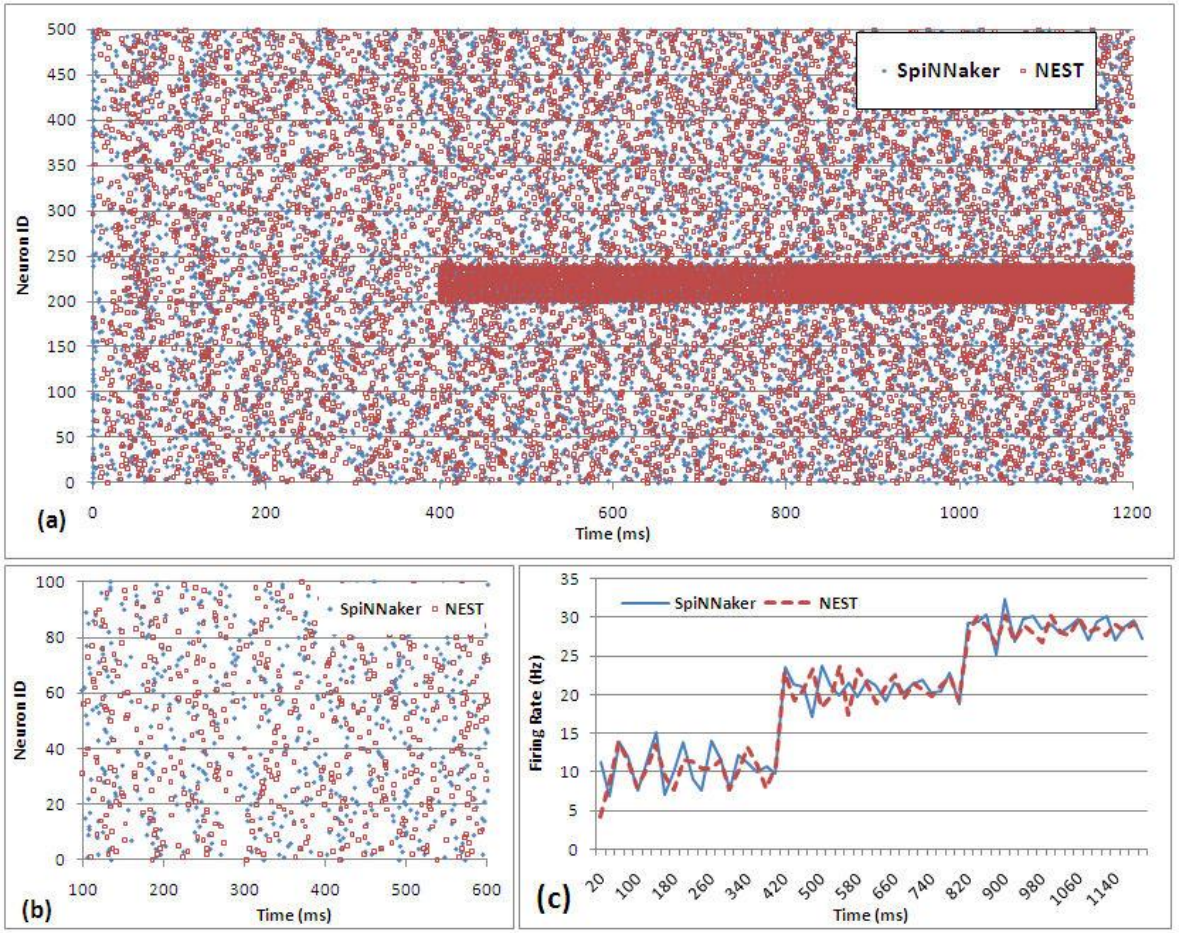
- Izhikevich



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

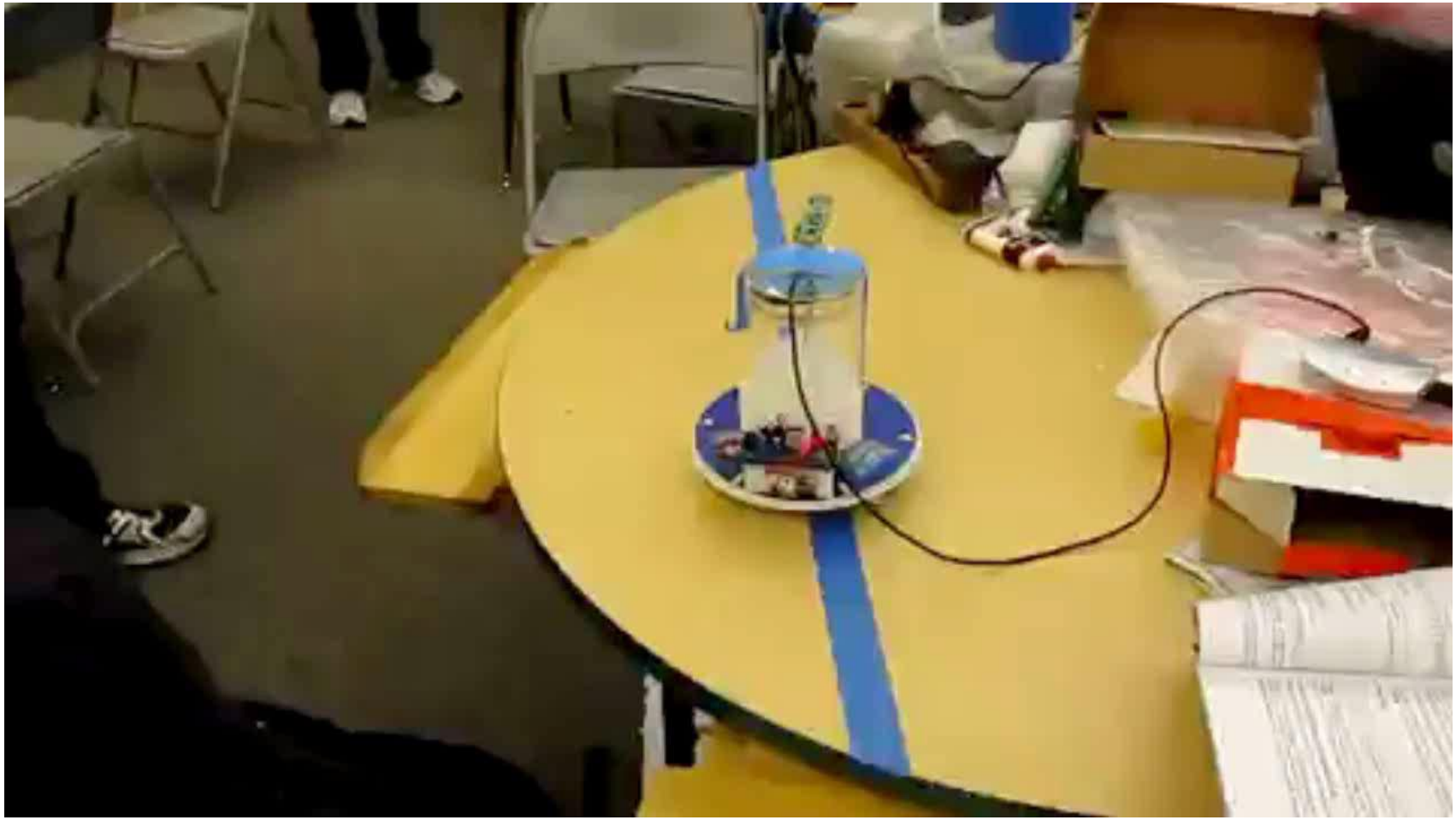
PyNN integration

- Vogels-Abbott benchmark
  - 500 LIF neurons



SpiNNaker
Biologically Inspired Massively Parallel Architectures

Telluride workshop

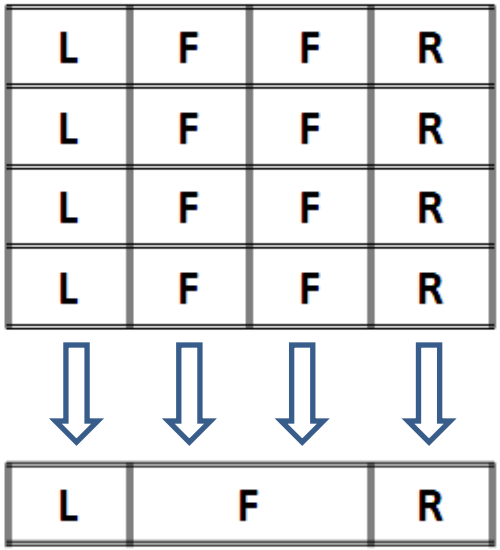


Robot control network

POLARITY RETINOTOPIC MAP
16x16x2

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

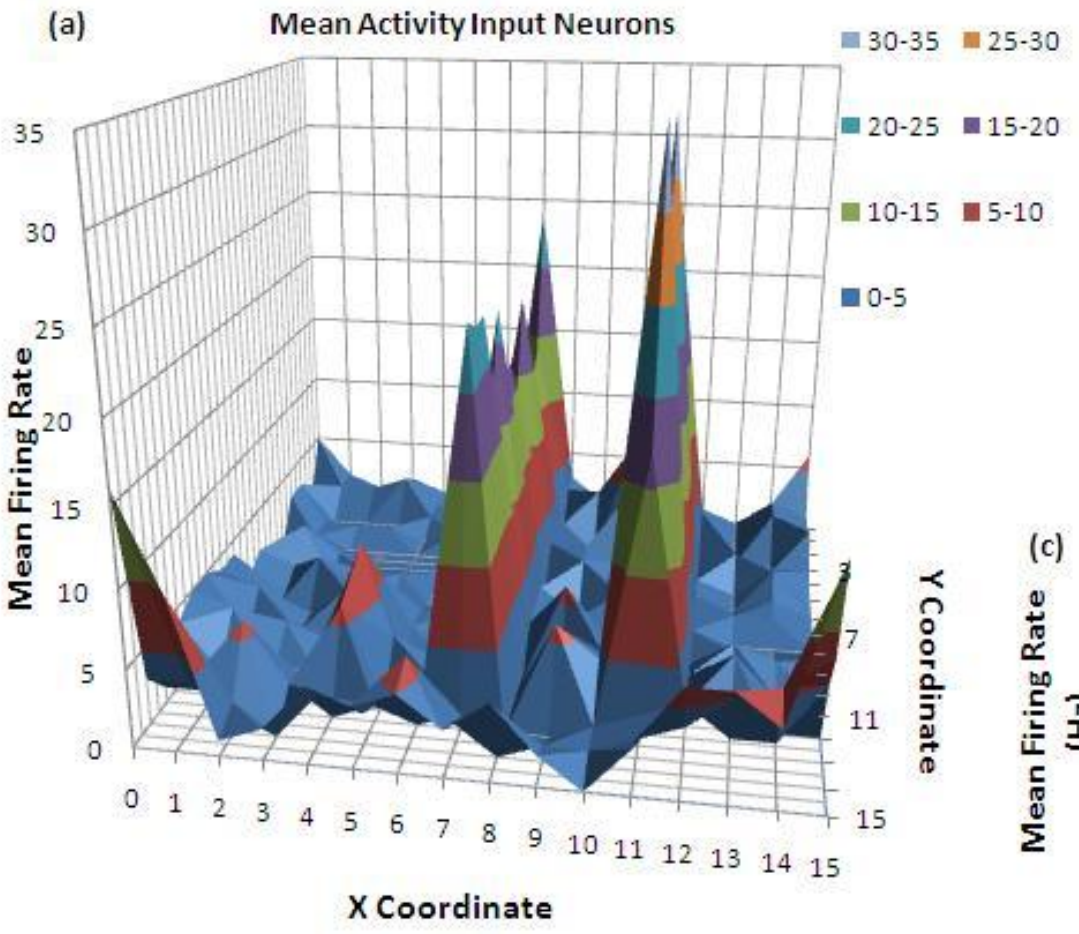
SUBSAMPLER
4x4
(1-Winner-Take-All)

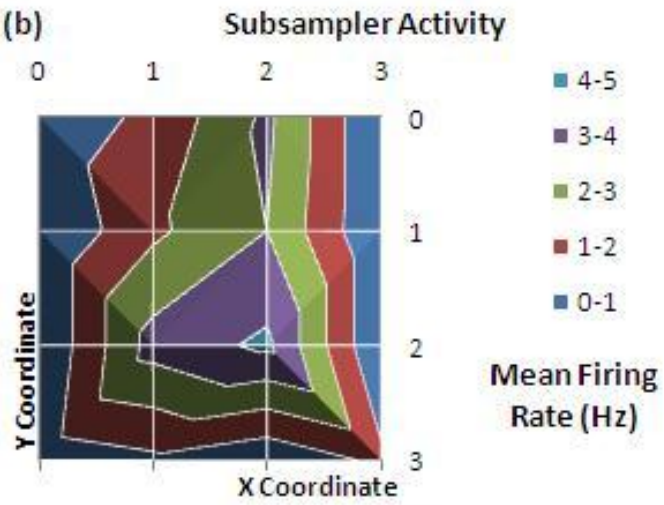


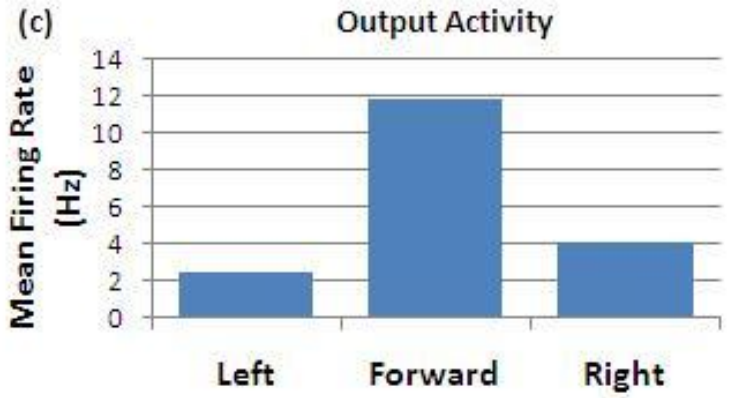
OUTPUTS
(1-Winner-Take-All)

SpiNNaker
Biologically Inspired Massively Parallel Architectures

Robot control network







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

Outline

- 60 years of progress
- Computer Architecture Perspective
- Building Brains
- The ***SpiNNaker*** system
- A generic neural modelling platform
- Conclusions

Current status…

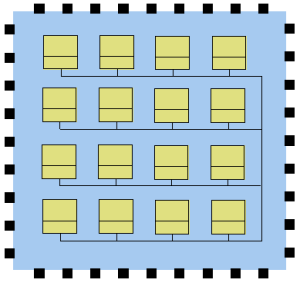
- Test chip: 2 ARM968 processors
- Test card: 4 test chips, 8 processors
  - Cards can be linked together
- Neuron models: LIF, Izhikevich, MLP
- Synapse models: STDP, NMDA
- Networks: PyNN -> SpiNNaker, various small tools to build Router tables, etc
- Monitor/debug: SpiNN doctor

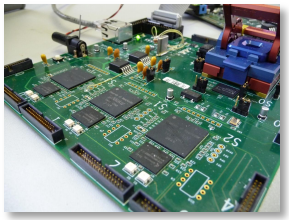
…and the next steps

- Full (18-processor) chip: tape-out end Oct'10
  - Package silicon with SDRAM in February 2011
  - Build 4-chip test card (Feb'11),
    50-chip 103 machine (Q2 2011), 500-chip 104 machine (H2 2011), 5,000-chip 105 machine (H1 2012), 50,000-chip 106 machine (H2 2012).
  - Rebuild event-driven software foundations
  - Extend PyNN -> SpiNNaker support
  - Monitor/debug tools,
    developmental models, intrinsic configuration, run-time fault-tolerance,…

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
- Fault-tolerance has been considered throughout
  - though the approach is rather 'ad hoc'
- We have prototype working hardware!





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