Neuromorphic Computing

Dr. Charles Kind

Computational Neuroscience Group, SCEEMS, Bristol University

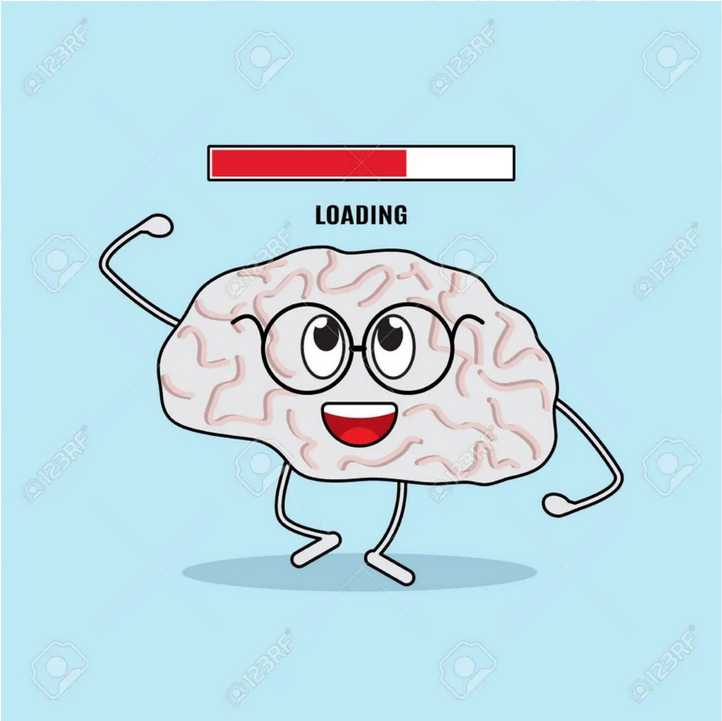
October 2022

Lecture series contents

- The state of computer hardware today.
- Challenges in hardware driven by societal needs (data science).
- What future solutions could answer current (and near future) data processing needs?
- How are brains like computers?
- How are computers like brains?
- What lessons learnt about human brains can we apply to computers?
- What is neuromorphic computing?
- The three main approaches to neuromorphic computing.
- Examples of the state of the art in neuromorphic computing.

Lecture series contents

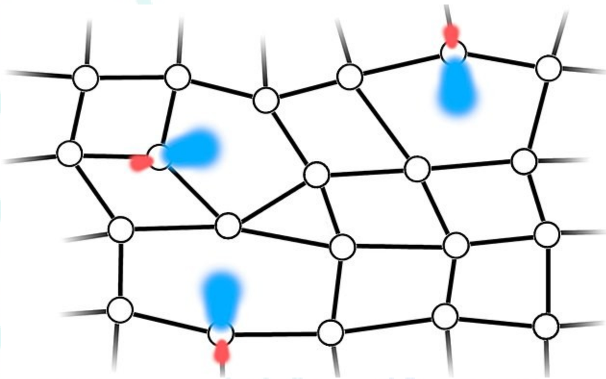
- Transistor improvements in size and power use have slowed and therefore the projected advances in computer power have slowed dramatically
- Given the laws what are the ways we can improve computer performance and why should we?
  - More transistors
  - Faster speeds (clock rate)
  - Multiple cores, parallelism
  - Optimised software
  - Integration of components
  - Specialisation of components
  - Our needs only increase

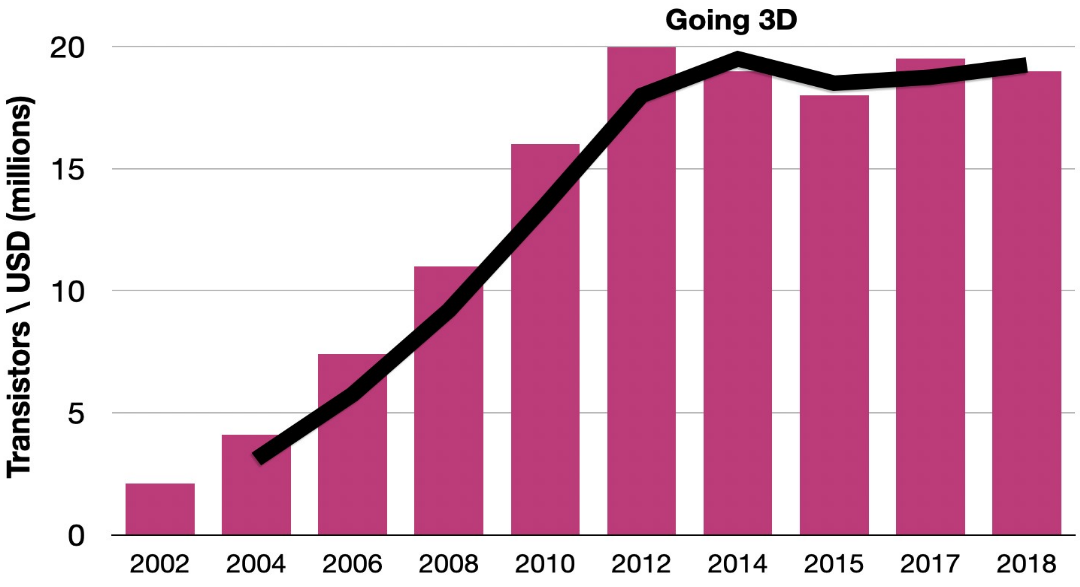


Integrated circuits, what really happened.

More transistors please

- Dangling bonds (an unsatisfied valence on an immobilized atom) trap electrons, limits transistor size reductions (around 5 nm)
- A partial solution is to move to 3D transistors
  - Increased complexity and cost
- Quantum effects: new and expensive annoyances?



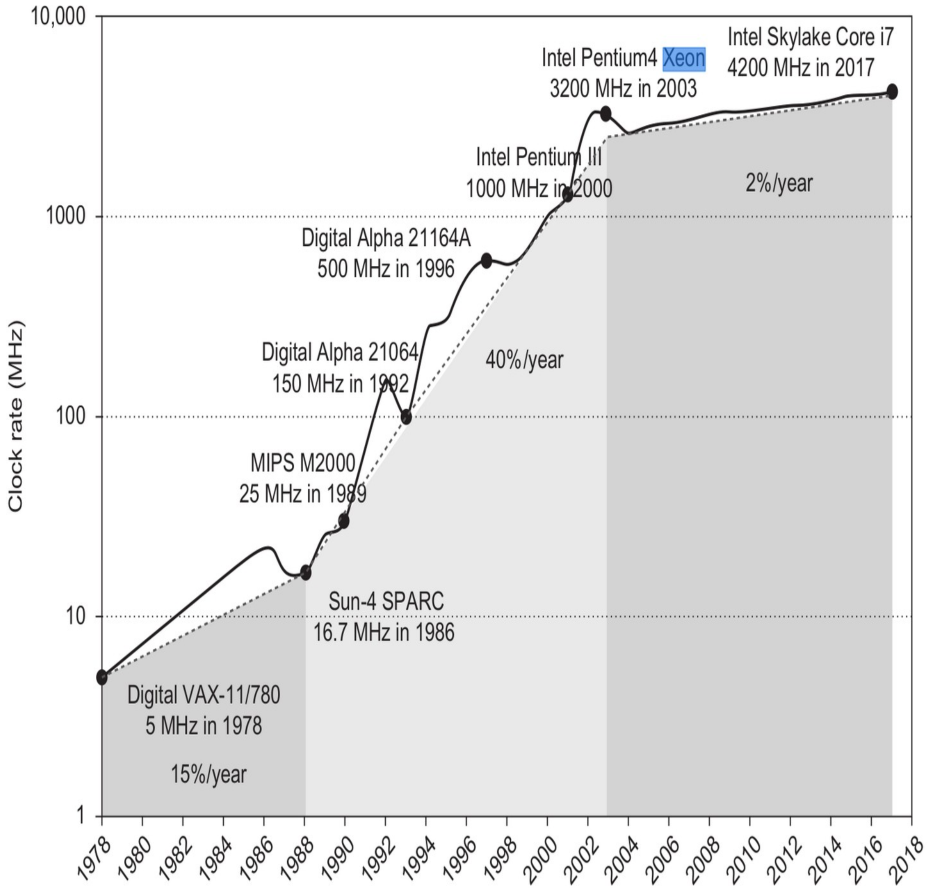


Data from Kwabena Boahen. A Neuromorph’s Prospectus. IEEE, 2017

Integrated circuits, what really happened.

Higher clock speeds please

- For CMOS chips, the primary energy consumption is the switching of transistors, also called dynamic energy.
- The first microprocessors (such as the Intel 4004) consumed less than a watt at around 100 Khz
- 32-bit microprocessors (such as the Intel 80386) used about 2 W at 12-40 Mhz
- Core i7-8700K uses around 100 W (3.7 Ghz) and can peak at 150 W (4.7 Ghz)
- We are at the limits of what can be air cooled and have been for around 15 years



Data from John L Hennessy and David A Patterson. Computer Architecture:

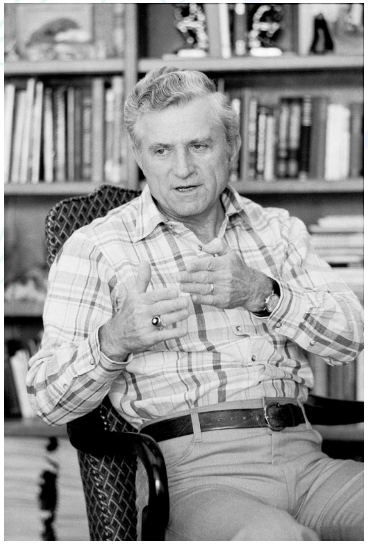
A Quantitative Approach. Elsevier, 2017

Integrated circuits, what really happened.

Distributed computing or more cores please

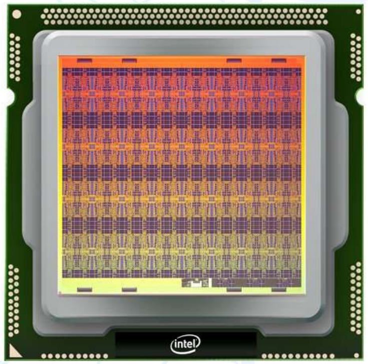
- This has been the predominant driver of increased performance since the failure of the LAWS
- Amdahl's law: the performance improvement to be gained from using some faster mode of execution is limited by the fraction of the time the faster mode can be used
  - Gene Amdahl, 1967
- Suppose 20% of the algorithm cannot be distributed for multicore processing. In that case, the maximum gain in performance is limited to 5x, no matter how many cores are available
- Power consumption and the rise of the GPU
- Von-Neumann bottleneck and the memory wall

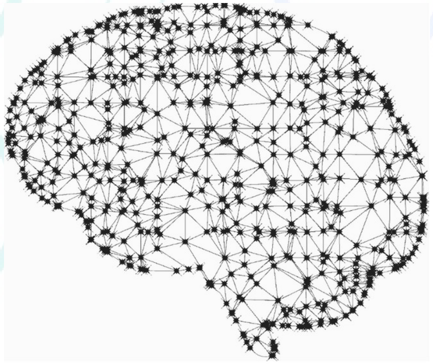
Amdahl



Integrated circuits: Summary

- More transistors
  - Energy and miniaturisation constraints
- Faster processors
  - Energy and heat dissipation constraints
- Distributed computing
  - Can be restricted by applications
  - See above
- General issues
  - Data transfer constraints
  - Power

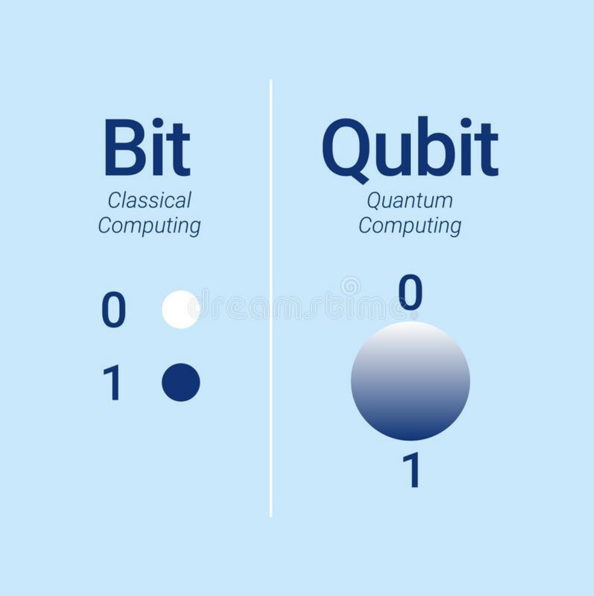




Computers vs brains

Blowfly retinal synapse vs your PC

- Energy consumed (Laughlin et al. 1998)
  - 1e4 ATP molecules per bit
  - Each synapse transmits 55 bits per second
  - 1e6-1e7 ATP molecules for spike coding
  - Single ATP molecule ~ 1e-19 joules (1 watt = 1 joule per second)
- Computer bit flipping (Prieto et al. 2022)
  - 10-14 joules per bit
- Comparison
  - Blowfly: 1e4 * 1e-19 = 1e-15 joules per bit
  - Therefore ~1e16 times more efficient?????
  - i9-9900K can manage 4e5 MIPS or 4e11 IPS
  - Note – an instruction can be more complex than a bit flip



Computers vs brains

The human brain

- Energy consumed
  - Kwabena Boahen (2017), current through ion channels ~ 20 W
  - Ferris Jabr (2012), using metabolic rates ~ 12 W

- Processing power and memory
  - Eugene Izhikevich (2004), ~ 10^11 neurons ~ 1e18 FLOPS
  - Sandberg and Bostrom (2008) ~ 1e18 bits (~1e8 GB)

The computer

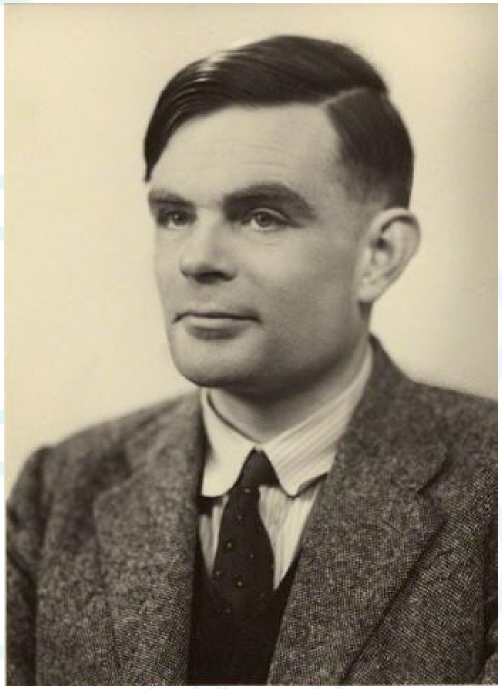
- Nvidia 3080, 30 TFLOPS (1e12) @ 320 W
- Fugaku, 1e18 FLOPS, 4e16 bits @ 30 MW and $1 Billion

Are brains computers?

Should we understand how brains work before asking this question or does asking this question tell us something about brains?

- Turing machine (Turing 1936)
  - An infinite tape containing a finite set of symbols and rules for moving and writing to the tape.
  - The protoype, mathematical, concept of the modern computer
- I can:
  - Emulate a Turing machine?
  - Factorise prime numbers
  - Calculate digits of pi

$$\frac{1}{\pi} = \frac{\sqrt{10005}}{4270934400} \sum_{k=0}^{\infty} \frac{(6k)!(13591409 + 545140134k)}{(3k)!\,k!^3\,(-640320)^{3k}}.$$



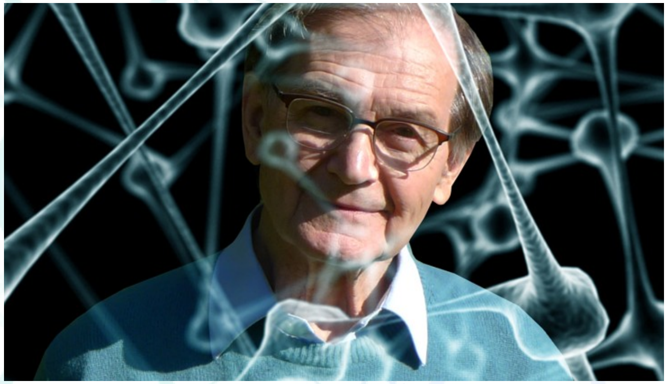
Alan Turing

Turing machine

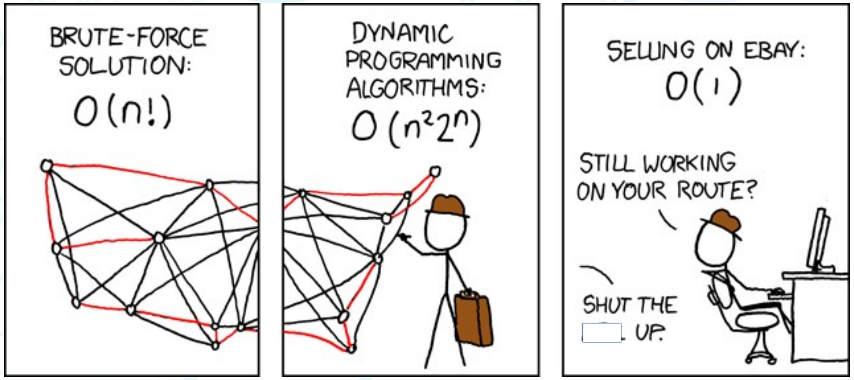
Are computers brains?

Some espouse the belief that a thermostat has intelligence as it changes state with respect to the world!

- Roger Penrose theorises that quantum events occur within the brain and hence classical computers cannot truly be conscious.
- We cannot agree on what intelligence is so therefore how should we measure this?
- Consider specific tasks
  - Travelling salesman and other NP (nondeterministic polynomial-time complete) complete problems. Here size matters!
- Can we train a neural network to solve or aproximate NP problems?



Roger Penrose



xkcd 399

Brains vs computers

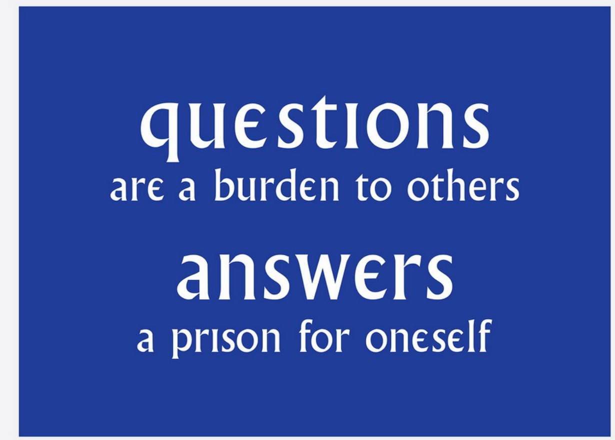
What can each do that the other cannot (currently)?

- Brains can change their hardware
- Brains can discover new science and mathematics through intuition and broad thinking
- Brains can emulate computers
- Computers can factorise prime numbers and do high precision floating point calculations
- Computers can beat humans at specific large problem space games such as chess
- Computers can react quickly ~ 1e-6s vs human peak 120 ms

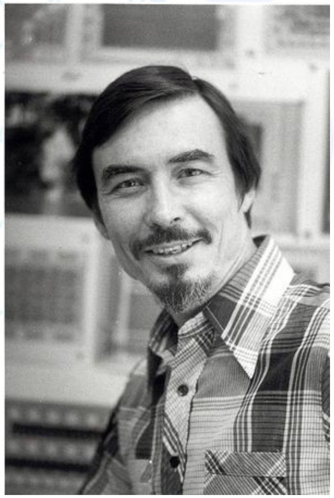
Neuromorphic Computing

The historical inspiration

- What to do about the obvious limits to semiconductor technology?
- Faster to scale old designs to smaller feature sizes than to innovate at the architecture level.
- 1981: Carver Mead, Richard Feynman and John Hopfield join together to teach: “Physics of Computation” at the CalTech
- “the brain is a factor of 1 billion more effective than our present digital technology and a factor of 10 million more efficient than the best digital technology that we can imagine”, Mead, 1990
- Lets make computers like human brains!



Patrick McGoohan, 1967



Carver Mead

Neuromorphic Computing

Goals

- Scalable architecture designed to run brain like computations
- Replace virtual neurons/synapses with physical, analog devices
- ‘In memory’ devices
- Reduce power consumption
- Enable ‘edge computing’
- Sit in the gap between high accuracy, high energy classical computing and low accuracy, low energy neural computation
- Become standard CPU/hardware extension like SIMD SSE
- Learn more about how the human brain works

Thank you :)