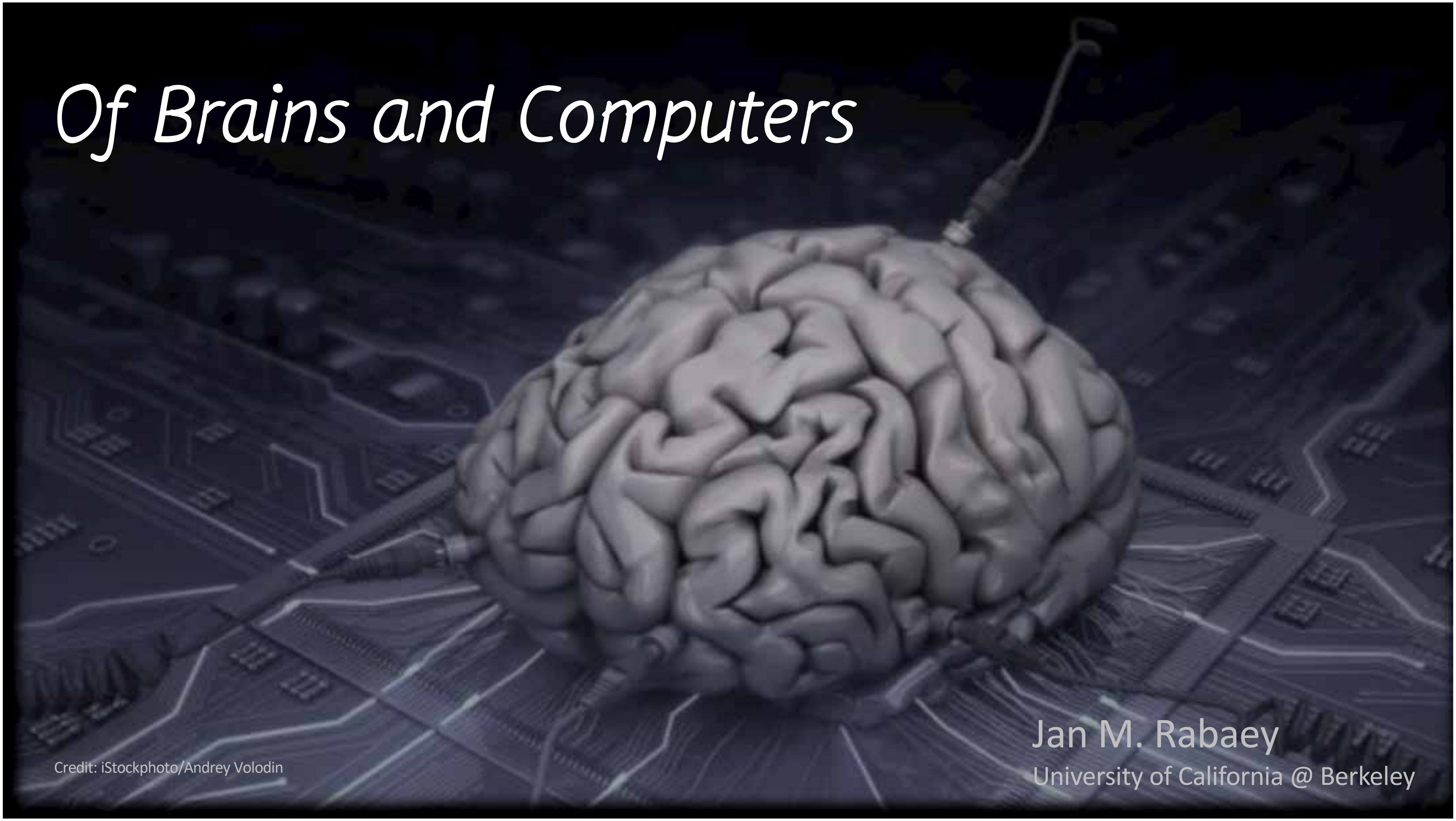
Of Brains and Computers

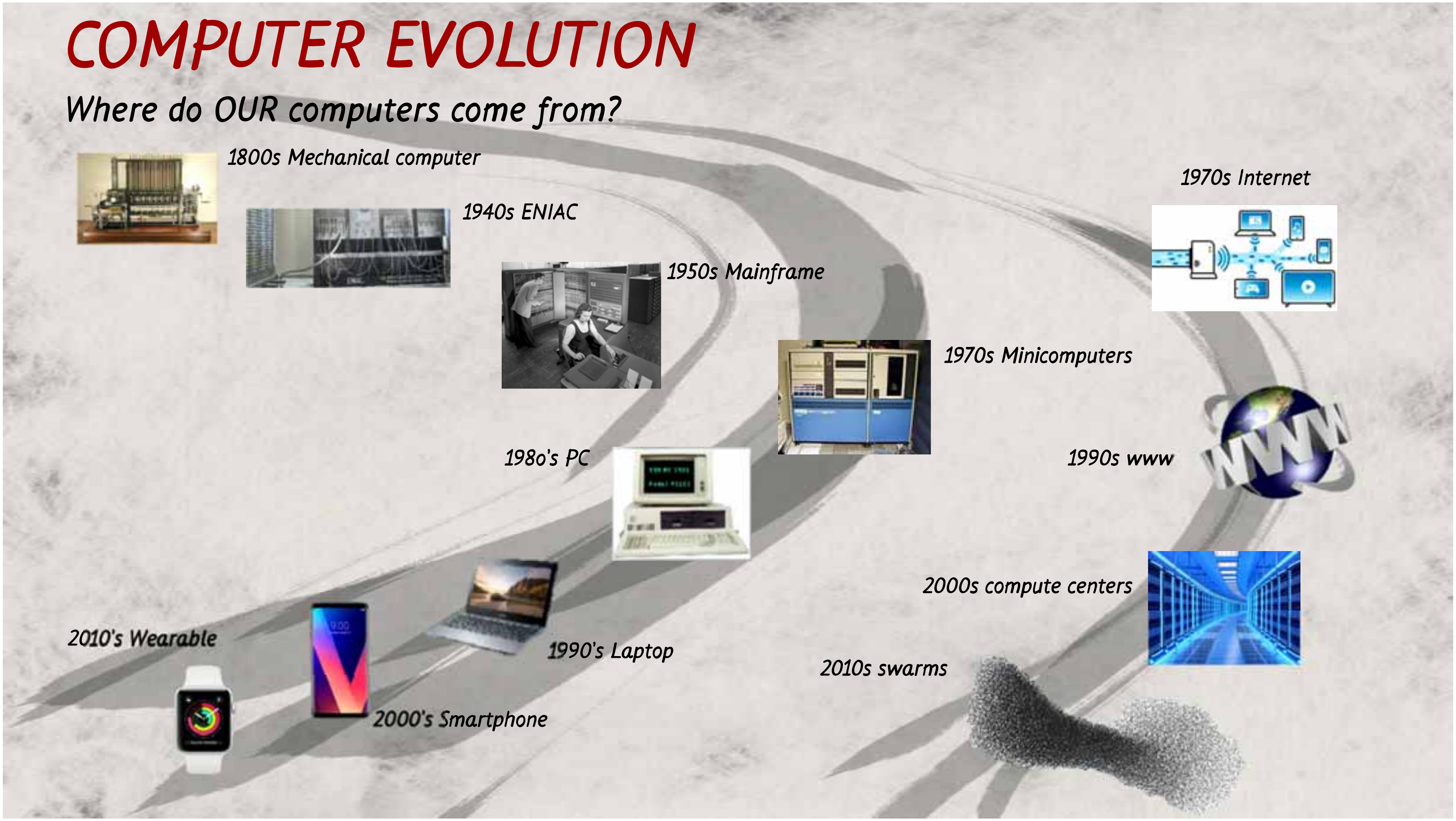


Credit: iStockphoto/Andrey Volodin

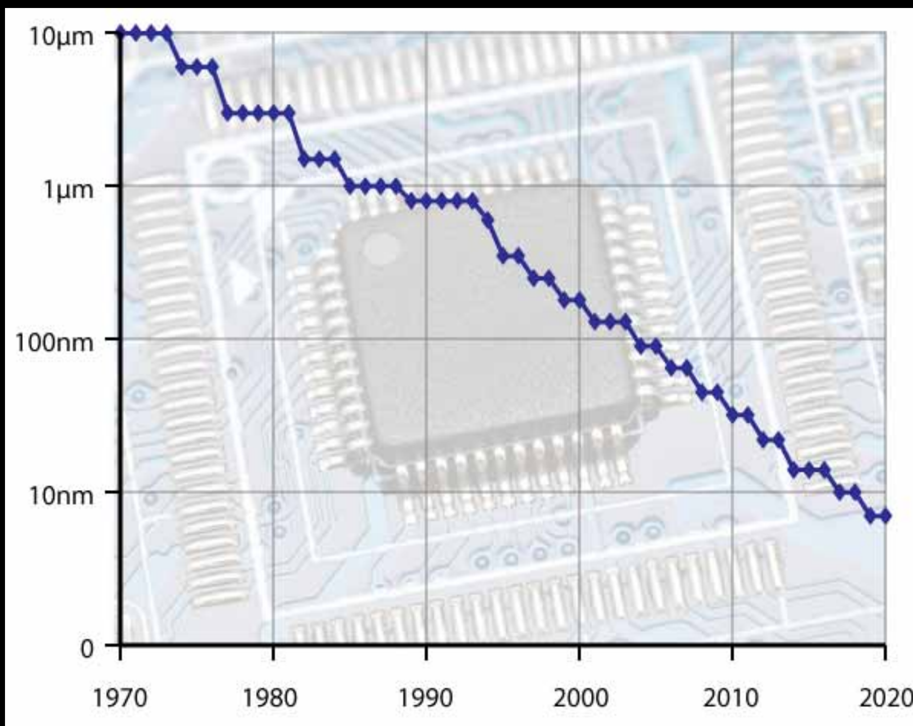
Jan M. Rabaey
University of California @ Berkeley

Congratulations Lou!

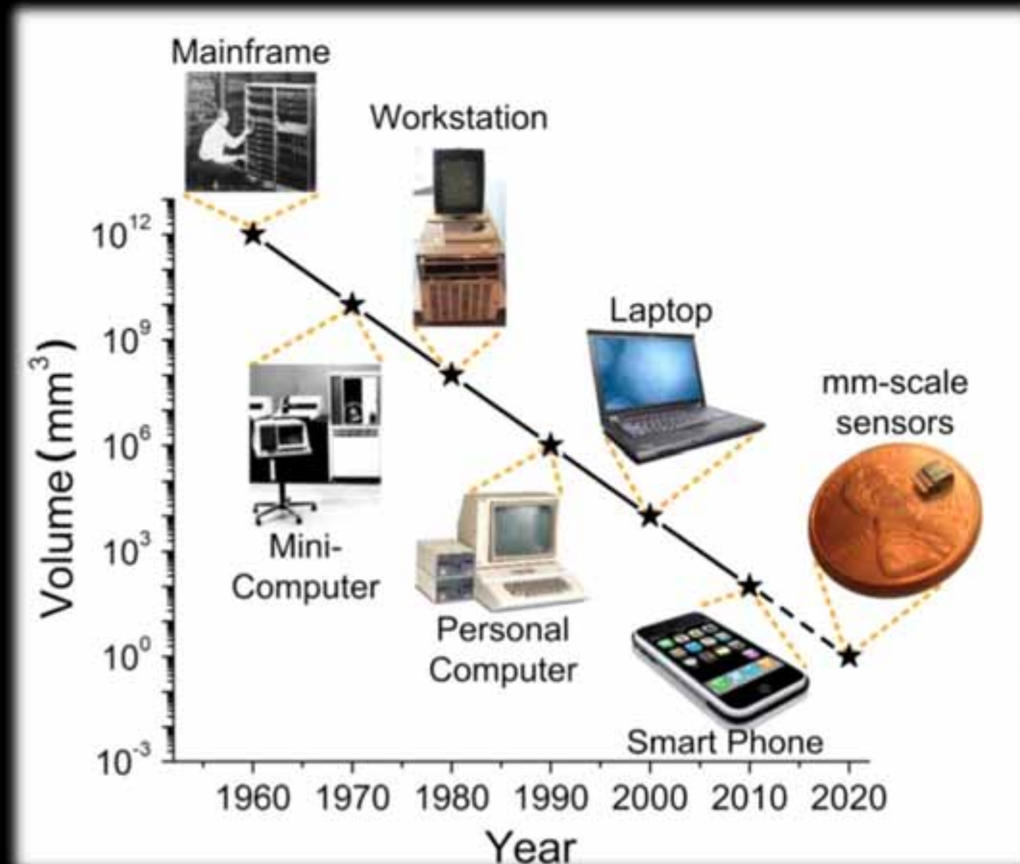
And thanks for all the fun moments!


COMPUTER EVOLUTION
Where do OUR computers come from?
1800s Mechanical computer
1940s ENIAC
1950s Mainframe
1970s Minicomputers
1980's PC
1990's Laptop
2000's Smartphone
2010's Wearable
1970s Internet
1990s www
WWW
2000s compute centers
2010s swarms

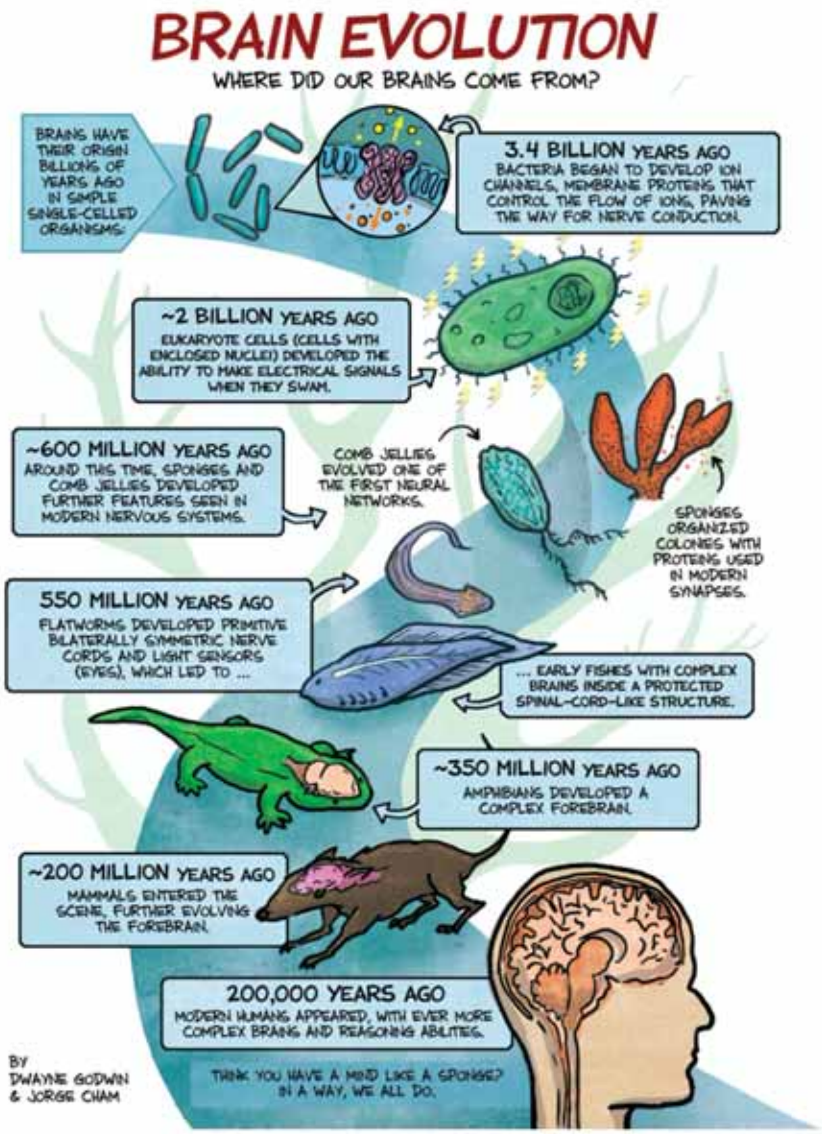
Computer Size Evolution



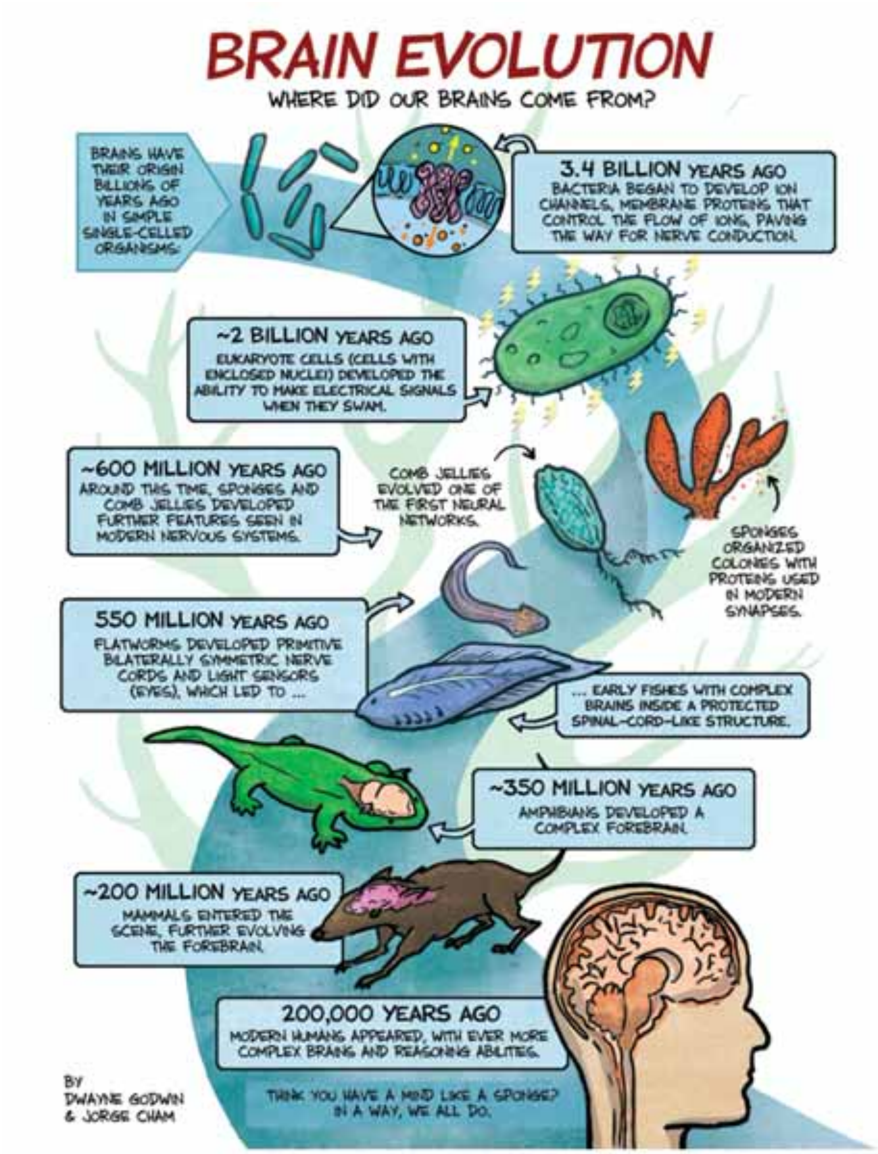
Transistor Size (log scale)

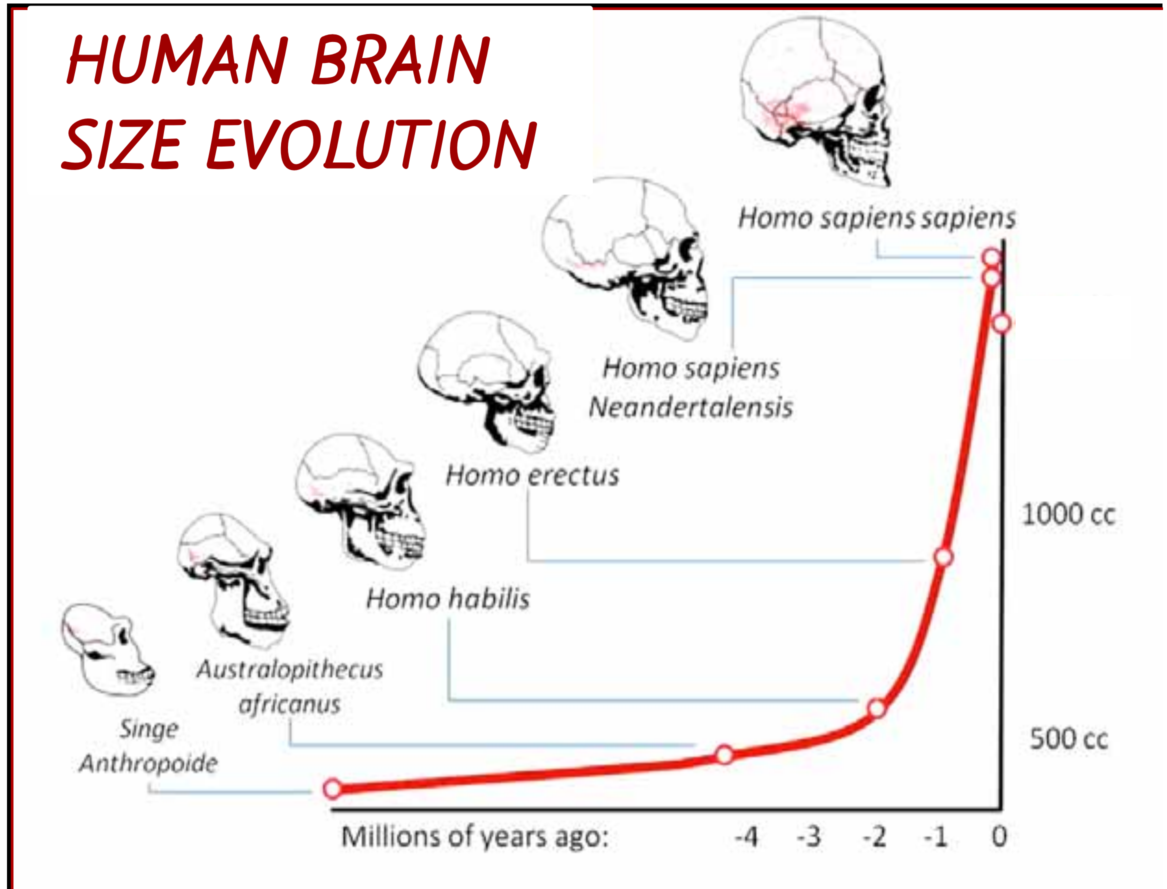


Futuretimeline.net



Scientific American, July 2014



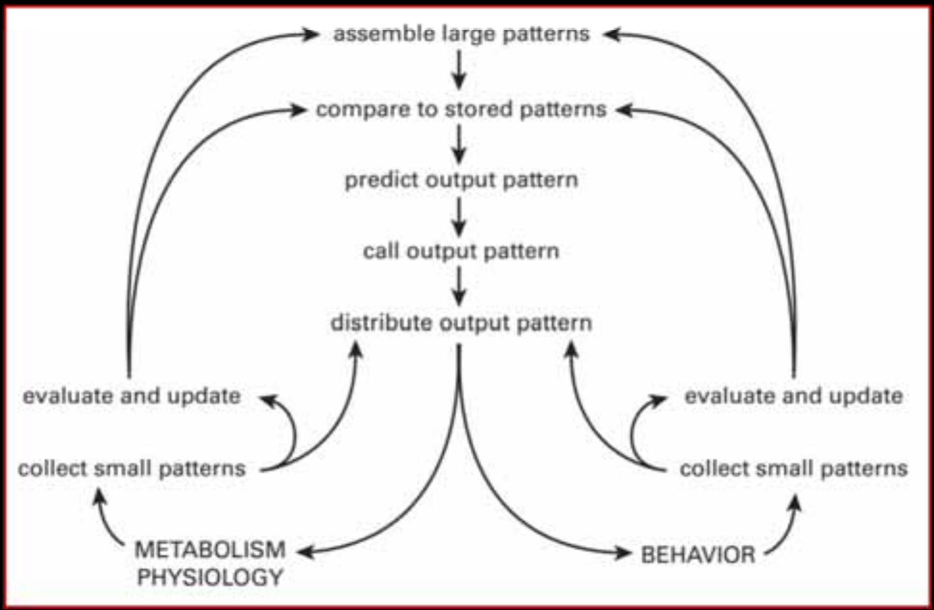


http://aquatic-human-ancestor.org/anatomy/images/brain-size.jpg

Scientific American, July 2014

Different goals and metrics

Brain mission: "survive, prosper, procreate"

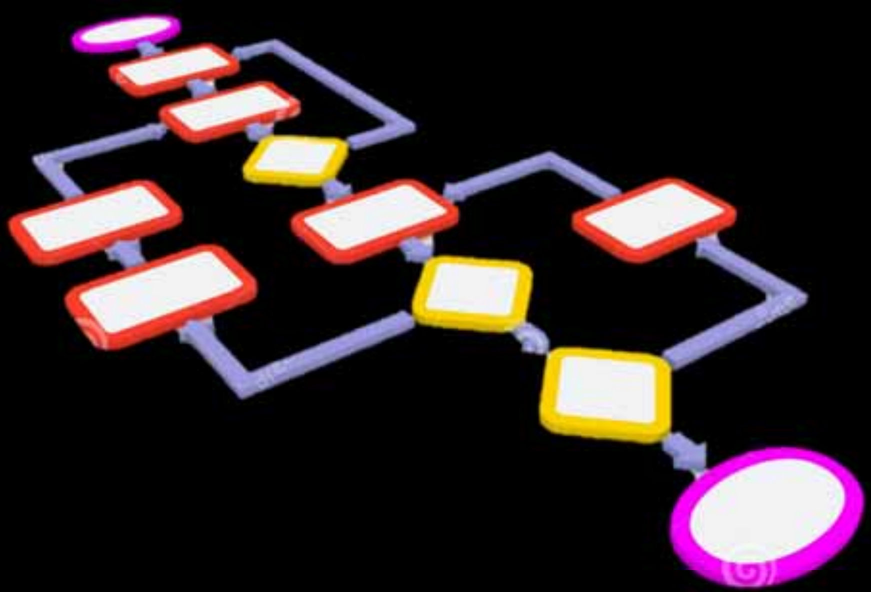


Inner loop

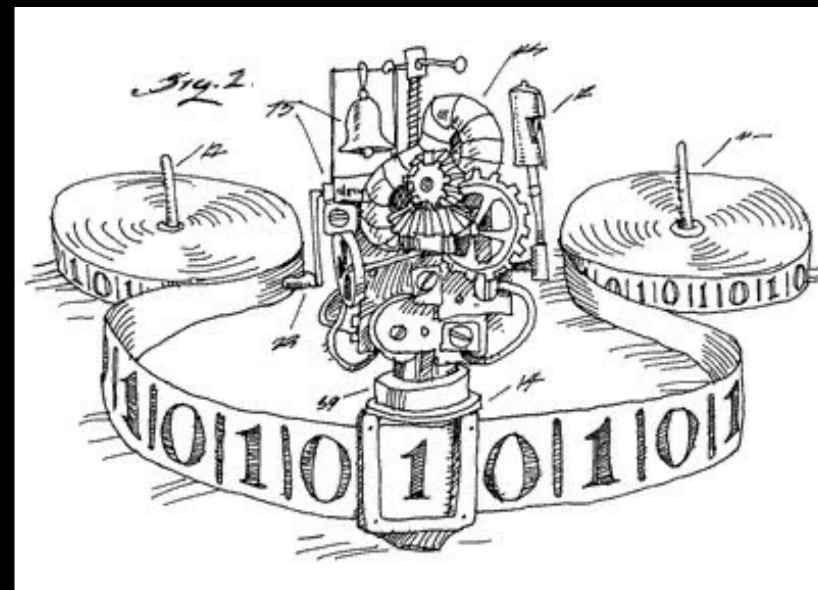
Outer loop

Maximize reward at minimum energy cost

"Computers execute algorithms"

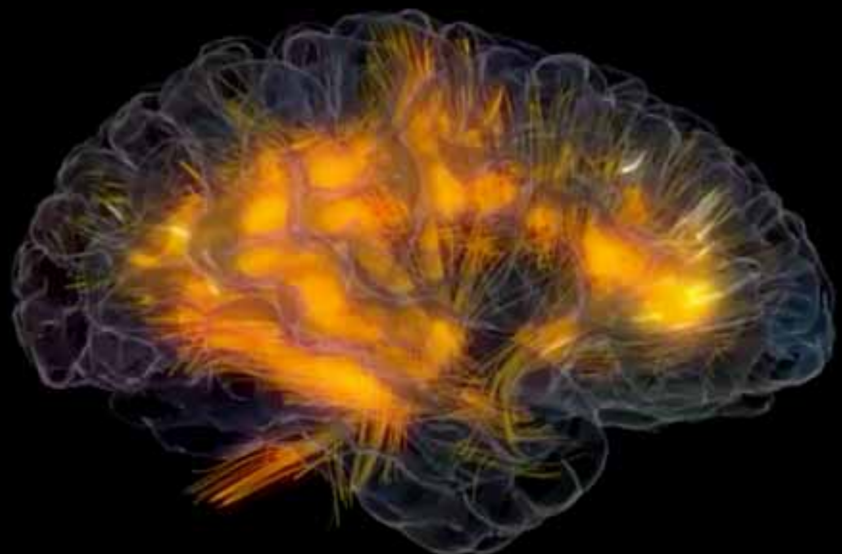


maximize performance,
minimize energy and cost

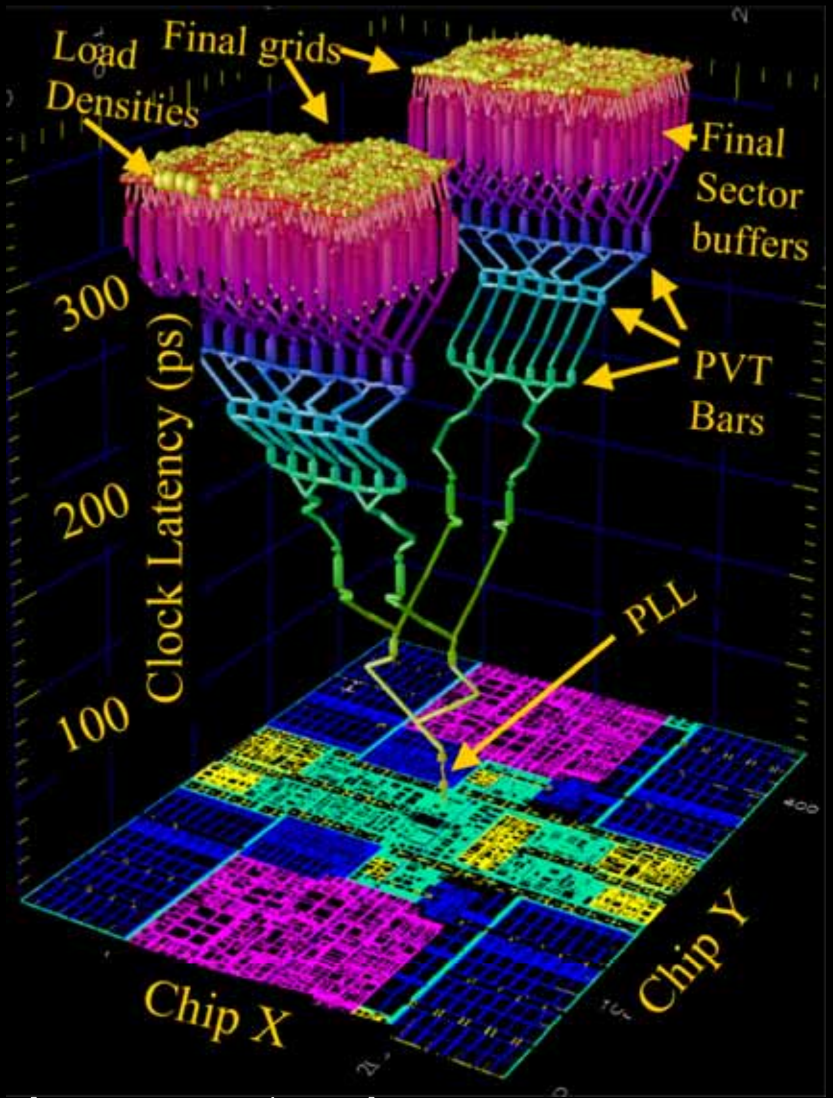


Different approaches

Different approaches



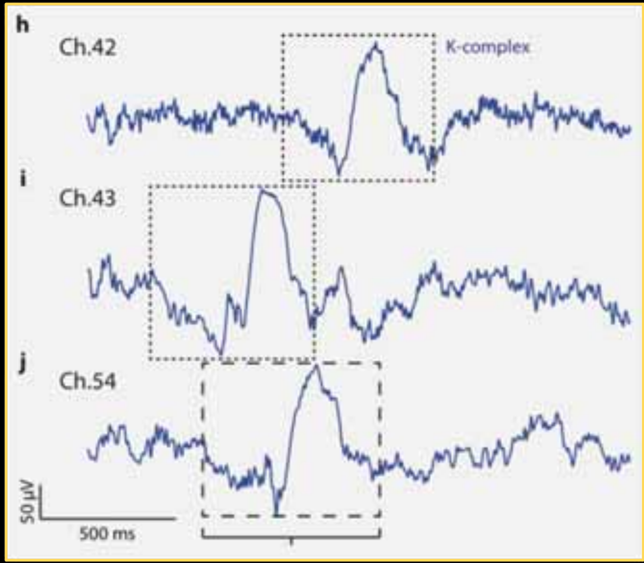
http://neuroscape.ucsf.edu/glassbrain

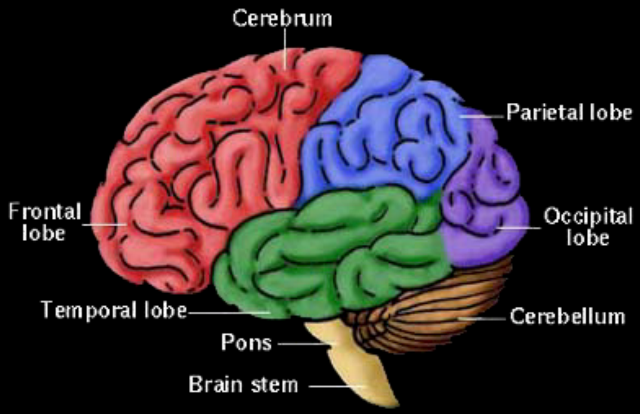


[Courtesy: P. Restle, IBM]

Brain-Computing: Distinguishing Properties

- Learning-based programming paradigm
- Approximate (statistical) & mostly analog
- Overcomplete and redundant
- Data represented in many ways
  - Patterns, phase relations, distributions
- Randomness as a feature

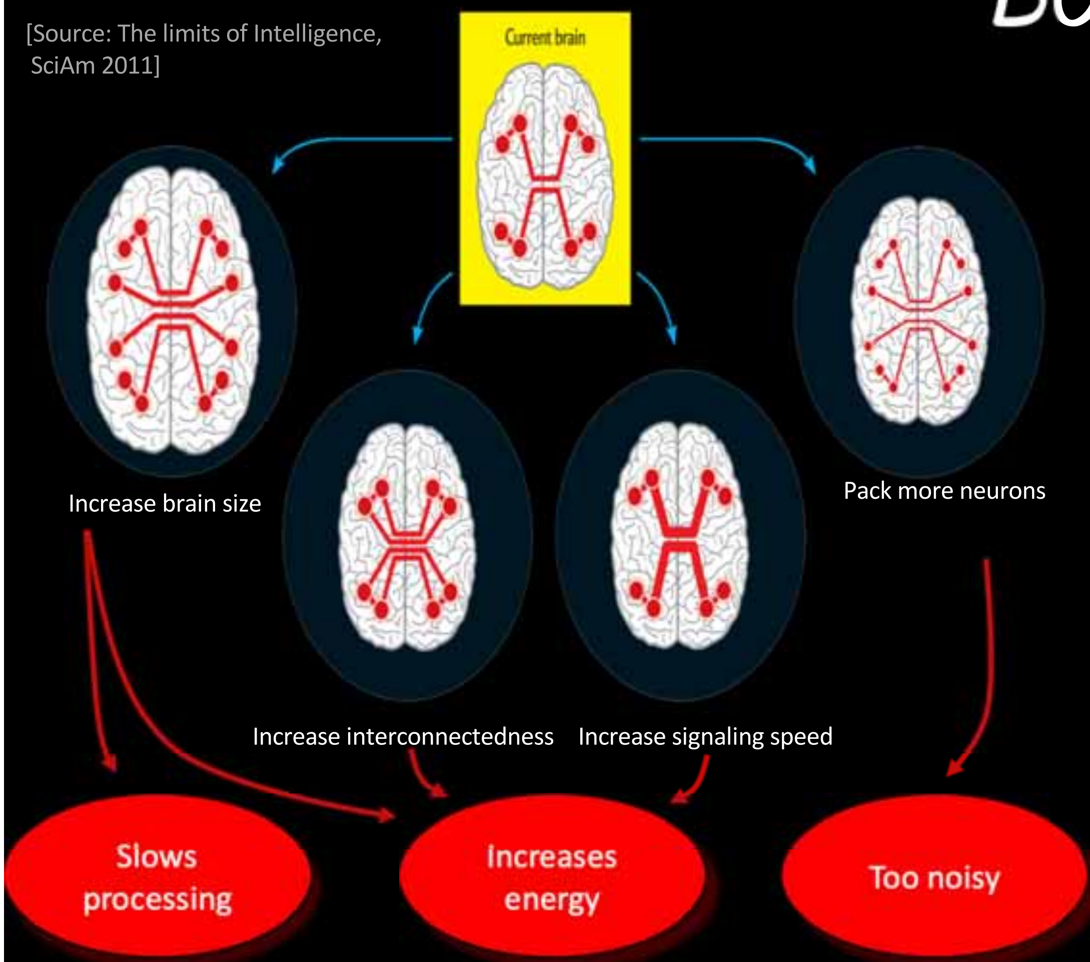


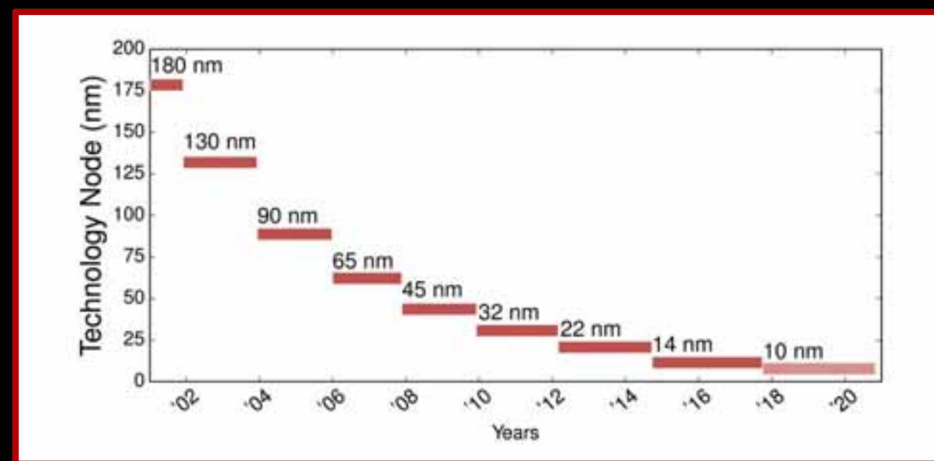


- Function mapped to space
  - no time multiplexing
- Intertwined memory and logic
- Embarrassingly parallel
- Sparse

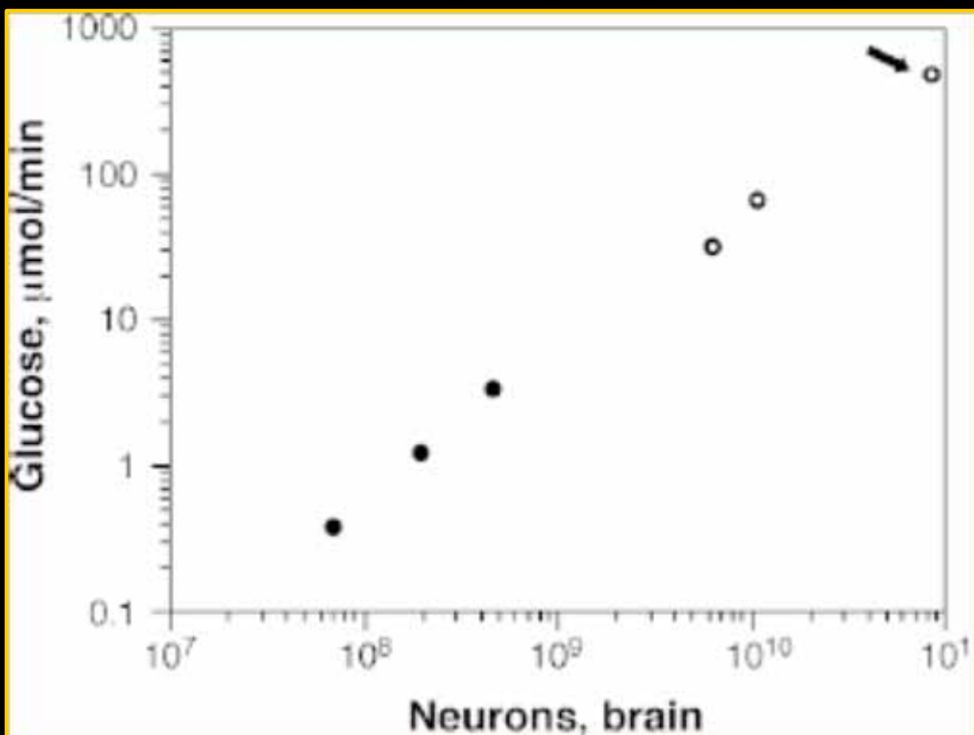
Both are at limits

[Source: The limits of Intelligence, SciAm 2011]





Energy/Power THE Limiting Factor



Energy/Power THE Limiting Factor

3D IC

Temperature range: -45 – 85 °C
Avg heat removal: up to ~ 100 W/cm^2

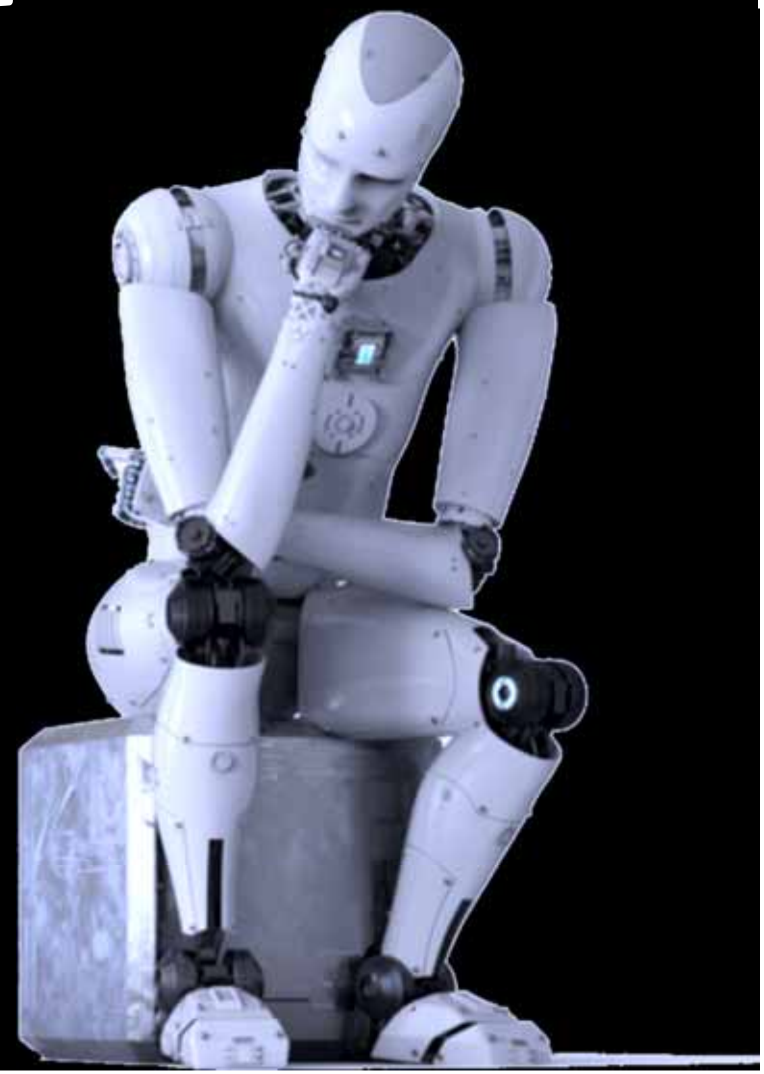
[news.gatech.edu]

Brain:

Temperature range: 36-37.5 °C
Avg heat removal: 15 mW/cm^3

THE OPPORTUNITY:
CROSS-CONTAMINATION

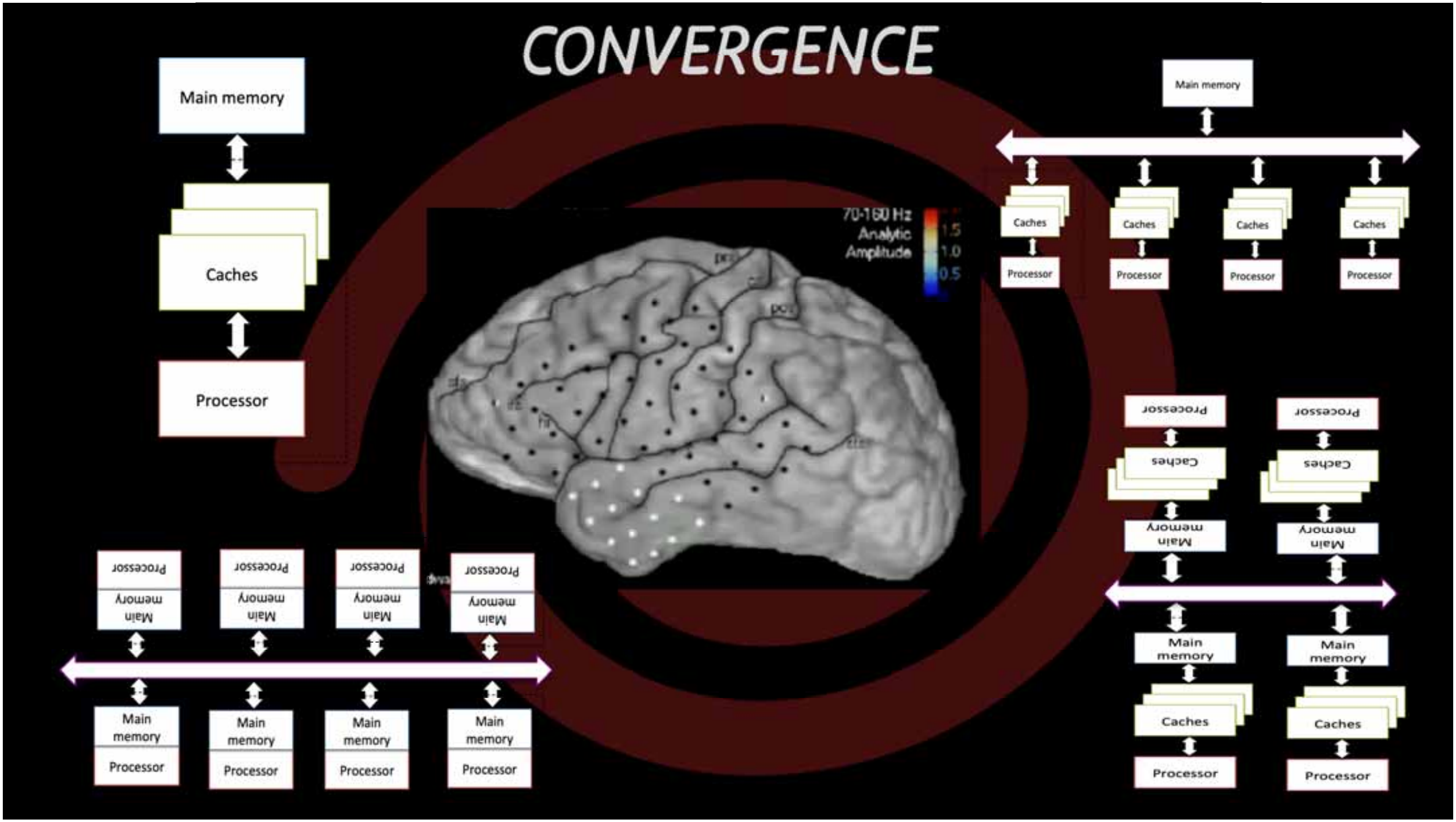
CONVERGENCE



CONVERGENCE

4-100 μm neurons
1.3-1.5 dm^3
~ 80 Billion Neurons
100 Trillion Synapses
20W Average
> 10,000 TOPS (??)

7 nm CMOS
425 mm^2
> 37 Billion Transistors
5-150 W
> 100 TOPS


CONVERGENCE
Main memory
Caches
Processor
70-160 Hz
Analytic
Amplitude
1.5
1.0
0.5
Main memory
Caches
Processor
Main memory
Processor

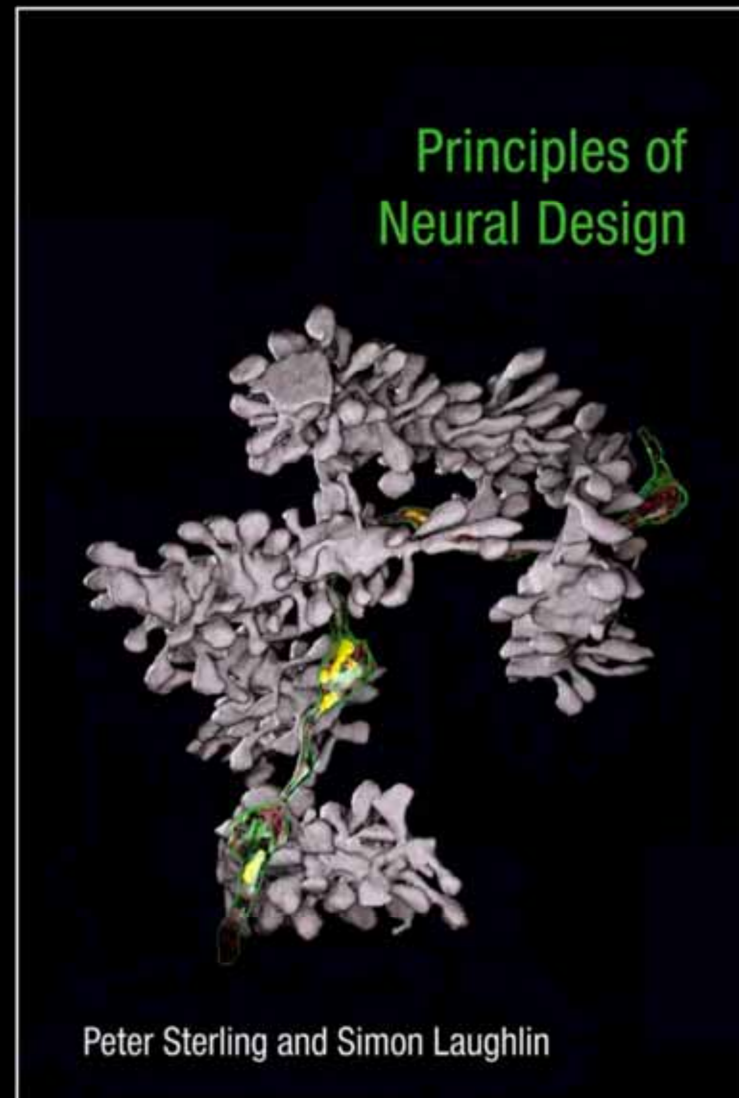
LESSONS FROM THE BRAIN

Caveat

'Biology is hiding secrets well. We just don't have the right tools to grasp the complexity of what's going on.'

— Bruno Olshausen

Wired Magazine, 2014

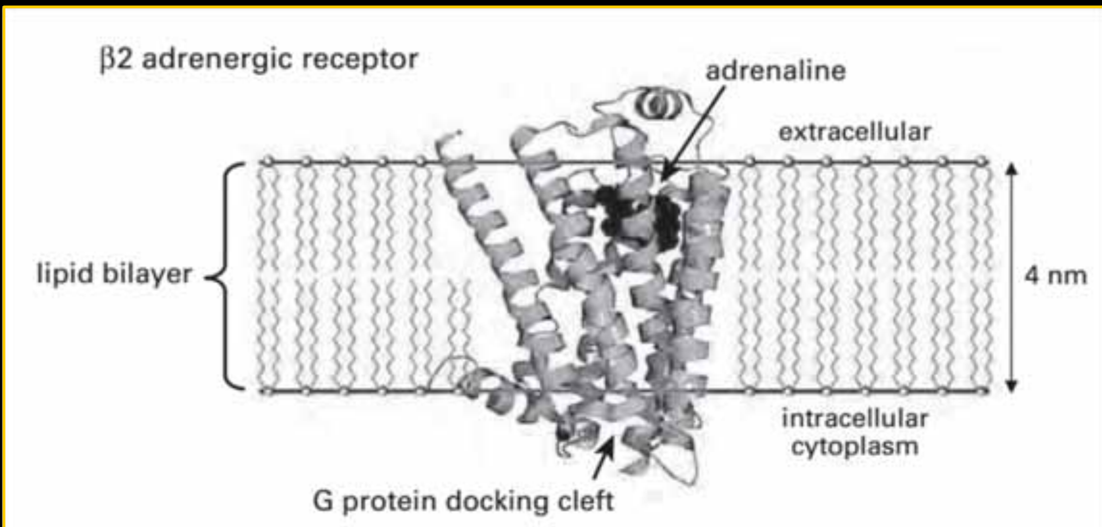


Neural Design Principles

"Compute with chemistry whenever possible"

From: Principles of Neural Design – P. Sterling, S. Laughlin

Computing with Proteins

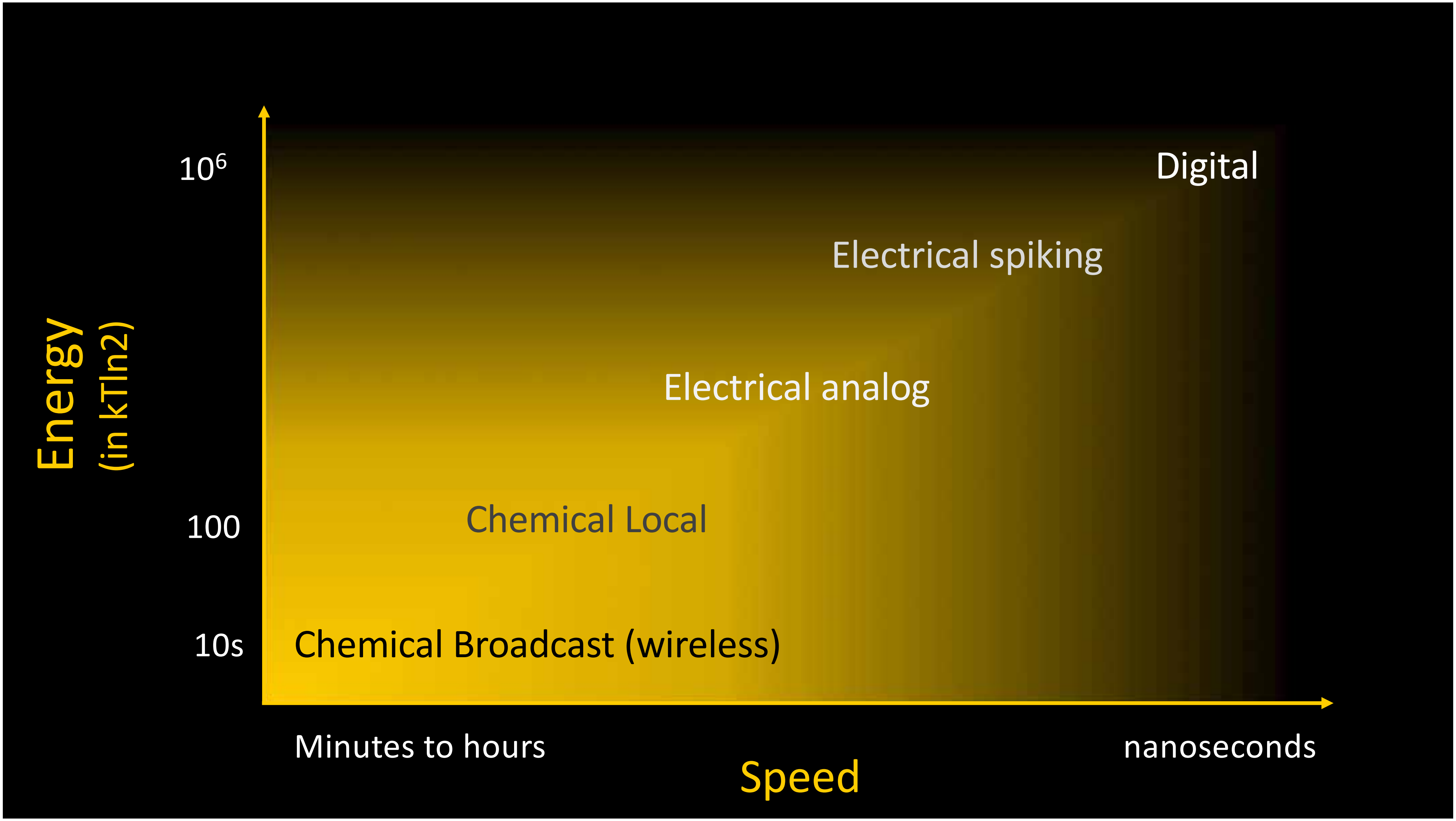


- Energy per transition: ~75 kT (e.g. opening an ion channel)
- Electrical energy flowing though ion channel (1 msec): 25 x 10^3 kT

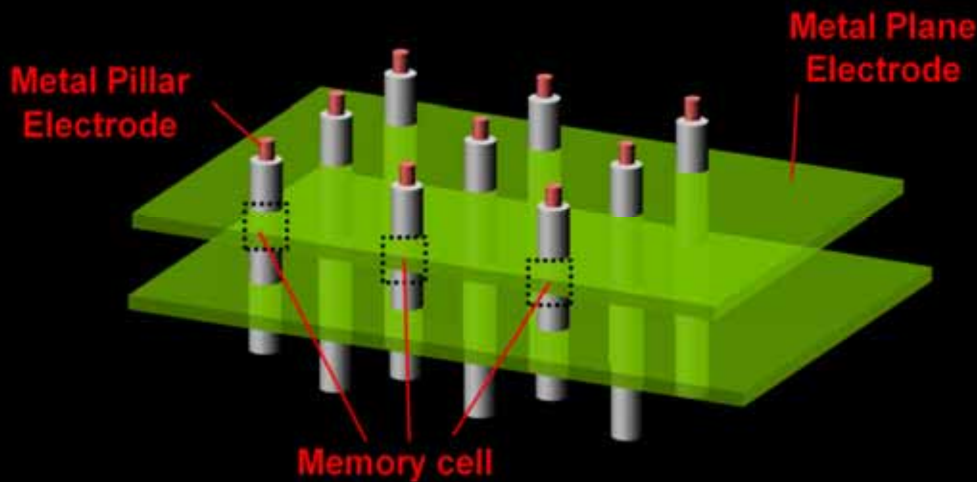
Supports analog (amplify, oscillate, integrate, differentiate, power, log, powers …) and digital (add, subtract multiply, divide)

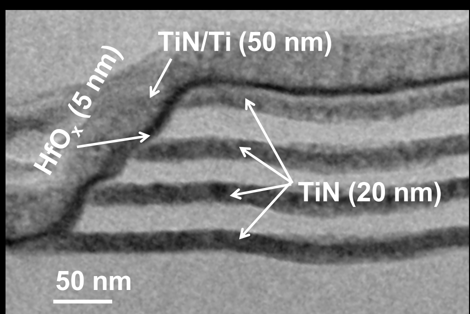
But:

- Slow: milliseconds to days …
- Noisy!



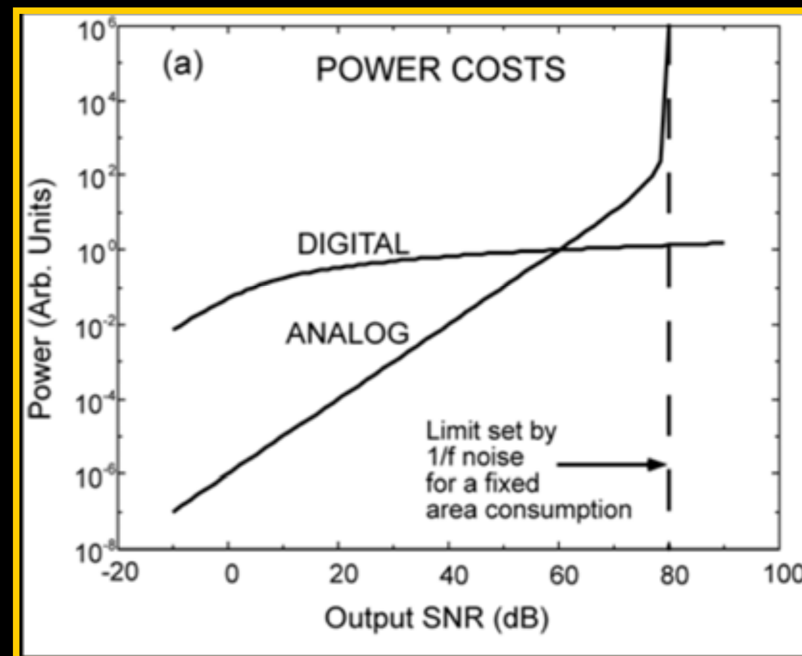
Energy
(in kTln2)
10^6
100
10s
Digital
Electrical spiking
Electrical analog
Chemical Local
Chemical Broadcast (wireless)
Minutes to hours
nanoseconds
Speed





[Courtesy: P. Wong]

PCRAM, RRAM



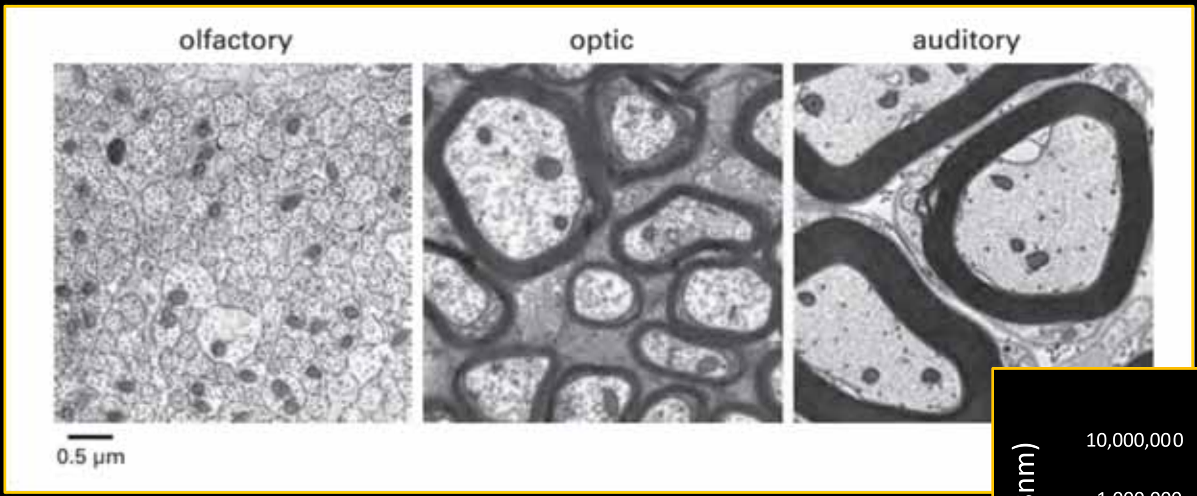
[Courtesy: P.C. Huang]

Neural Design Principles

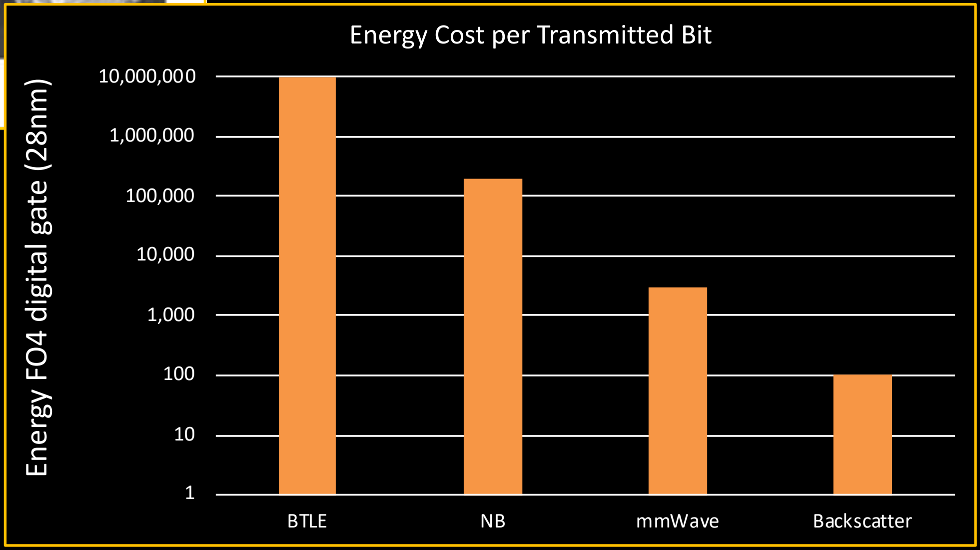
"Send only information that is needed, and send it as slowly as possible"

From: Principles of Neural Design – P. Sterling, S. Laughlin

Communication is expensive

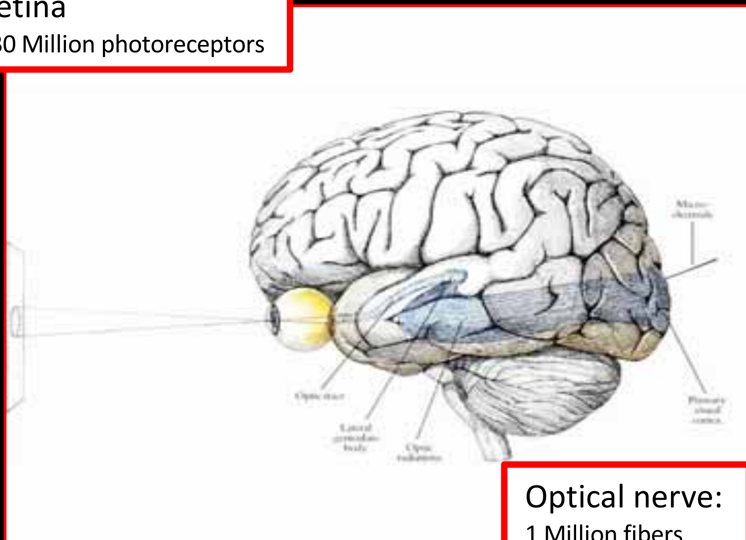


High speed comes at
area and energy cost



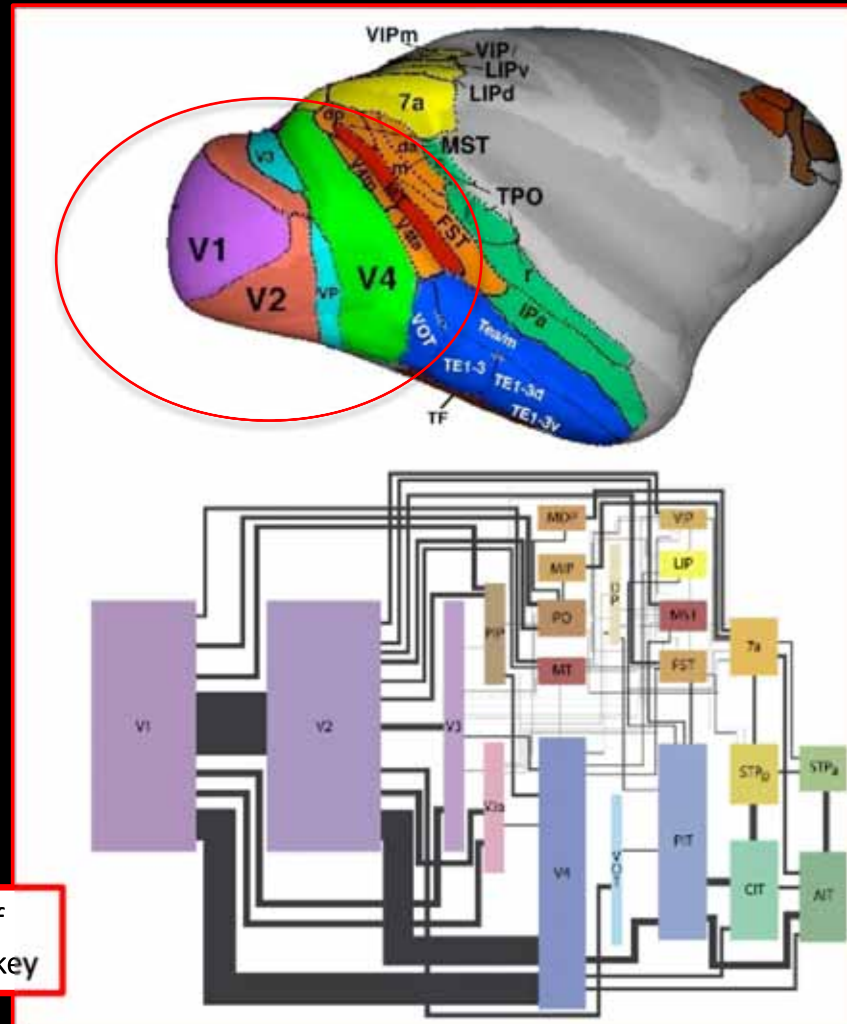
The visual sensory pathway

Retina
130 Million photoreceptors



Optical nerve:
1 Million fibers
10-100 Mb/sec

Data compression in retina

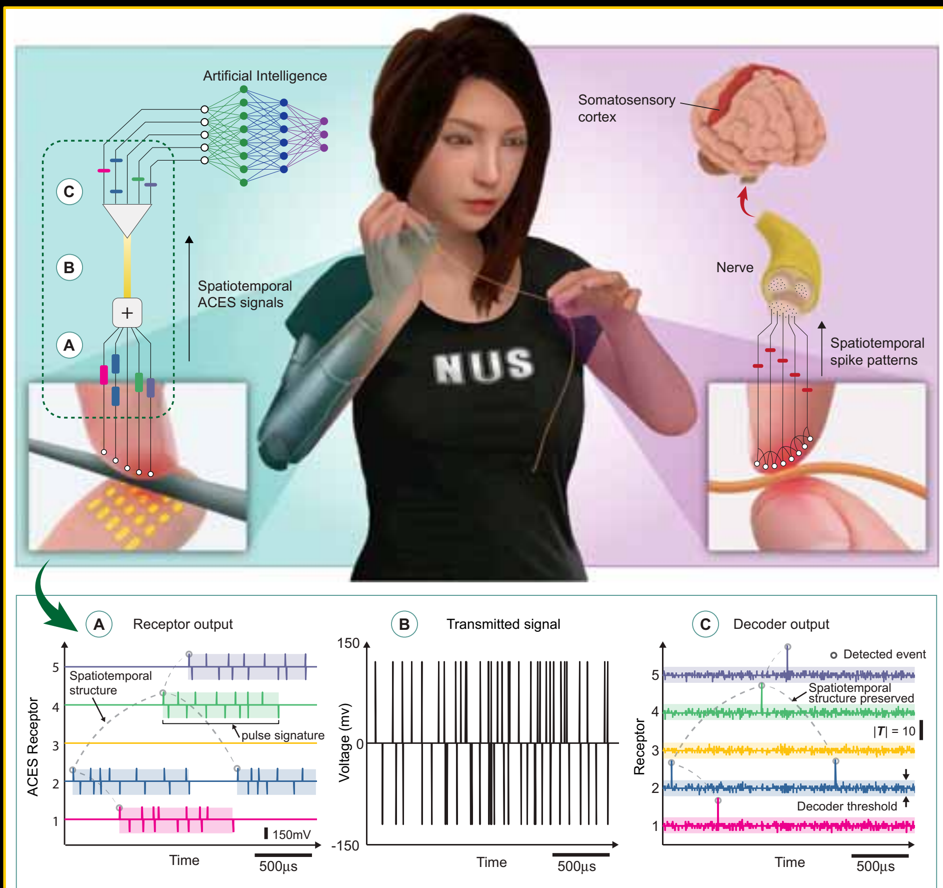


Massive expansion in V1 and V2

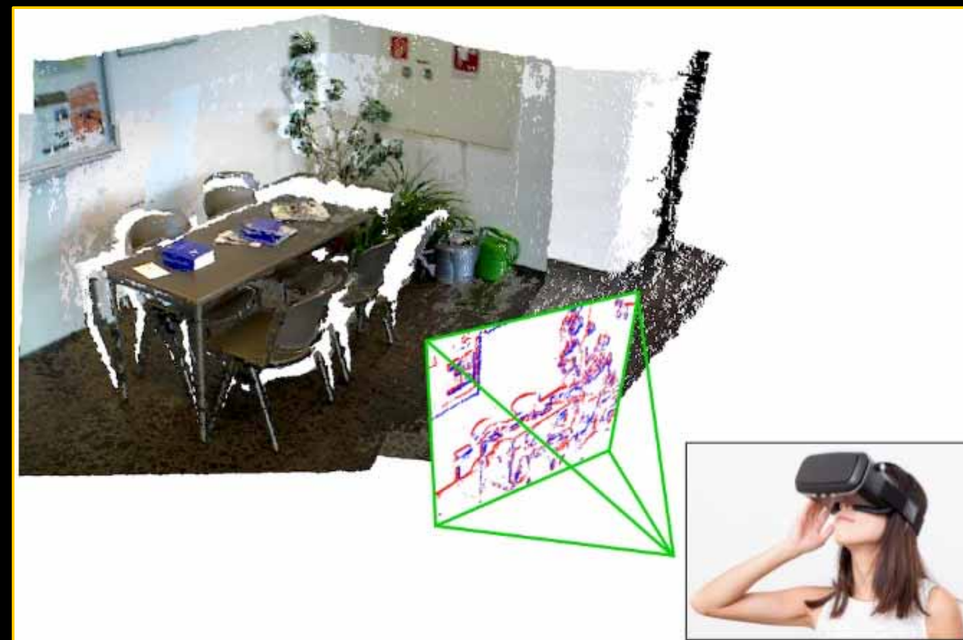
Visual cortex of Macaque monkey

[Courtesy: B. Olshausen, UCB]

Send only information that is needed



Neuromorphic electronic skin [B. Tee, NUS]



Event Camera [D. Scaramuzza, Univ of Zurich]

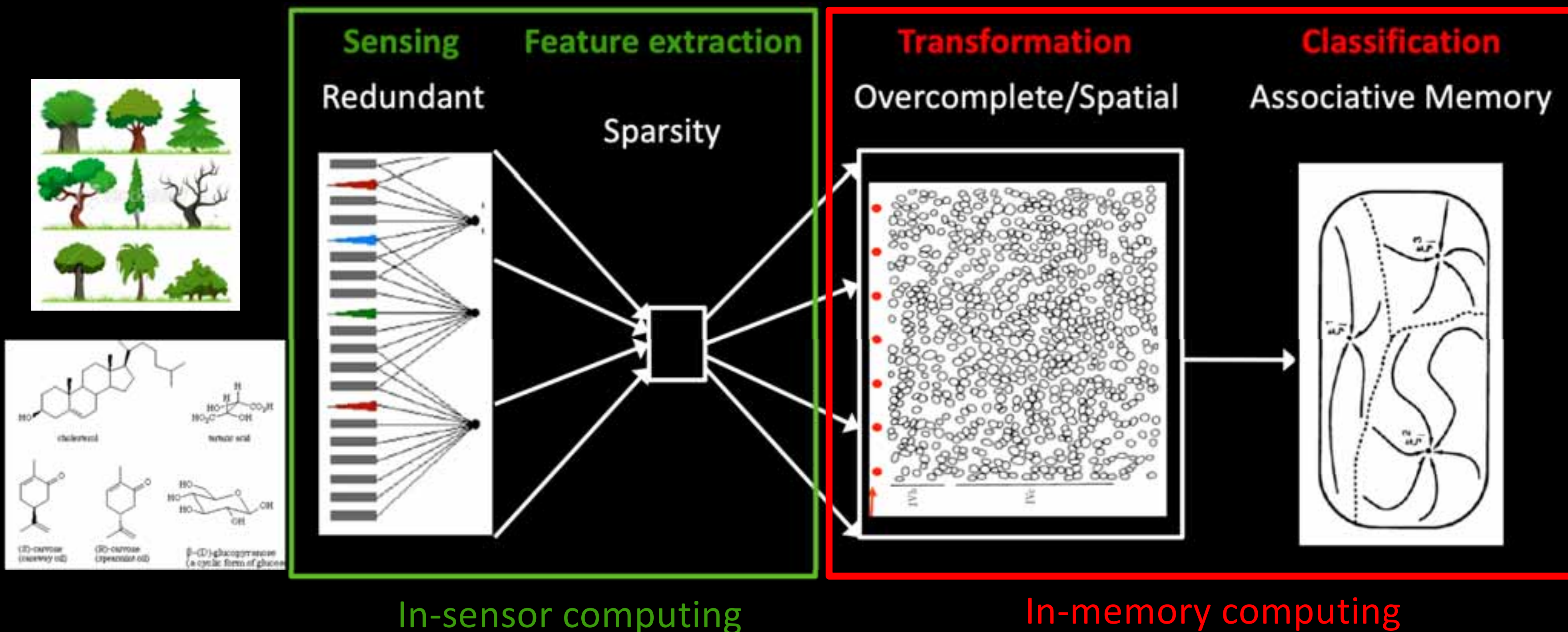
Neural Design Principles

*“To conserve space, time and energy, **new information** should be stored at the site where it is processed and from whence it can be recalled …”*

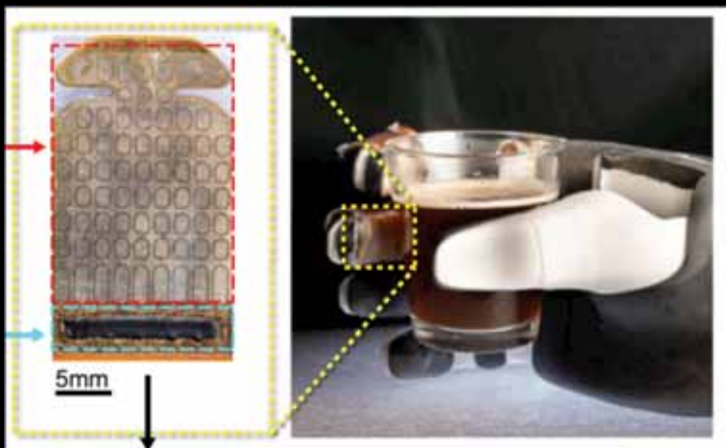
From: Principles of Neural Design – P. Sterling, S. Laughlin

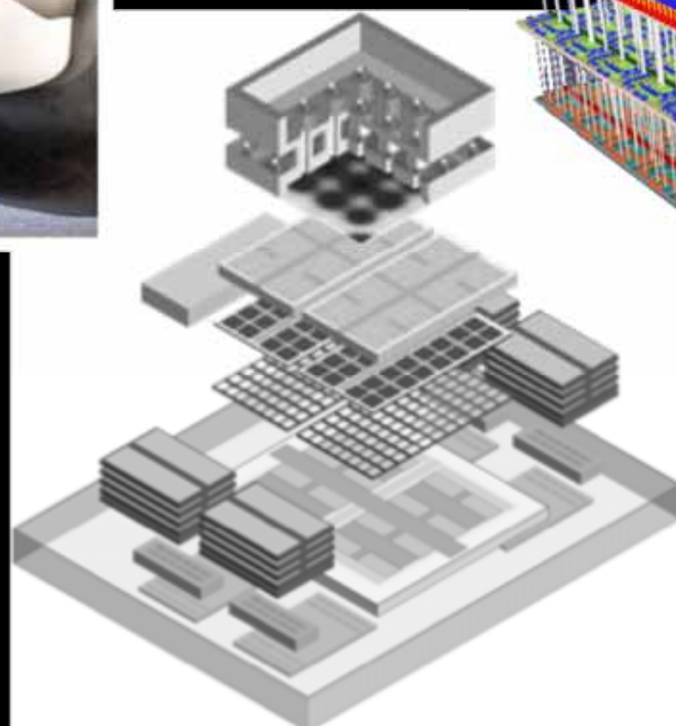
In-sensor and in-Memory Computing

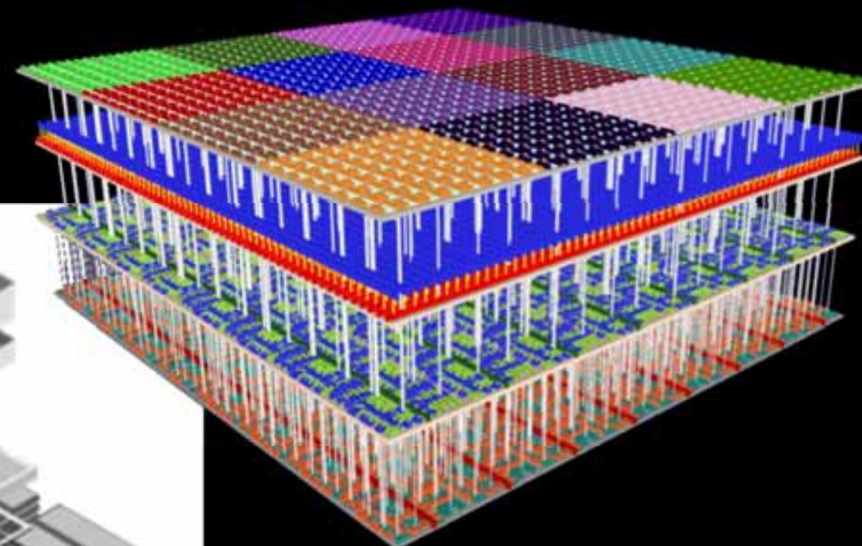
The Sensory Pathway – An engineering perspective



Intertwining sensing, processing and memory







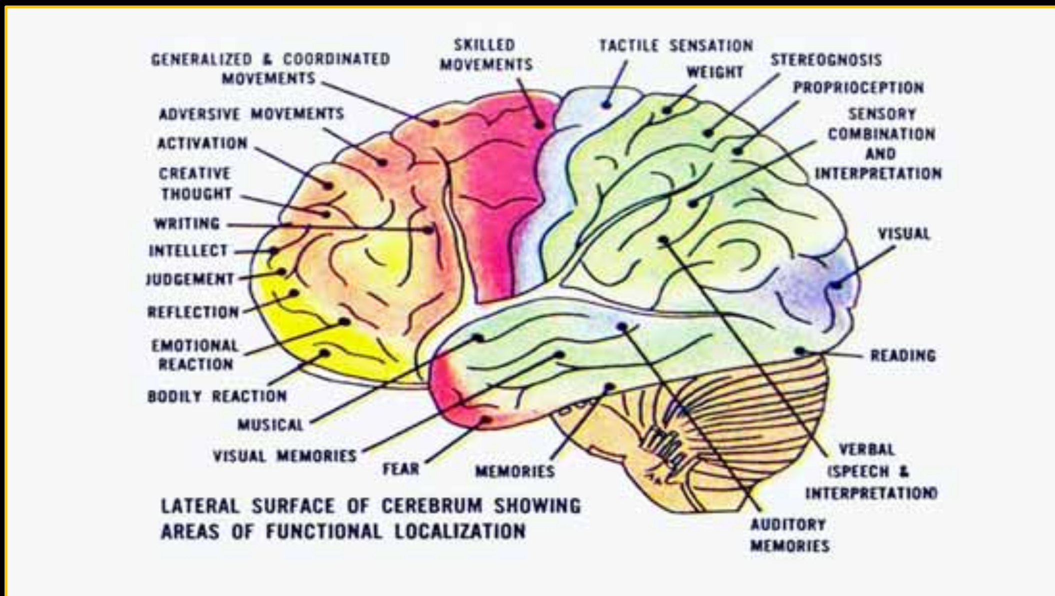
A new look at system partitioning and packaging

Neural Design Principles

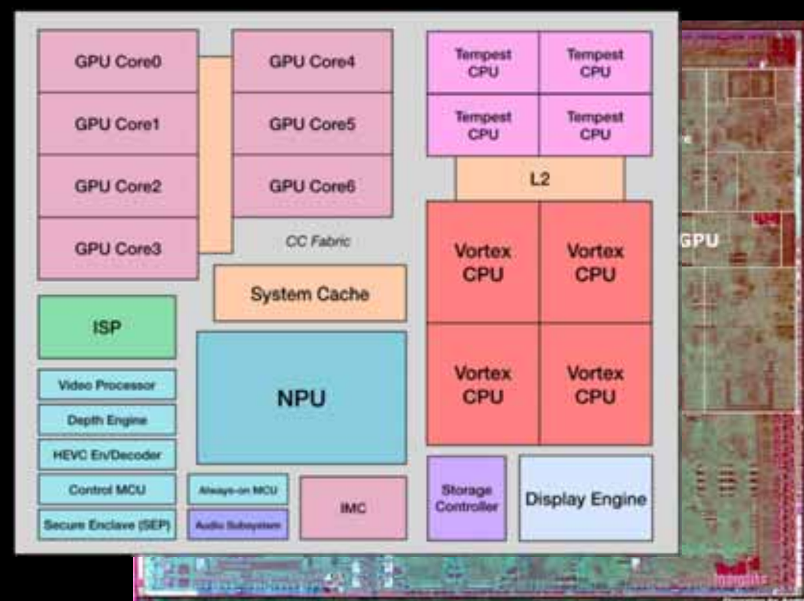
“Complicate”
(in engineering terms: Customize)

“if one design is simple and another complicated, choose the complicated”
(Glegg, 1969; Pahl et al., 2007).

Broad interpretation of Principles of Neural Design – P. Sterling, S. Laughlin

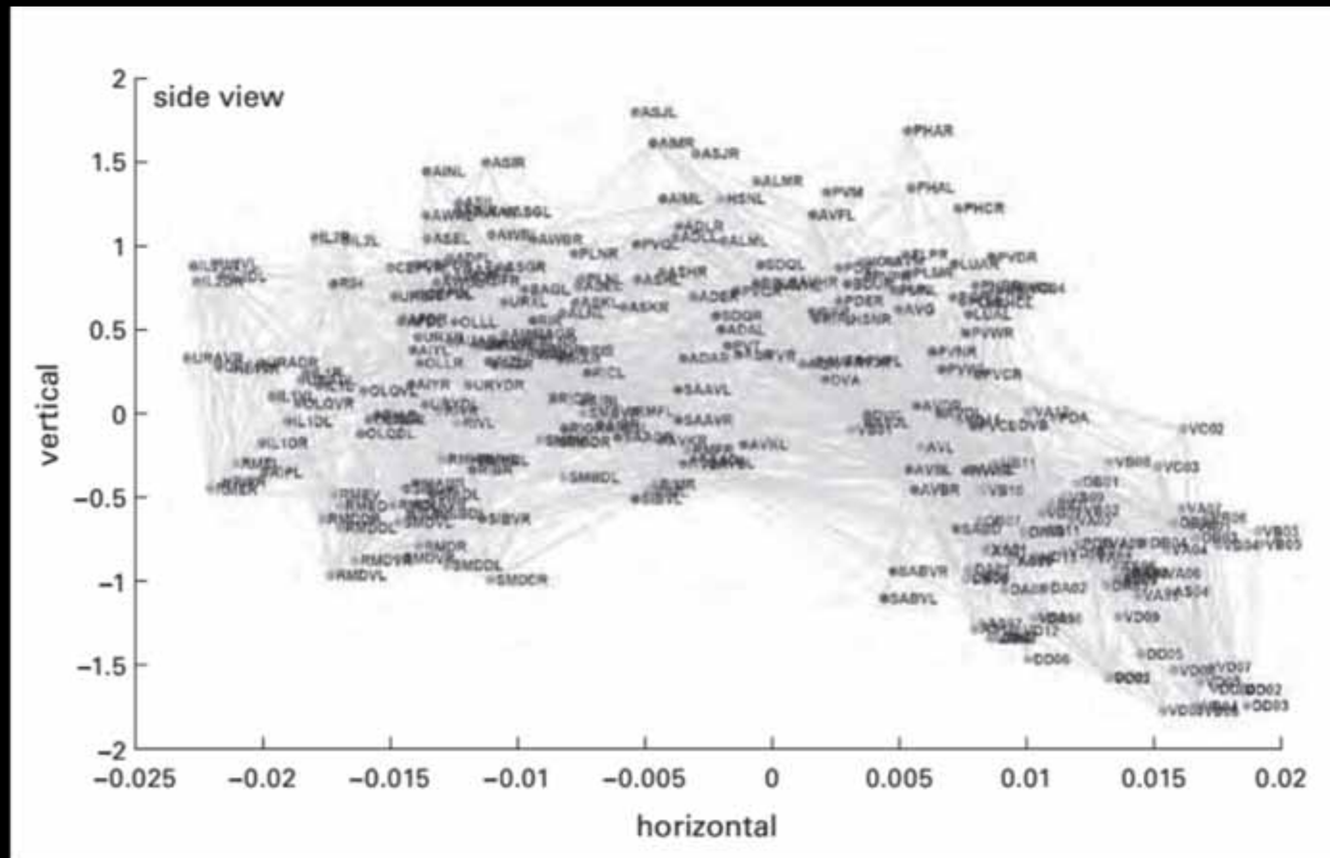


Apple A12 SOC



Placement of neurons in C. elegans

C. Elegans is a free-living transparent nematode about 1 mm in length that lives in temperate soil environments.



Optimal in reducing wire length (Varshney, 2012)

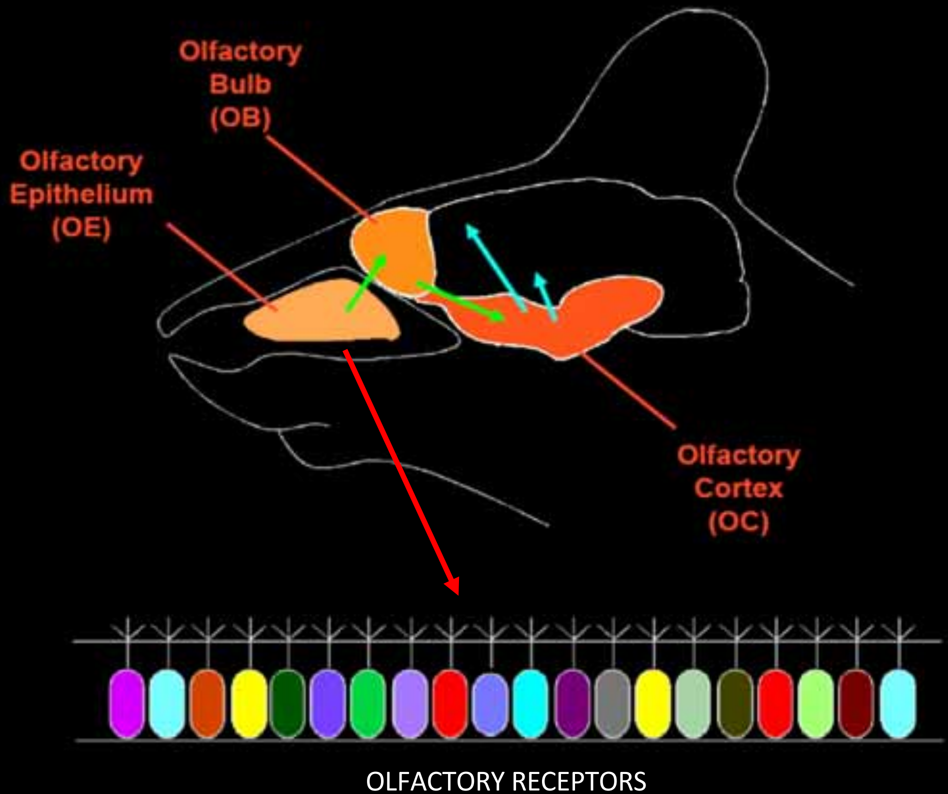
Neural Design Principles

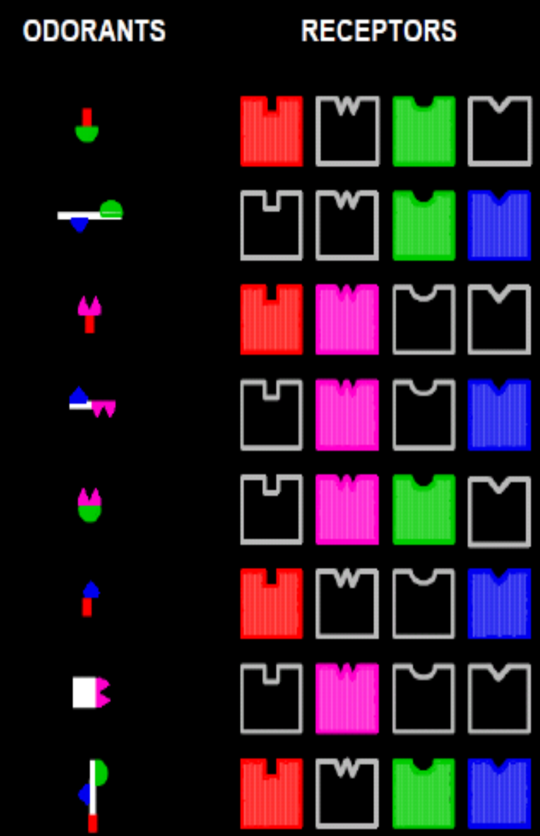
"Play the number game"

or

"embrace randomness"

Dealing with Low SNR and Variability

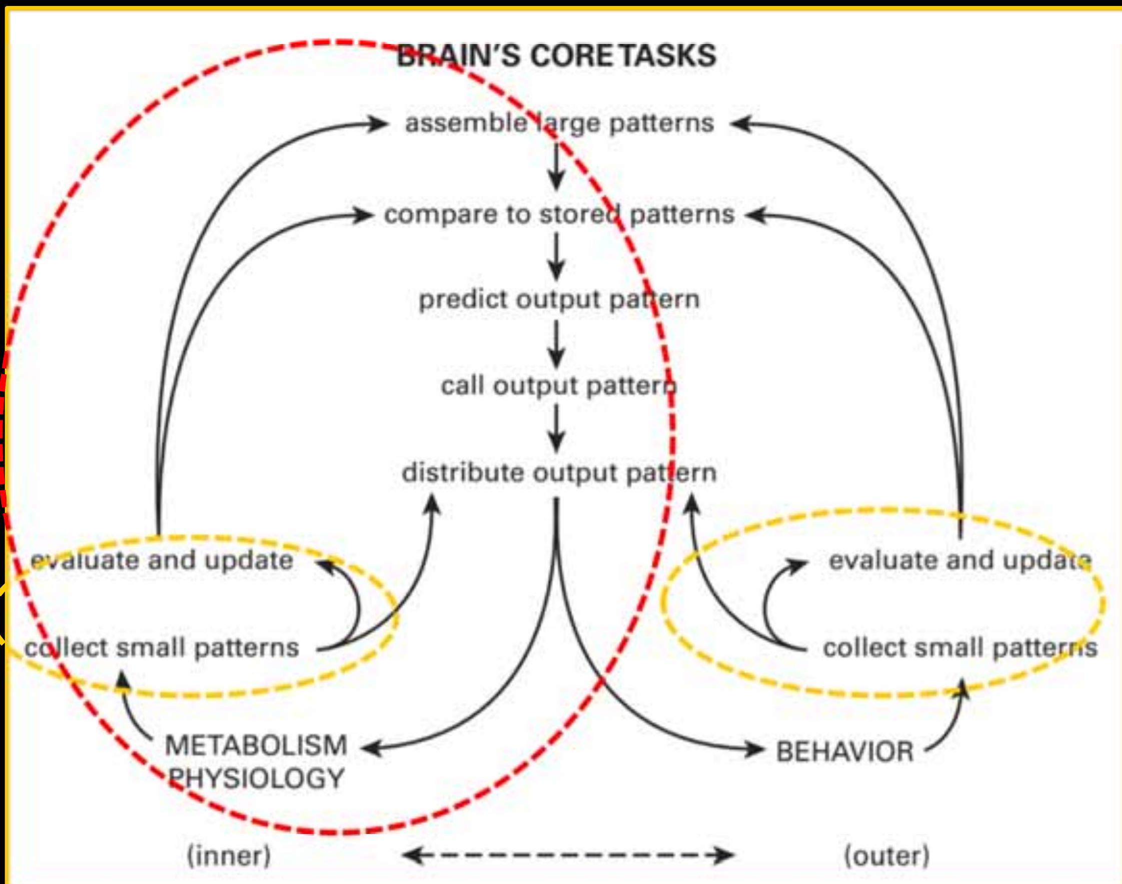


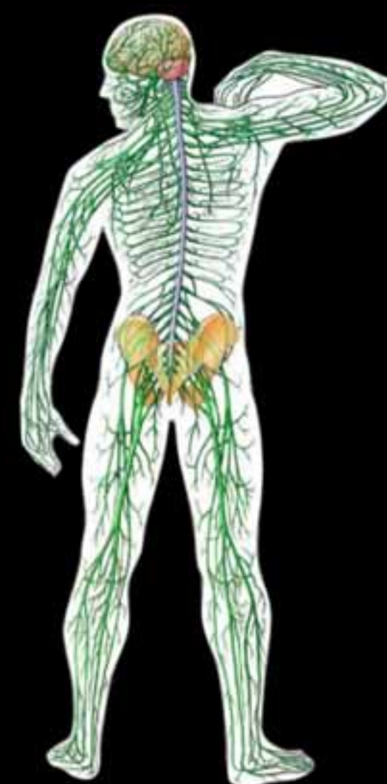


Population Coding: Diversity and redundancy

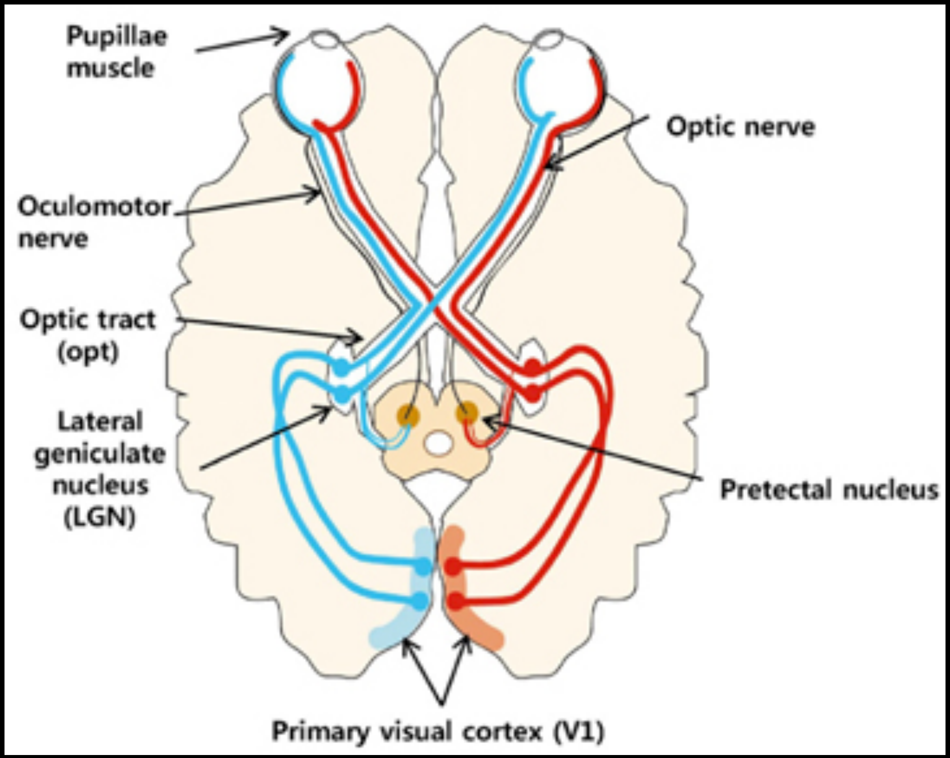
Neural Design Principles

"Self-calibrate,
adapt and heal"


BRAIN'S CORE TASKS
assemble large patterns
compare to stored patterns
predict output pattern
call output pattern
distribute output pattern
evaluate and update
collect small patterns
evaluate and update
collect small patterns
METABOLISM
PHYSIOLOGY
BEHAVIOR
(inner)
(outer)



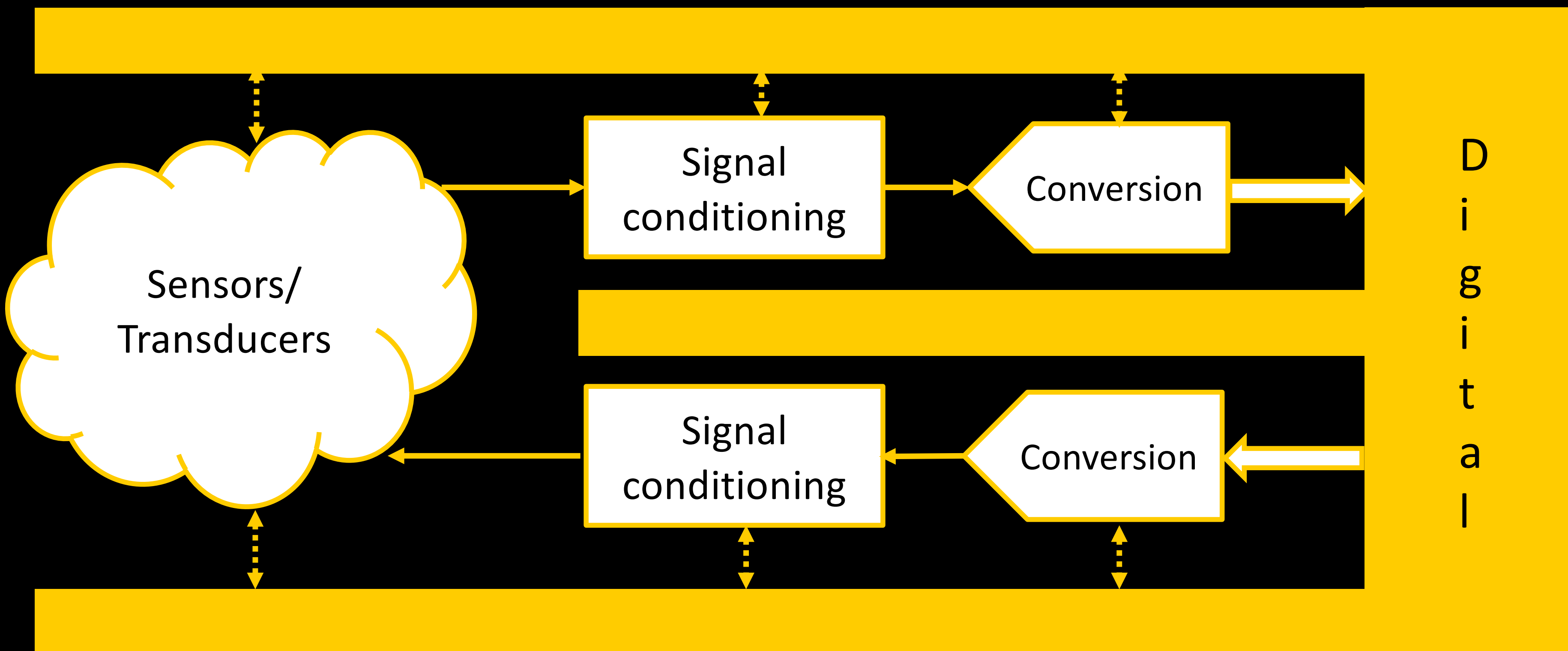
Maximizing sensory efficiency (auto-tuning)



Feedback loop (colliculus) orients eye (light) to highest-resolution area of retina (fovea)

Combines high-level intent with observation

Digitally-Assisted Mixed-Signal



Some musings

Time for a New (Moore's) Law?

- Nature and evolution have found in the biological human brain a (optimal?) solution that balances size, power budget, cell complexity and size, wiring density, and robustness
- Given our knowledge about physics, materials, nano-devices, noise and energy constraints, can't we derive what such an optimal solution could look like for a physical computer?
- Or is this too simplistic??

$E > kT\ln 2$

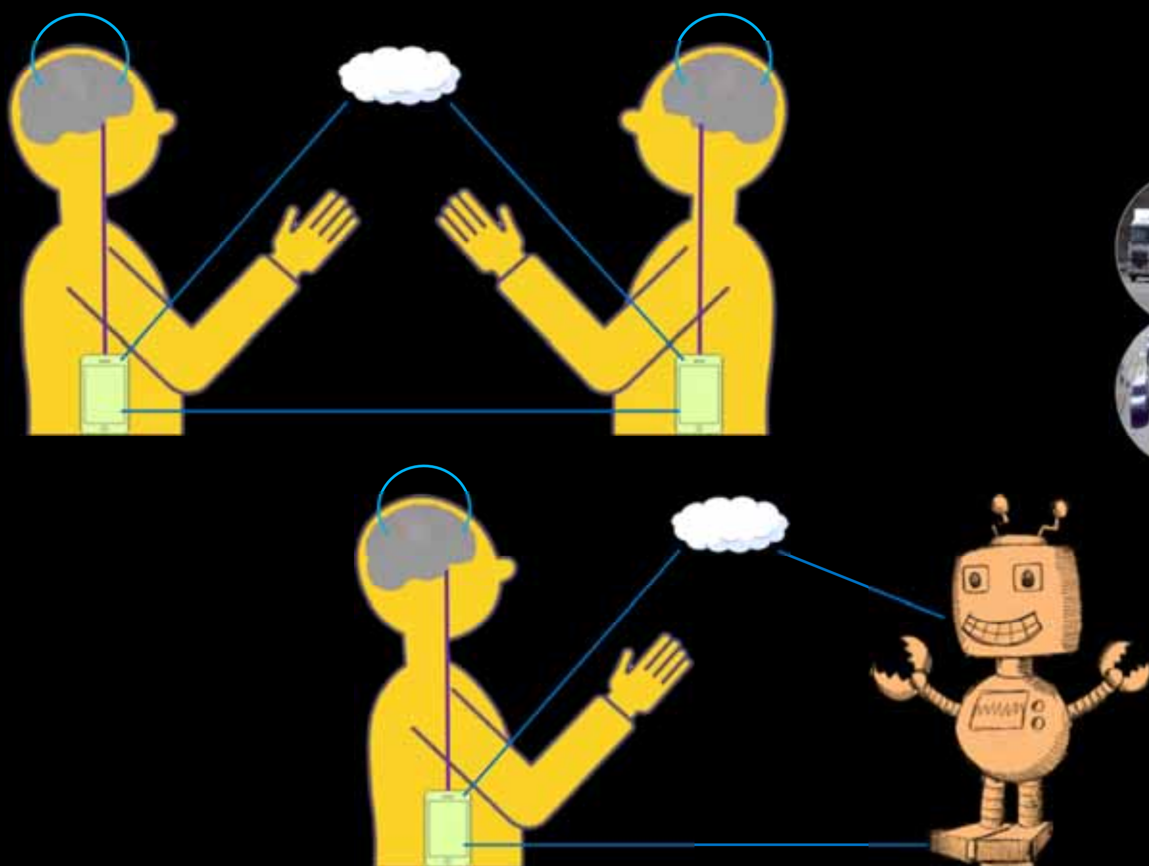
The great disconnect

10^6: Ratio between computer and brain frequencies (Brain in sync with nature)

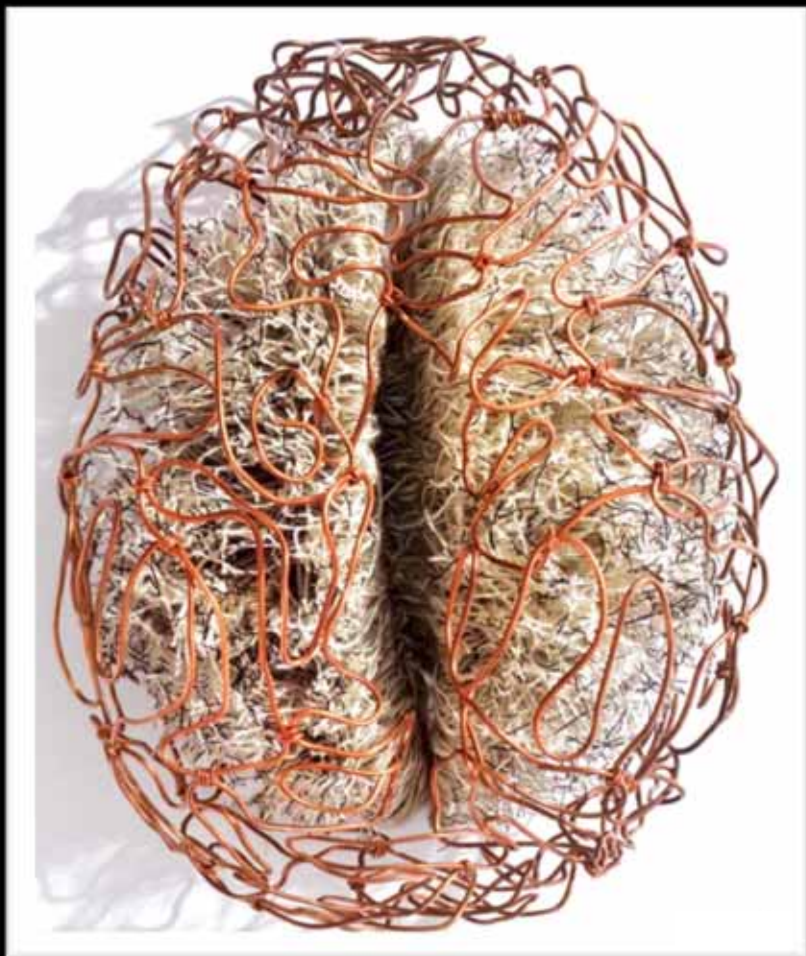
$E > kT \ln 2$

Computers to outpace brains and nature, or …

The great disconnect, really?



The true convergence



Brain sculpture by Maggie McDonald

Thank you!