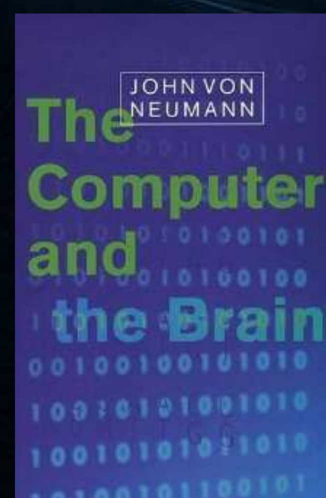
NANOSEMINAR TUDRESDEN 11 JULY 2013

AND THE COMPUTER

WHY BIOLOGY CAN AND SILICON … CAN'T

VALERIU BEIU

جامعة الإمارات العربية المتحدة
United Arab Emirates University



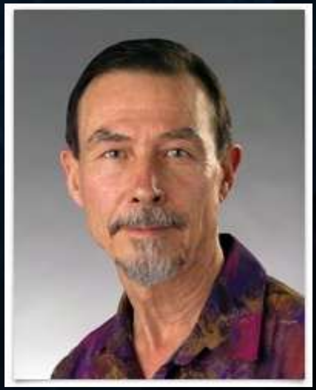
Structure

- Motivation
- 101 ... on Brain's "single ion transistors" et al.
  - Power … device(s) … wire(s) … system
  - Reliability … device(s) … wire(s) … system
- Architectural organizations
  - Low level … arrays (of ion channels)
  - Intermediate … columnar structures
  - Higher levels … sparsely connected (hierarchical)
- Conclusions
  - Device-level redundancy
  - Non-linear (active) "wires" (solitonic-like waves)
  - Communication/computation separated form power consumption

It is clear to me
that we will develop
silicon neural systems
and that learning
how to design them

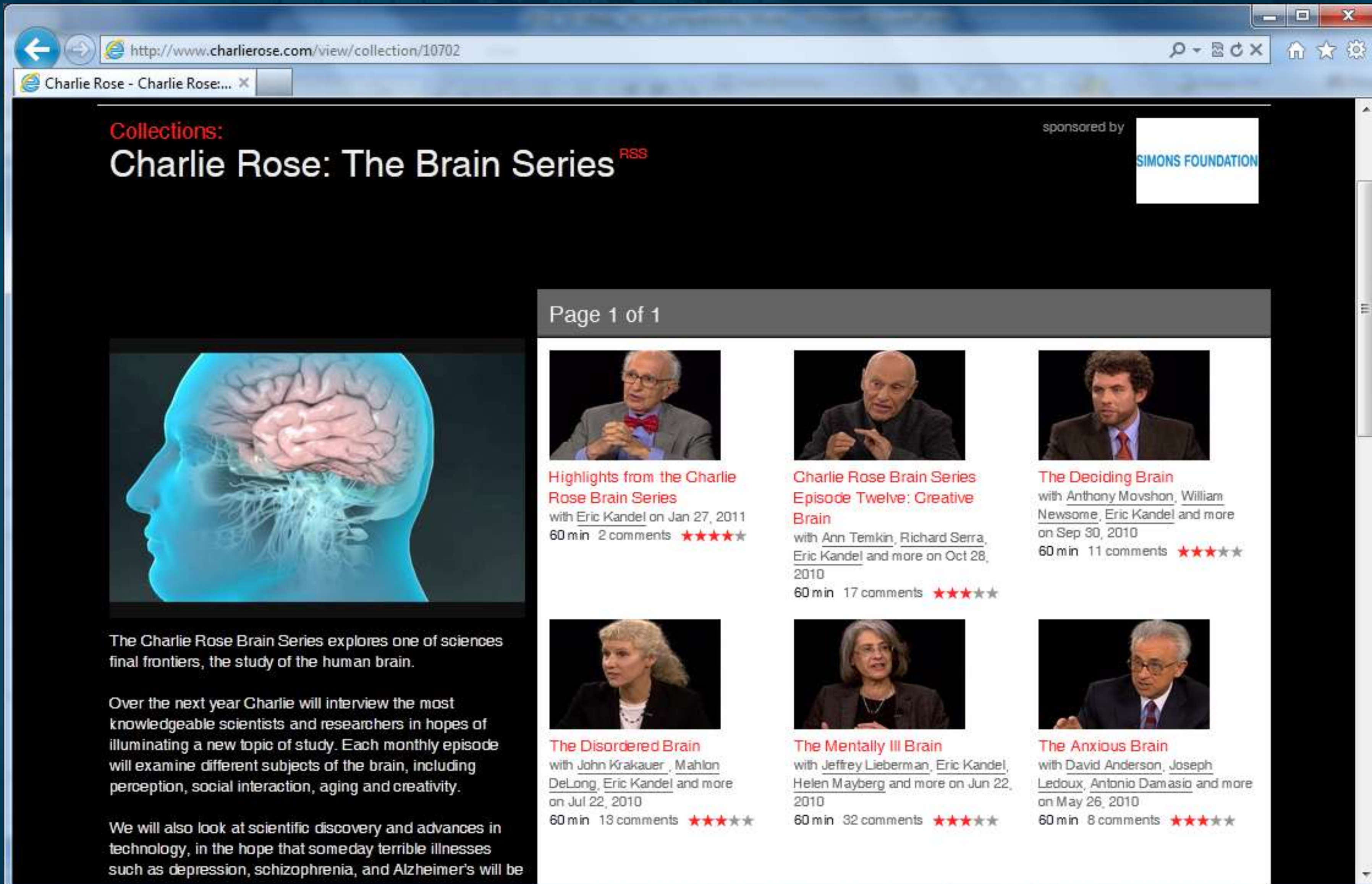
is

one of the greatest
intellectual quests
of all times

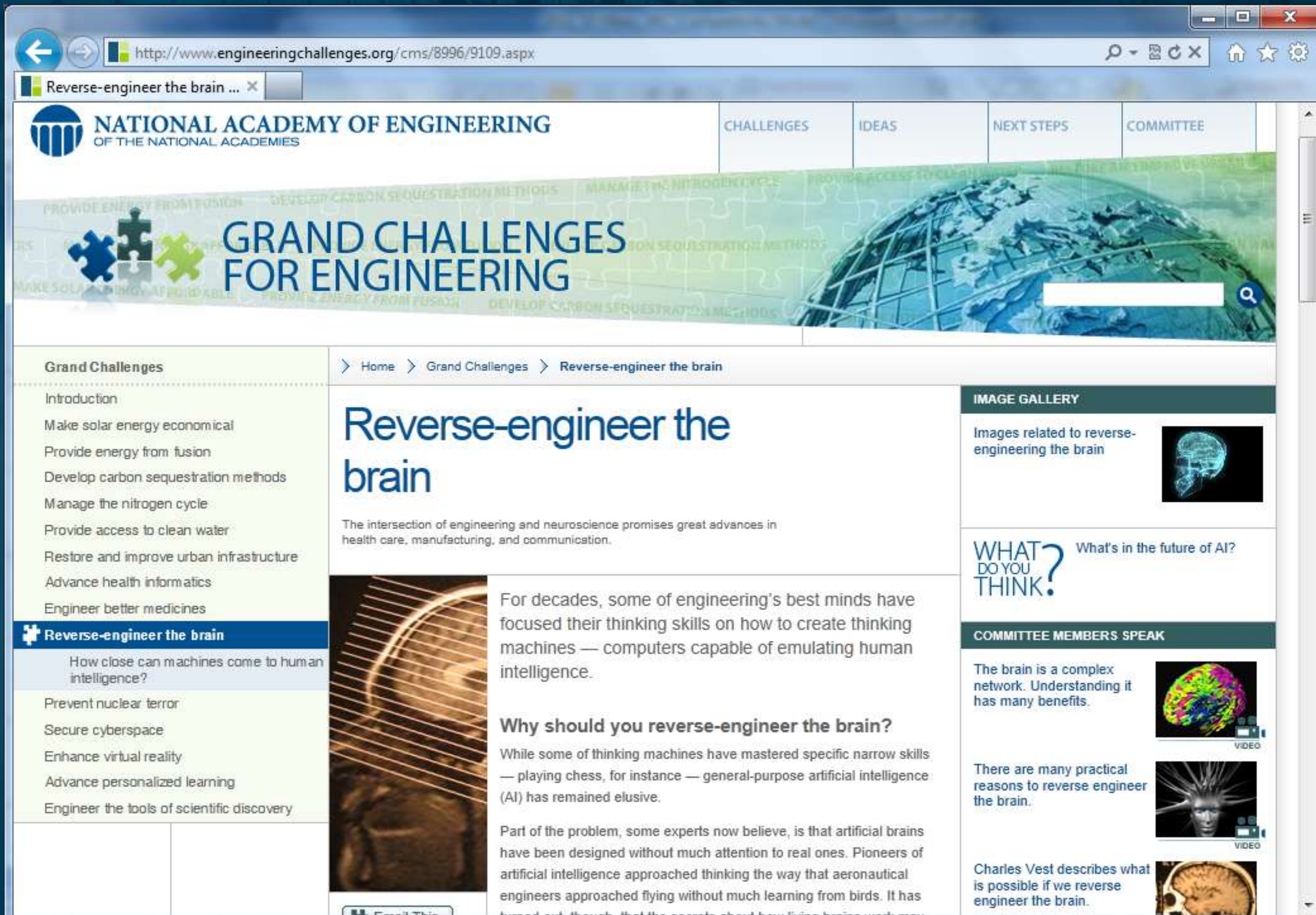


Carver Mead

… this is the century of the Brain



... it's a grand challenge


http://www.engineeringchallenges.org/cms/8996/9109.aspx

Reverse-engineer the brain ...

NATIONAL ACADEMY OF ENGINEERING
OF THE NATIONAL ACADEMIES

CHALLENGES | IDEAS | NEXT STEPS | COMMITTEE

GRAND CHALLENGES FOR ENGINEERING

Grand Challenges
- Introduction
- Make solar energy economical
- Provide energy from fusion
- Develop carbon sequestration methods
- Manage the nitrogen cycle
- Provide access to clean water
- Restore and improve urban infrastructure
- Advance health informatics
- Engineer better medicines
- **Reverse-engineer the brain**
  - How close can machines come to human intelligence?
- Prevent nuclear terror
- Secure cyberspace
- Enhance virtual reality
- Advance personalized learning
- Engineer the tools of scientific discovery

Home > Grand Challenges > Reverse-engineer the brain

Reverse-engineer the brain

The intersection of engineering and neuroscience promises great advances in health care, manufacturing, and communication.

For decades, some of engineering's best minds have focused their thinking skills on how to create thinking machines — computers capable of emulating human intelligence.

Why should you reverse-engineer the brain?

While some of thinking machines have mastered specific narrow skills — playing chess, for instance — general-purpose artificial intelligence (AI) has remained elusive.

Part of the problem, some experts now believe, is that artificial brains have been designed without much attention to real ones. Pioneers of artificial intelligence approached thinking the way that aeronautical engineers approached flying without much learning from birds. It has

IMAGE GALLERY

Images related to reverse-engineering the brain

WHAT DO YOU THINK? What's in the future of AI?

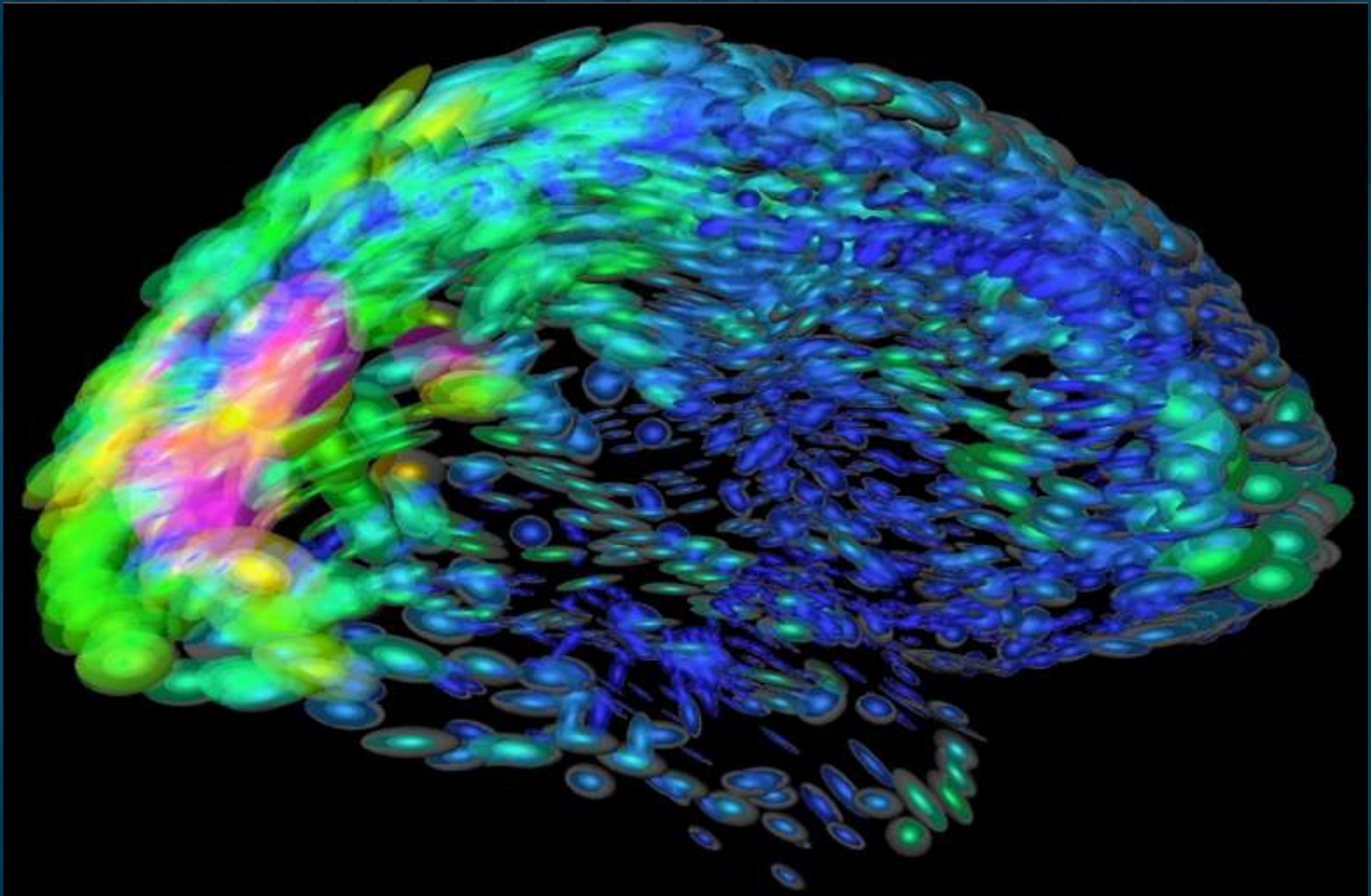
COMMITTEE MEMBERS SPEAK

The brain is a complex network. Understanding it has many benefits. VIDEO

There are many practical reasons to reverse engineer the brain. VIDEO

Charles Vest describes what is possible if we reverse engineer the brain.

The Brain

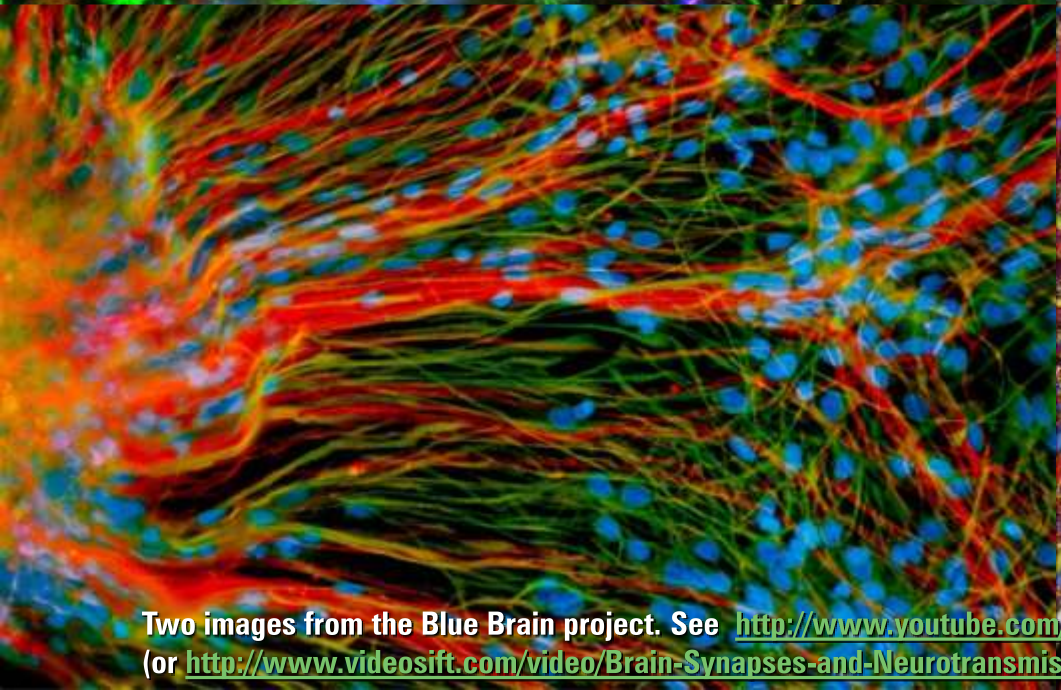


Let's take a walk through …

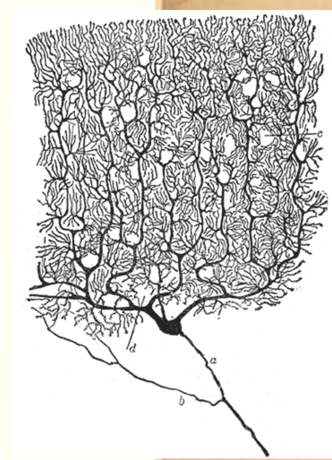
- … the "Machinery of Mind"

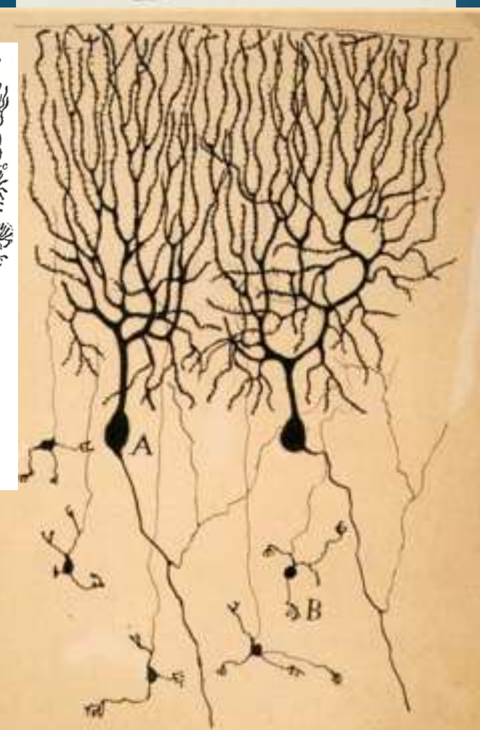
  http://smithlab.stanford.edu/Smithlab/AT_Movies_files/Machinery_of_Mind_SD.mov

Measurements … and simulations

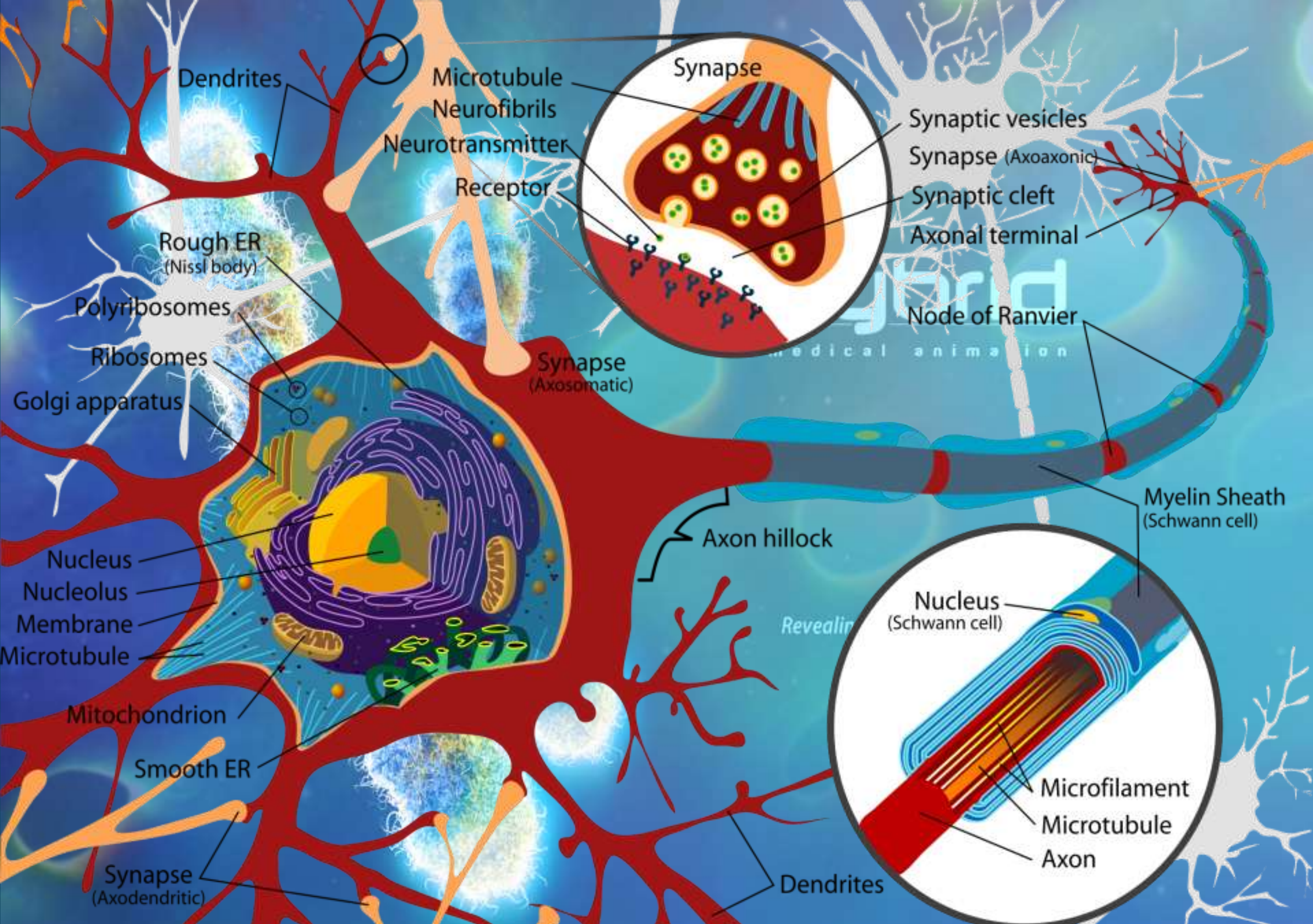


Two images from the Blue Brain project. See http://www.youtube.com/watch?v=90cj4NX87Yk
(or http://www.videosift.com/video/Brain-Synapses-and-Neurotransmission-3D-Animation)



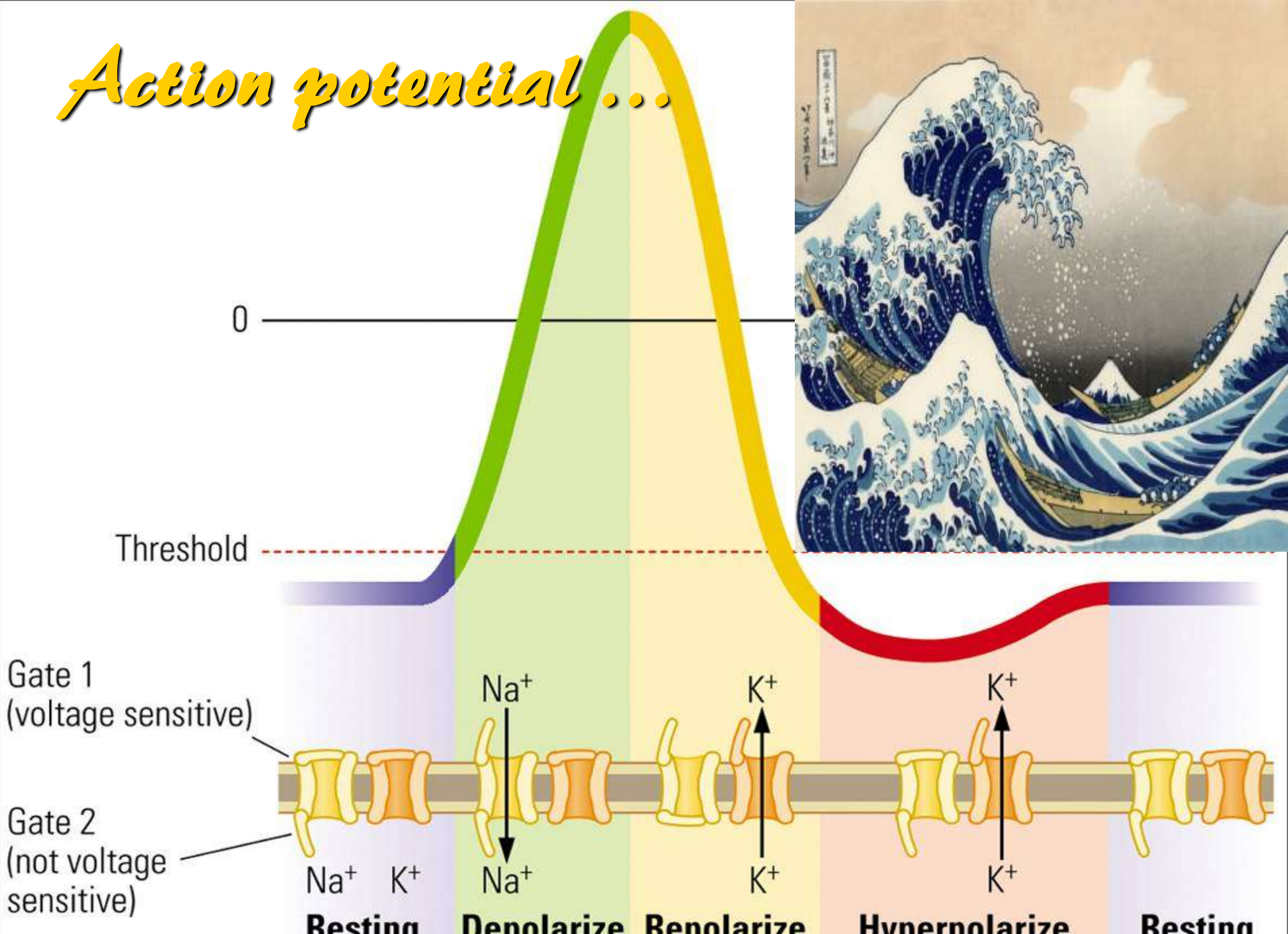


Many different biological neurons … corresponding to cores



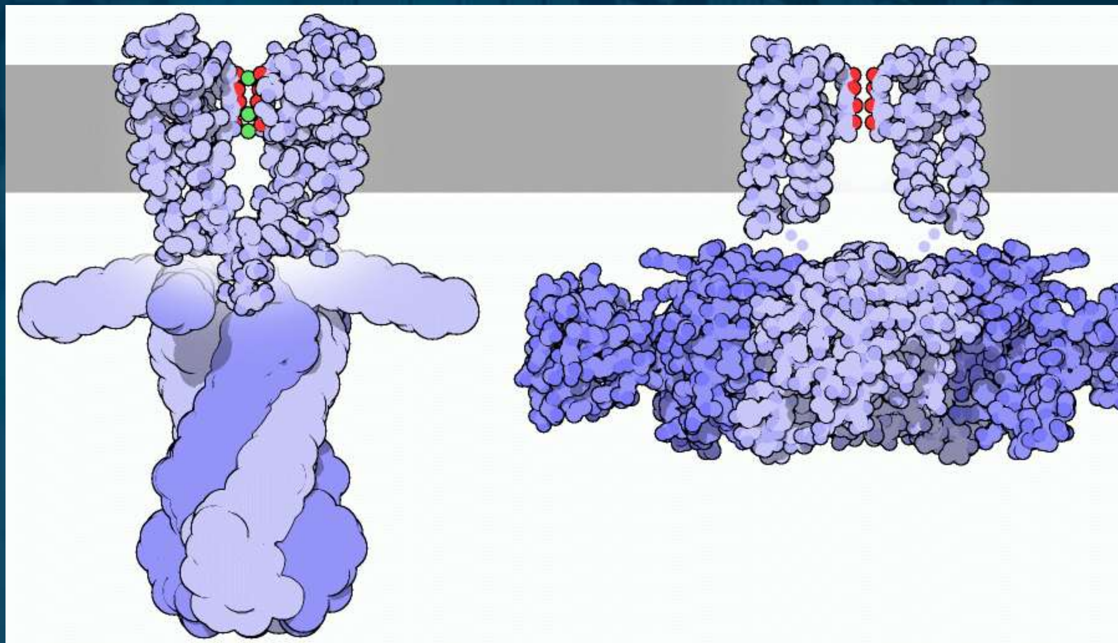
Drawing from Wikipedia http://en.wikipedia.org/wiki/Neuron
Background from Hybrid Medical Animation http://www.hybridmedicalanimation.com/

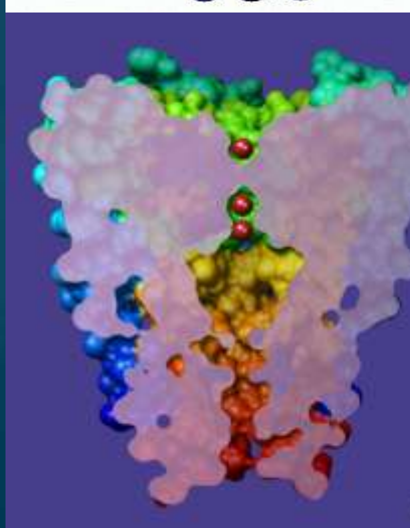
Action potential …

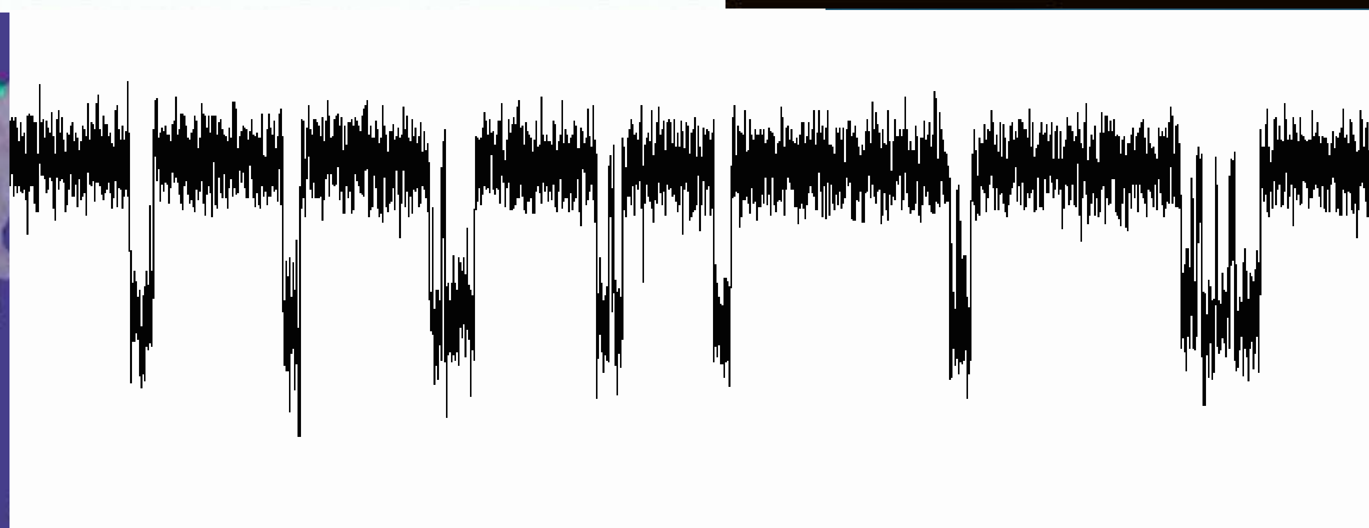


See http://www.youtube.com/watch?v=DJe3_3XsBOg (from 2:08 till 3:30)

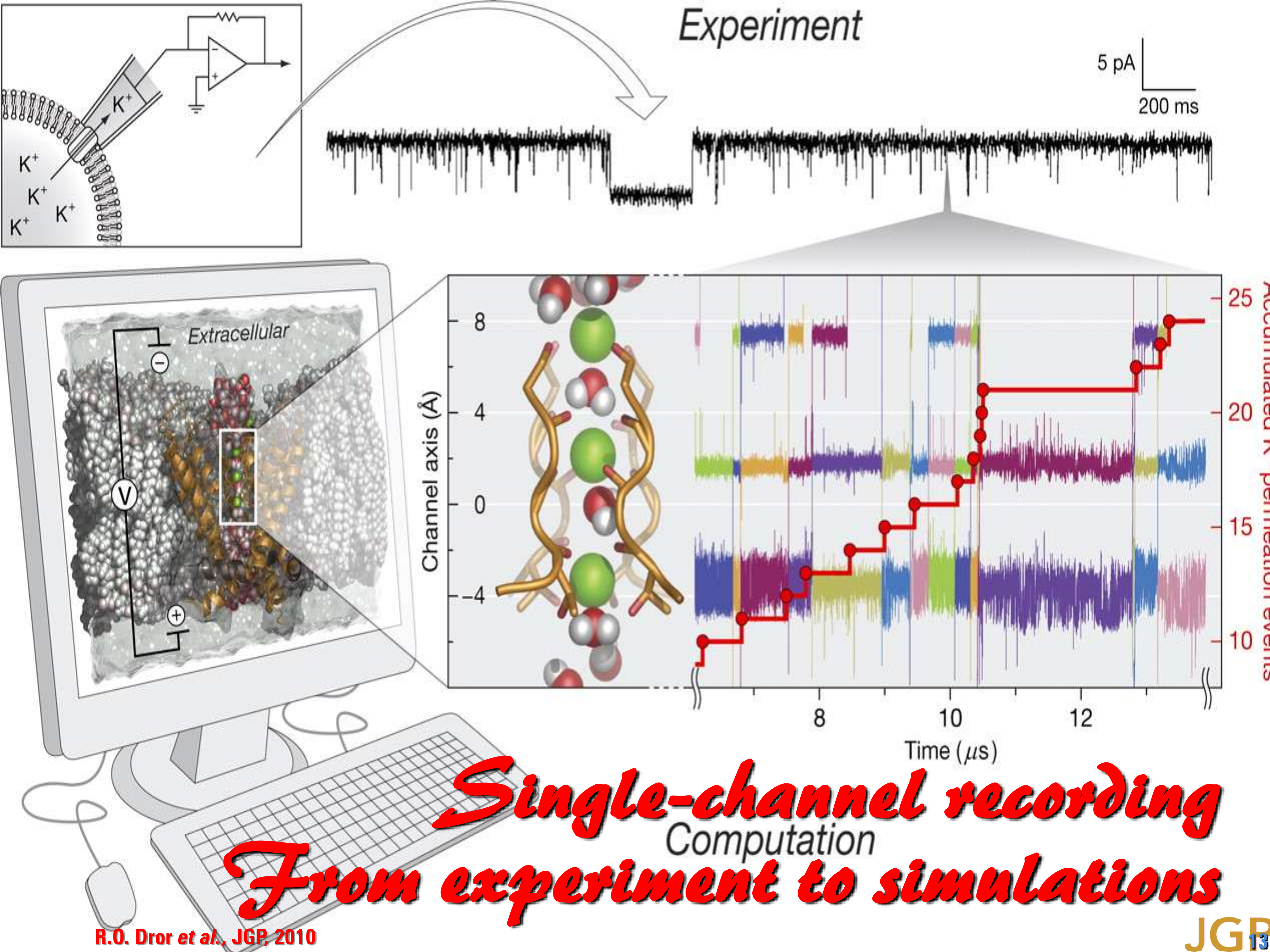
Brain's nano-switches





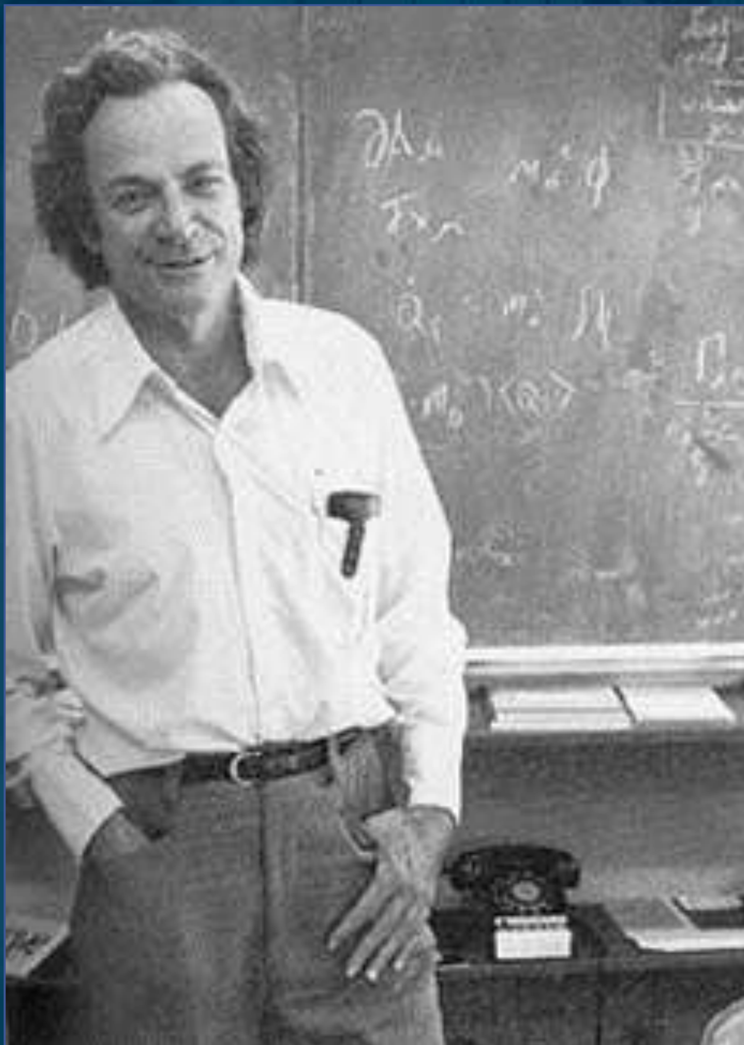


Single-channel recording
From experiment to simulations



R.O. Dror *et al.*, JGP, 2010

Biomorphic



"In an actual cell, the pyrophosphate concentration is kept low by hydrolysis, ensuring that only the copying process occurs, not its inverse. The whole RNA polymerase system is not particularly efficient as far as energy use goes:

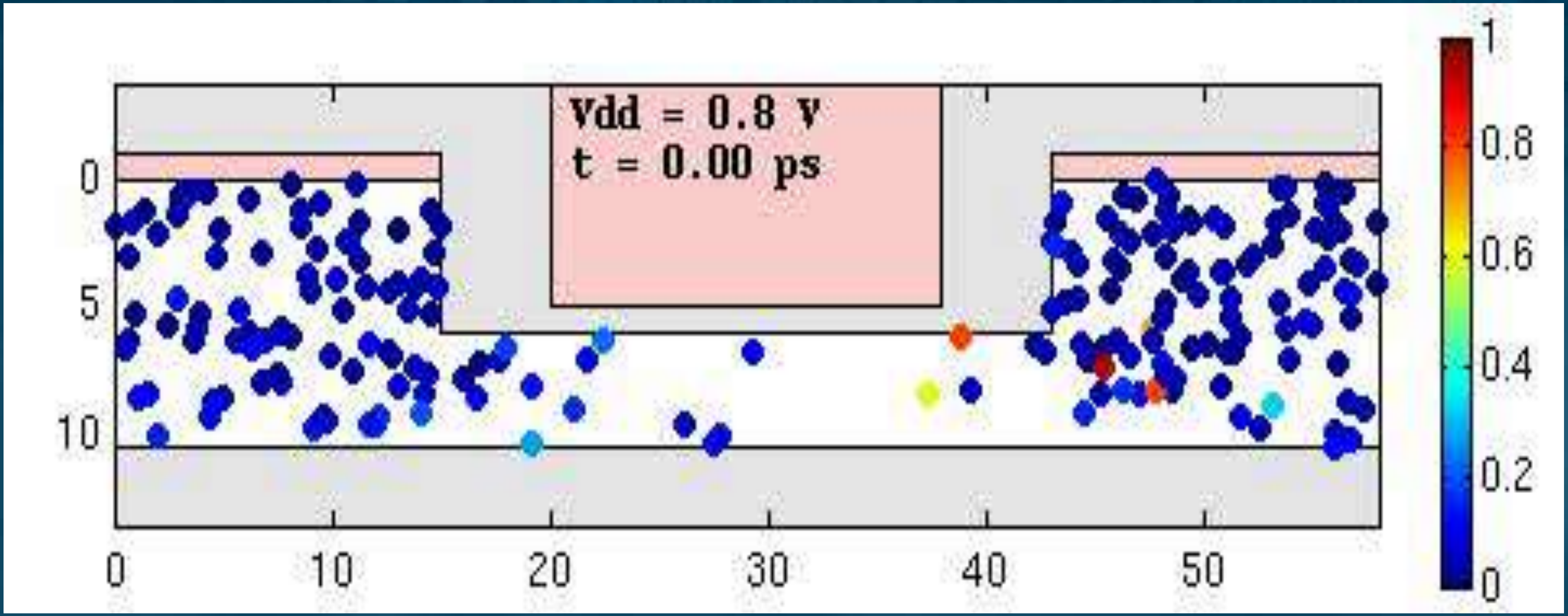
… it dissipates about 100 kT per bit.

Less could be wasted if the enzyme moved a little more slowly (and of course, the reaction rate does vary with concentration gradient), but there has to be a certain speed for the sake of life!

Still, 100kT per bit is considerably more efficient than the 10^8 kT thrown away by a typical transistor!"

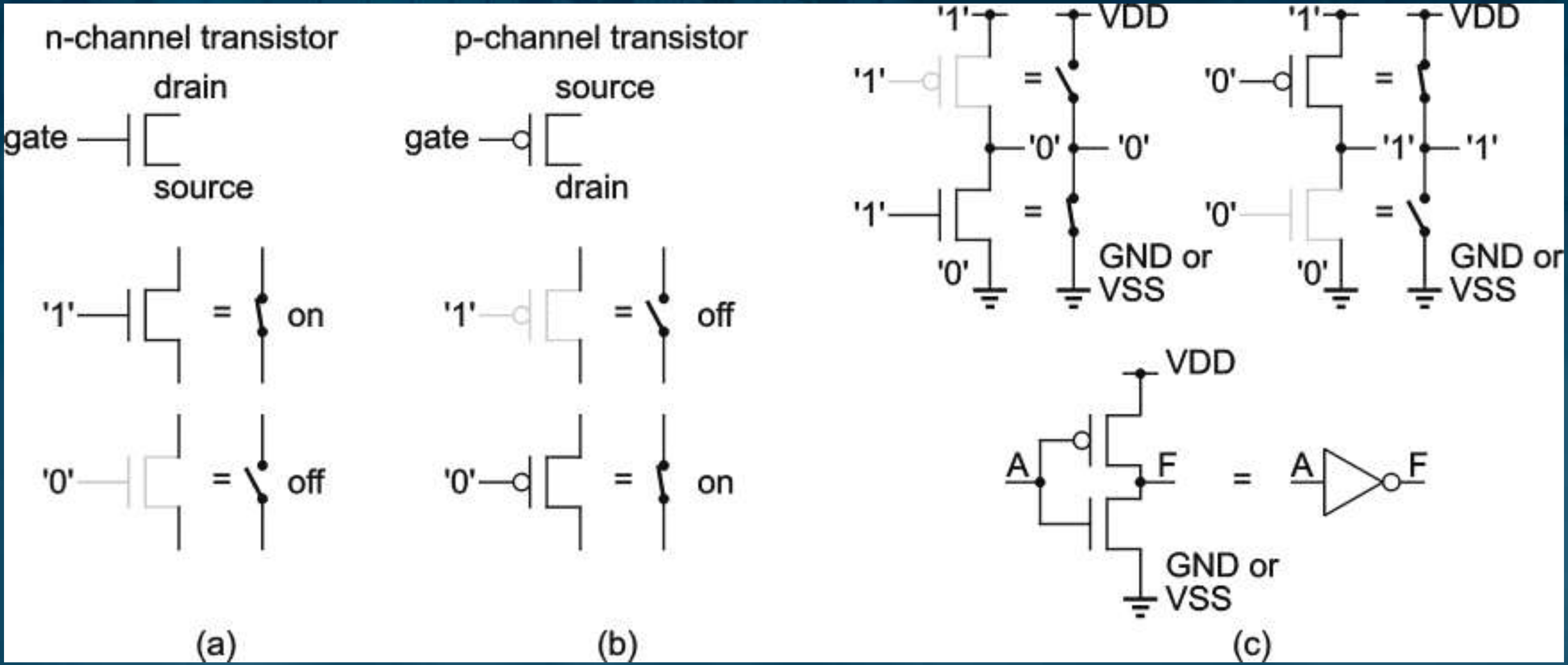
(37,600 kT in 90nm ... 15,300 kT in 65nm …)

18nm SOI transistor



E. Pop (using MONET & Medici) http://poplab.stanford.edu/files/soimovie.mpg

Inverter

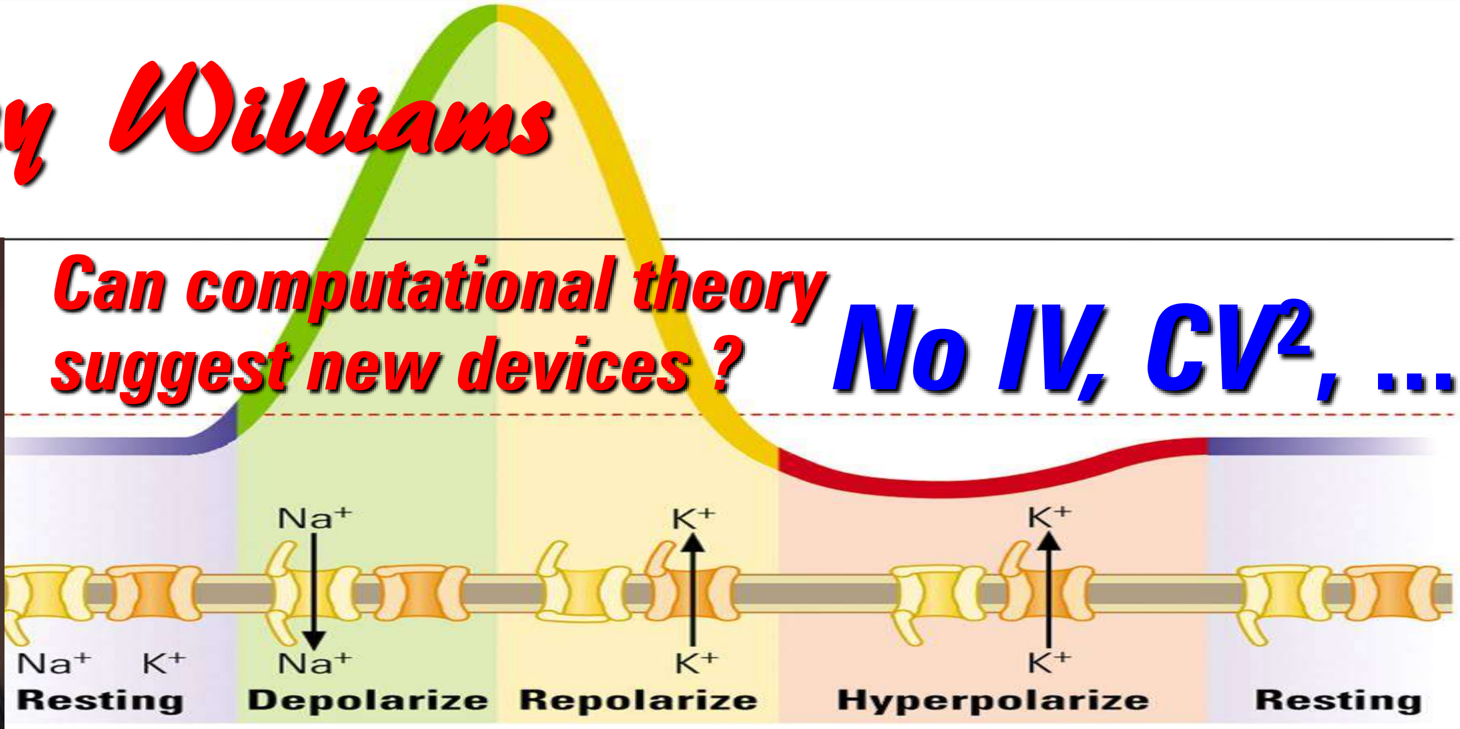


- CMOS inverter moves charges from V_{DD} to C_{load} (pMOS) … and then from C_{load} to GND (nMOS), so **charges are lost**
- Ion channels move charges from outside to inside (Na^{+}) … and then back from inside to outside (K^{+})
  Equivalently … from V_{DD} to C_{load} … and back to V_{DD} (!)

Stanley Williams

Can computational theory suggest new devices ?

No IV, CV^2, ...



Yes! **Here is the brain's ultra low-power solution:**

- Two different information carriers (but same type of charge)
- Two different switches (one for each carrier) + asynchronous
- Carriers move (through these switches) in opposite directions
- A third device consumes (very slowly / quasi-adiabatically)

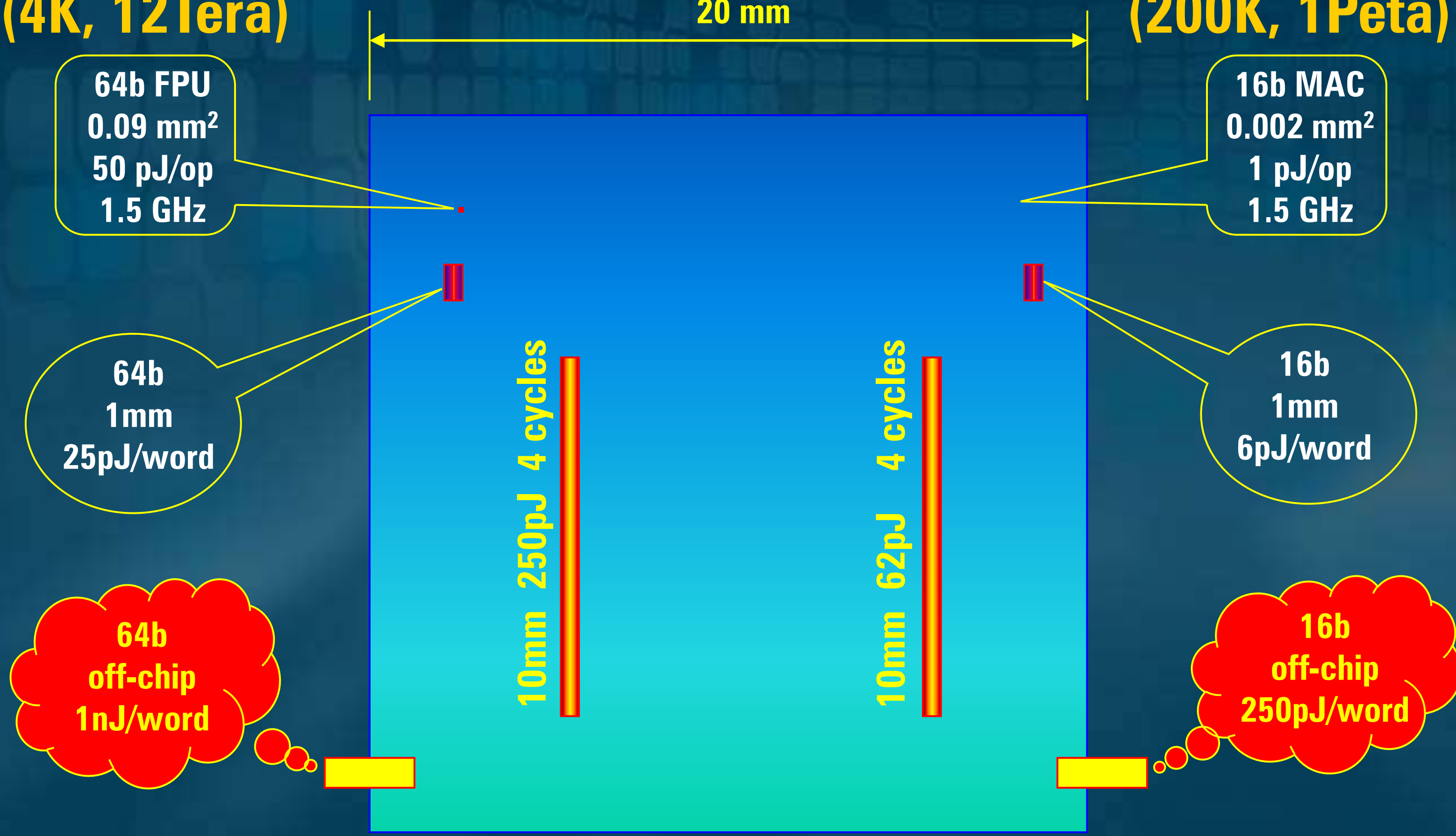
V. Beiu, *MEES'10* http://www.src.org/calendar/e003960/SessionV-Beiu.pdf
Tutorial at *WCCI'10* http://education.ieee-cis.org/lectures/Conference-Tutorials/On-Brain-Inspired-Nano-Interconnects

The fault, dear Brutus, lies … not in our gates, but in our wires!

Wires: A view from the top

64b FPUs (4K, 12Tera)

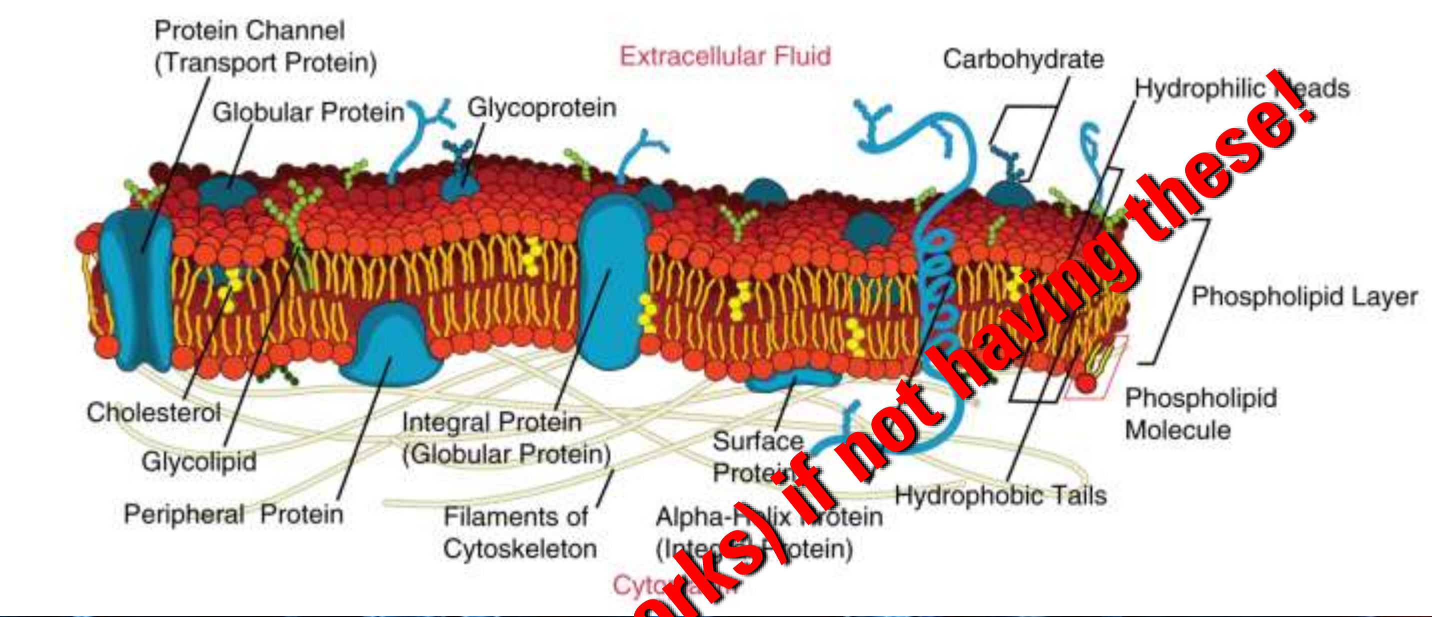
16b MACs (200K, 1Peta)

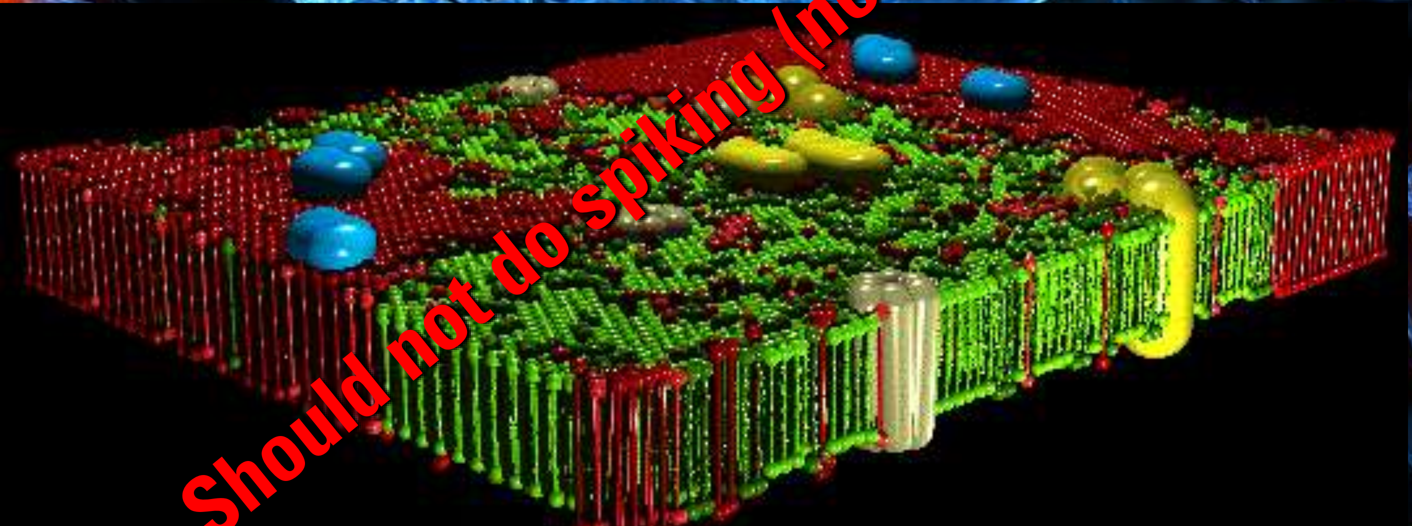


Moving a word across die = 124 MACs or 10 FMAs
Moving a word off chip = 250 MACs or 20 FMAs

W.J. Dally at DAC'09 http://videos.dac.com/46th/wedkey/dally.html

Brain's non-linear "wires"



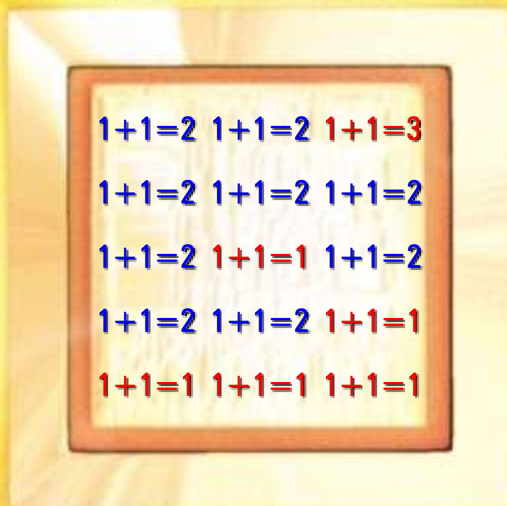


Should not do spiking (neural networks) if not having these!

See http://nobelprize.org/nobel_prizes/chemistry/laureates/2003/chemanim1.mpg

There are two reasons why WSI is very difficult

- **First**, a typical digital chip will fail if even a single transistor or wire is defective
- **Second**, the power dissipated by several hundred chips is over 100W and getting rid of all that heat is a major problem



These two problems have prevented even the largest companies from deploying WSI systems successfully.

Carver Mead

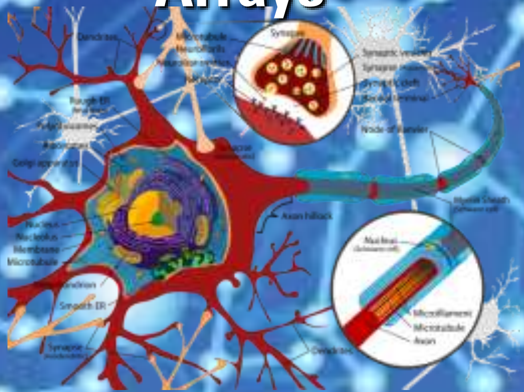
Modeling

- Cable theory 1930
- Hodgkin-Huxley 1952
- Fitzhug-Nagumo 1961 traveling wave
- Morris-Lecar 1981 traveling wave
- Turner and Sens 2004 (mechanical)
- Heimburg 2005 solitons
- Quite a few “simplified” models (latest trend)

None of these can be used for estimating either power or reliability

The brain … one level at a time

- Sub-cellular — Arrays — Boolean
- Cell/neuron — Inner product, Threshold, WTA

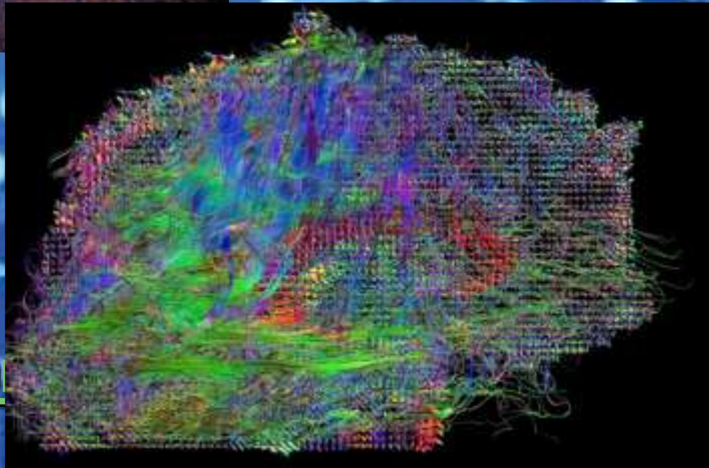


- Columnar — Almost fully connected — Interesting functions …

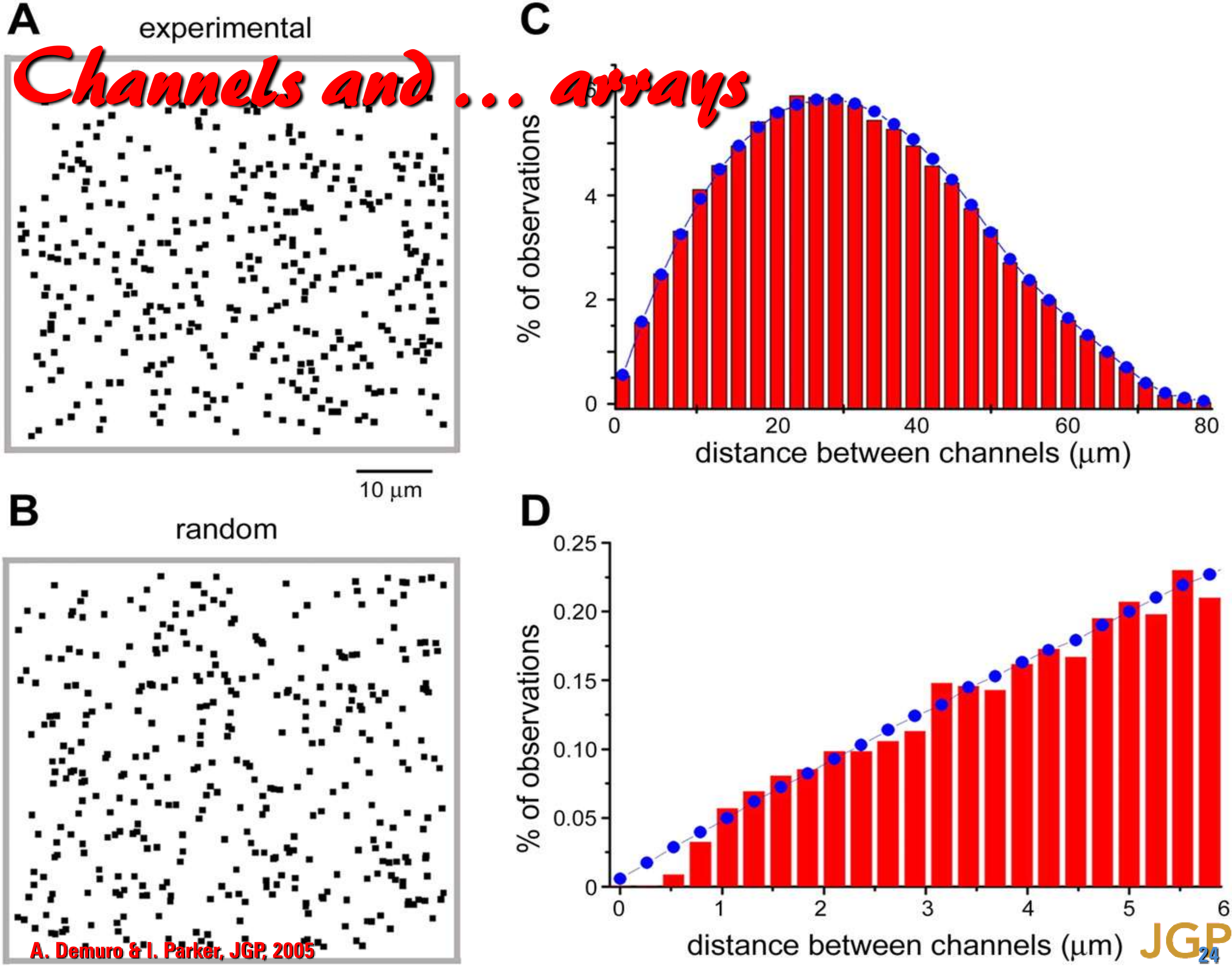
  BBP http://bluebrain.epfl.ch/

- Region/area — ???
- Brain/white — Sparsely connected

  HCP http://www.humanconnectomep

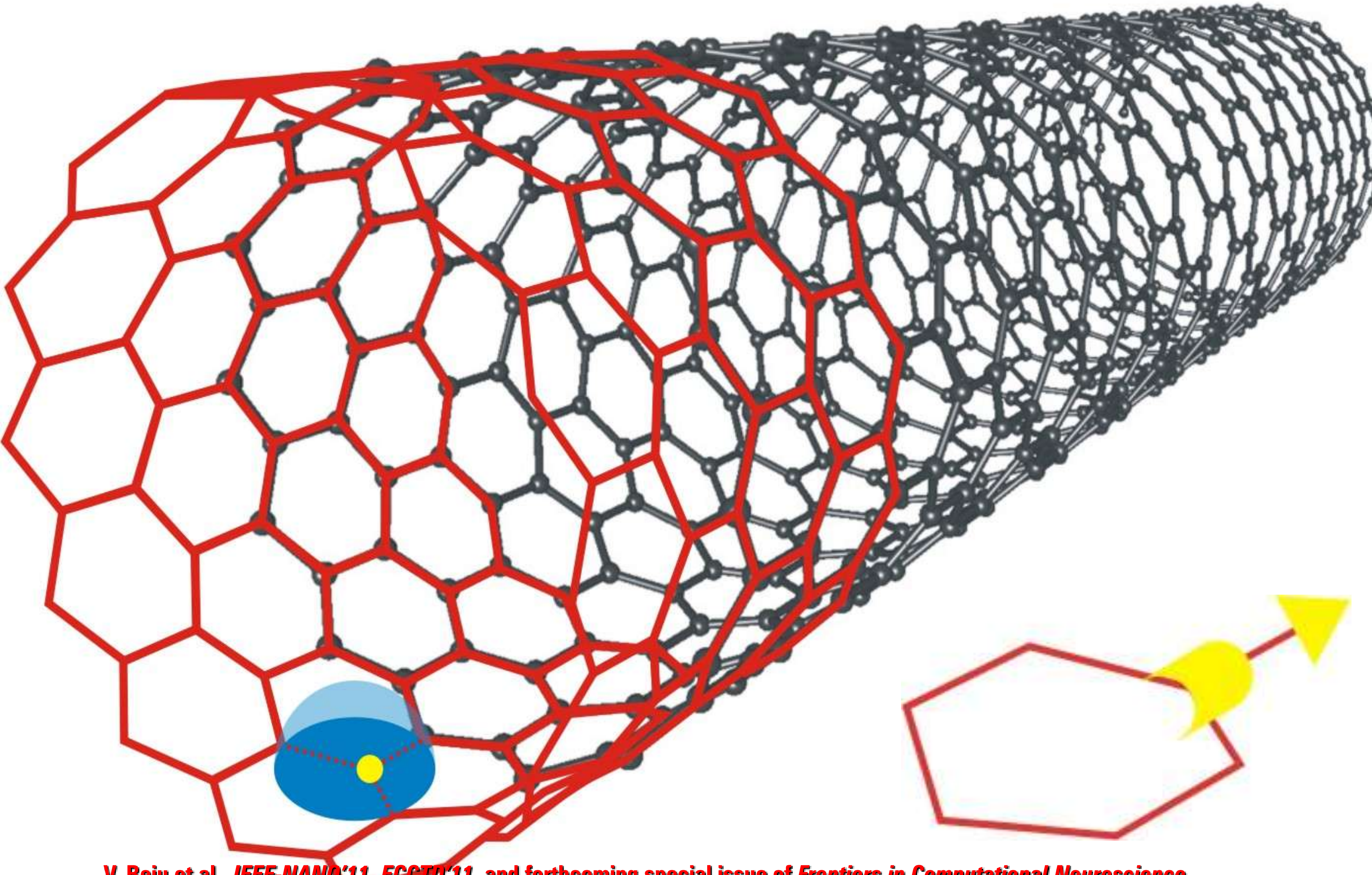


Channels and ... arrays



A. Demuro & I. Parker, JGP, 2005

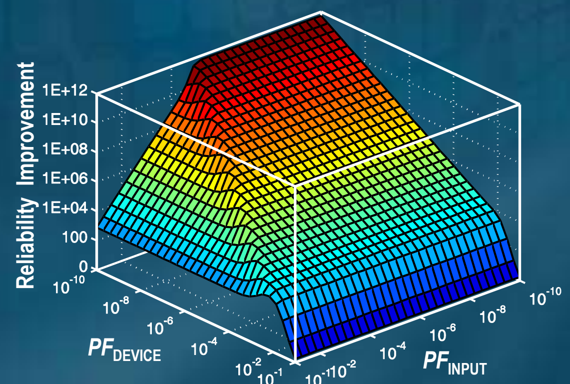
Axon-like (simplistic) communication



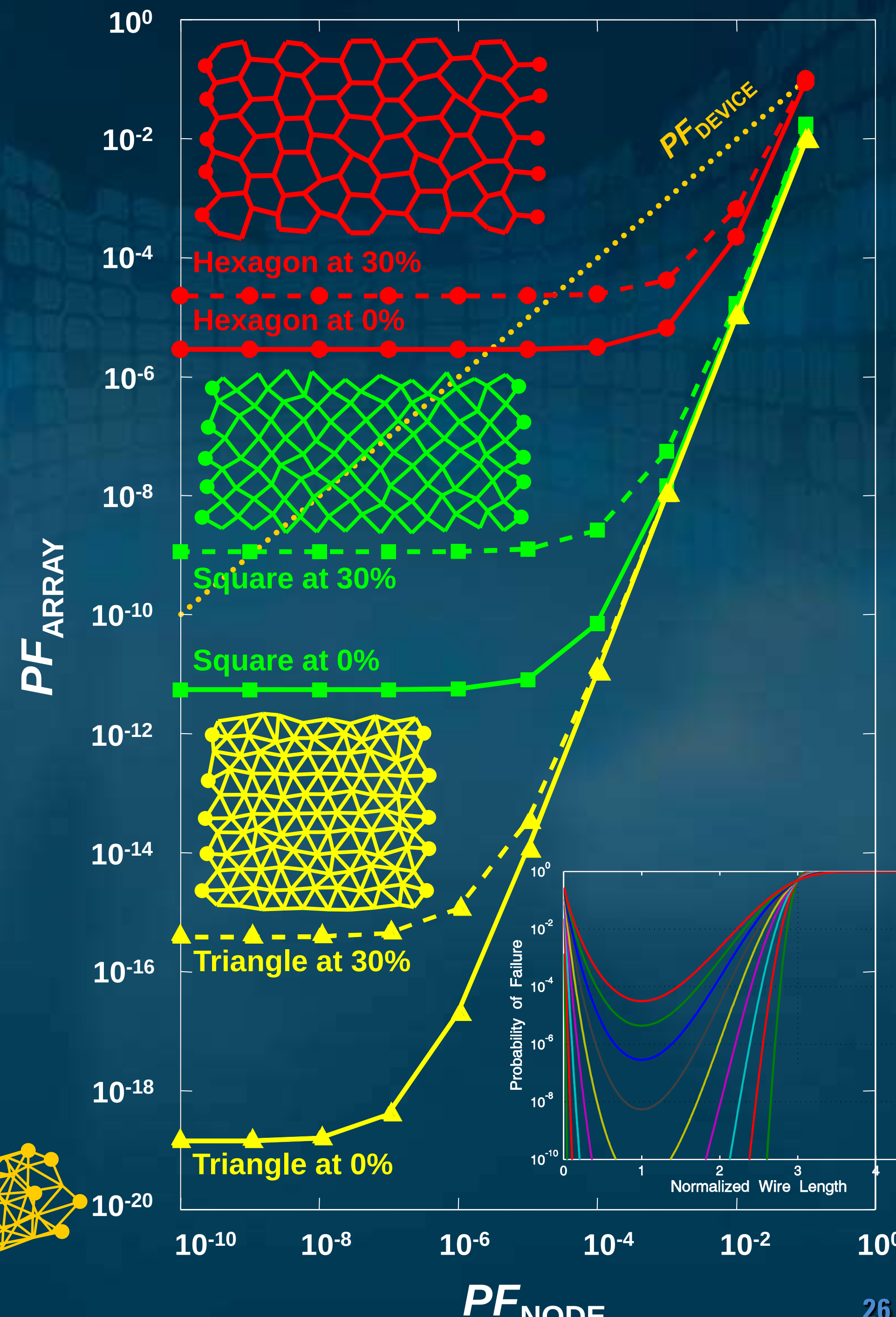
V. Beiu et al., *IEEE-NANO'11*, *ECCTD'11*, and forthcoming special issue of *Frontiers in Computational Neuroscience*

Latest results

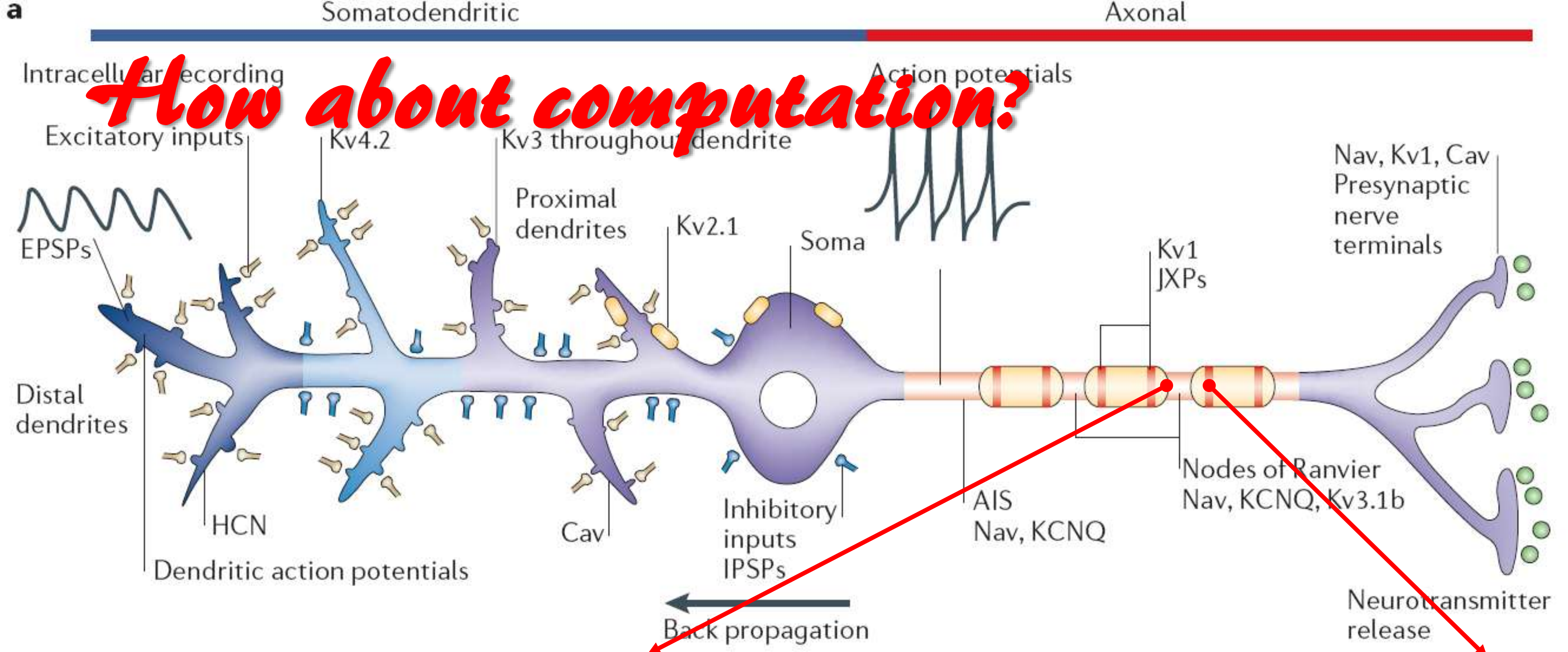
- **2D arrays can be extremely reliable**
  - At least 100 better than von Neumann multiplexing

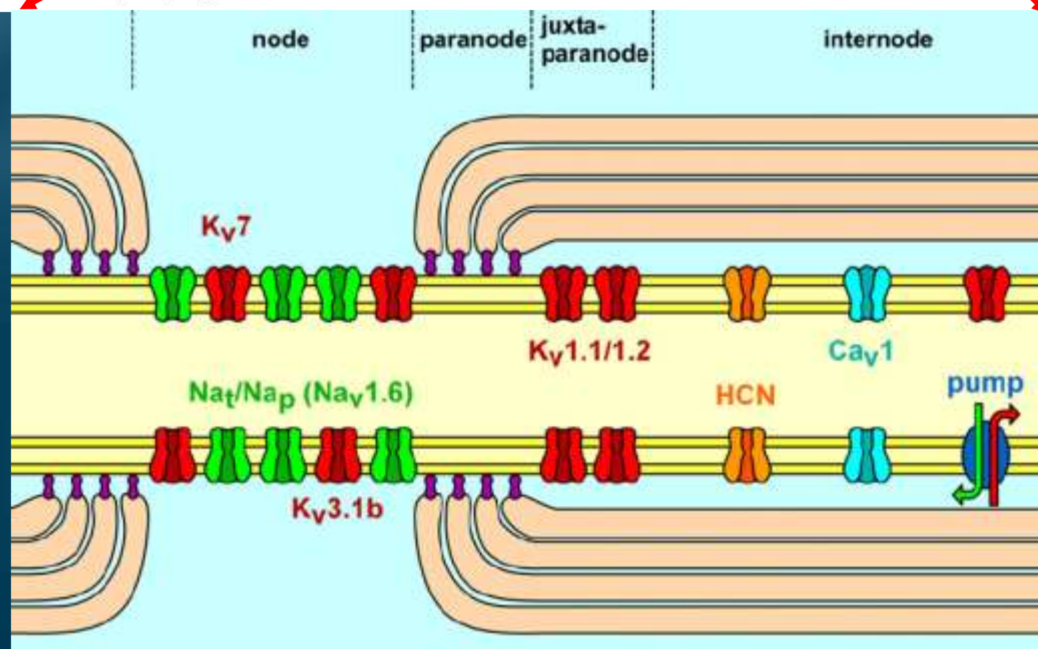


  - 10% failing nodes + 10% failing inputs + 30% placement (x,y) + failing links …
  - Could achieve 0.01nJ/cm
- **3D even better**



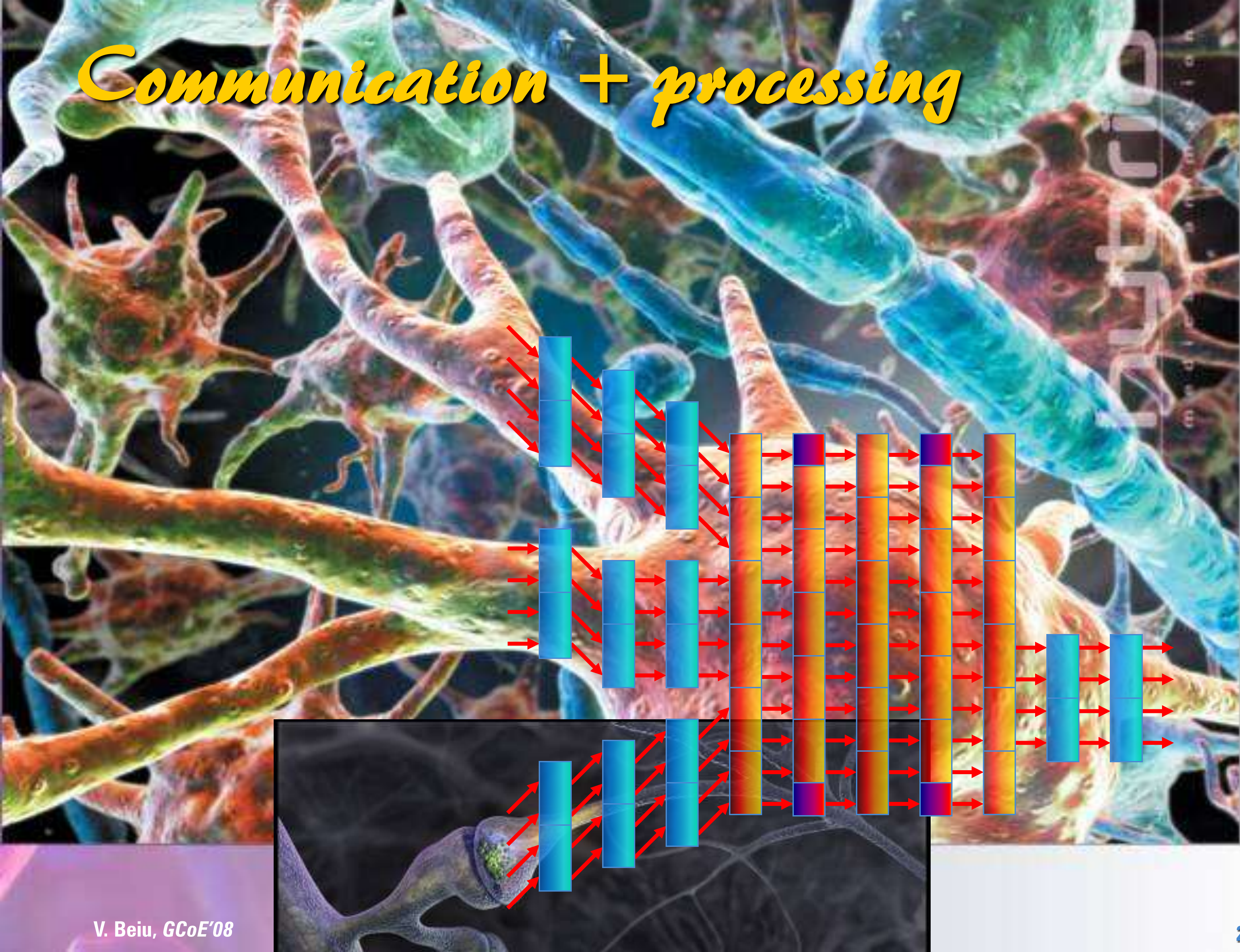
How about computation?



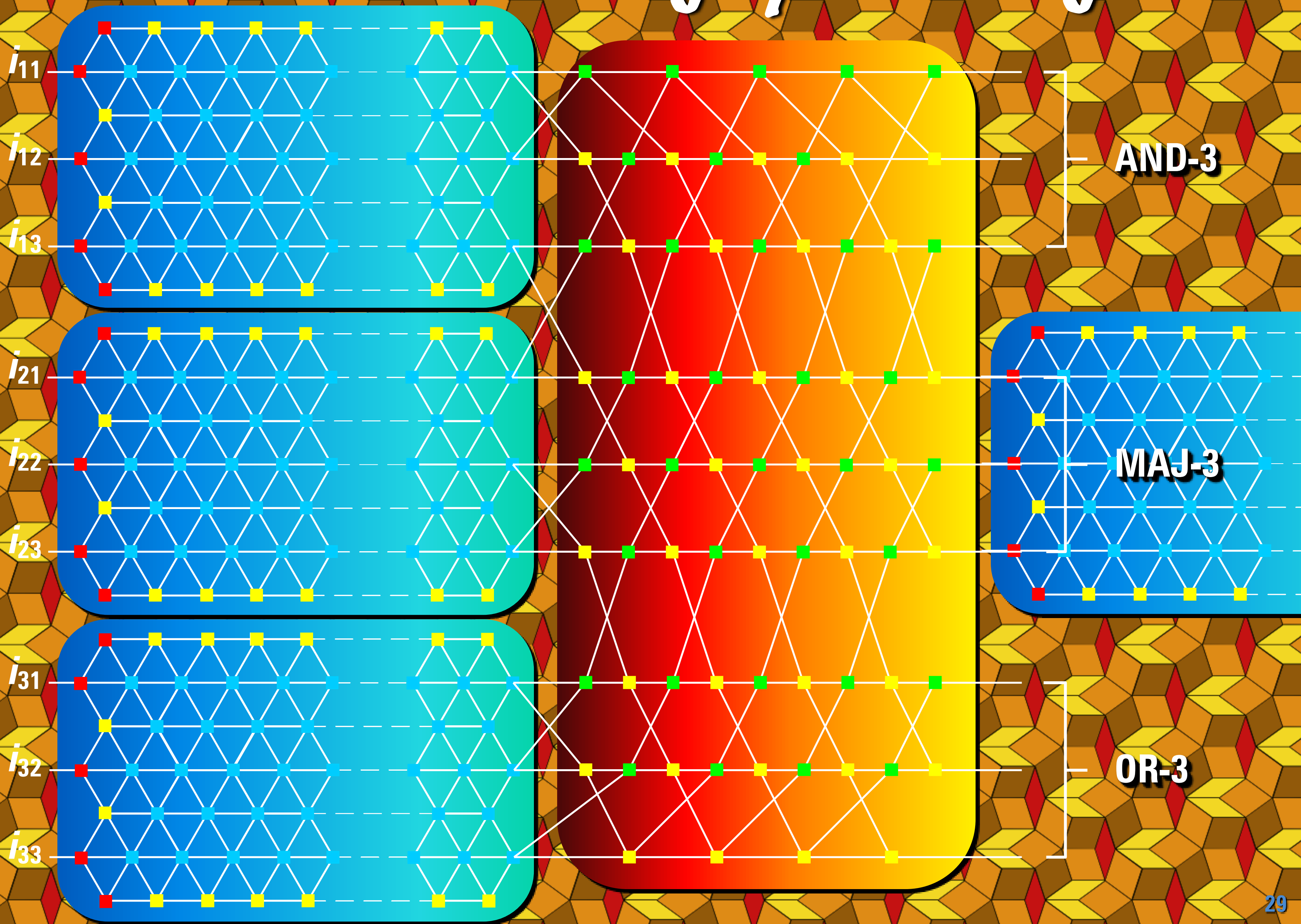


H.C. Lai & L.Y. Jan, *Nature Rev. Neurosci.*, 2006

Communication + processing



Hex (wire) + sorting (processing)



Intermezzo

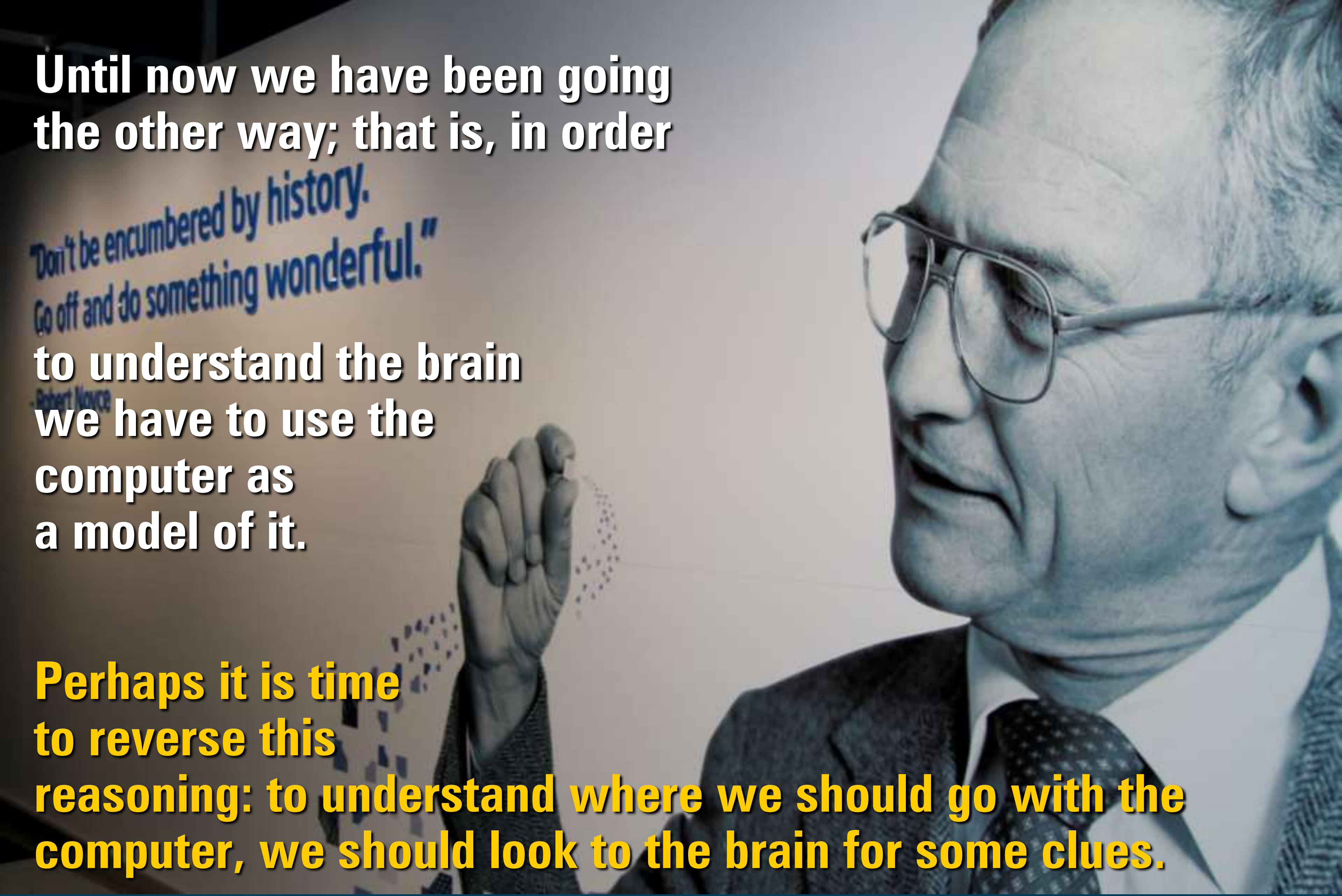
- **Europe**
  - FACETS and BrainScaleS — EU — ~ 30 M$
  - BlueBrain — Switzerland
  - Human Brain Project (HBP) — EU — ~ 1 BEuro

- **USA**
  - SyNAPSE — DARPA — ~ 50 M$
  - Human Connectome — NIH — ~ 40 M$
  - Brain Research through Advancing Innovative Nanotechnologies = BRAIN (functional connectomics) — NIH, NSF, DARPA — > 300 M$

Until now we have been going the other way; that is, in order

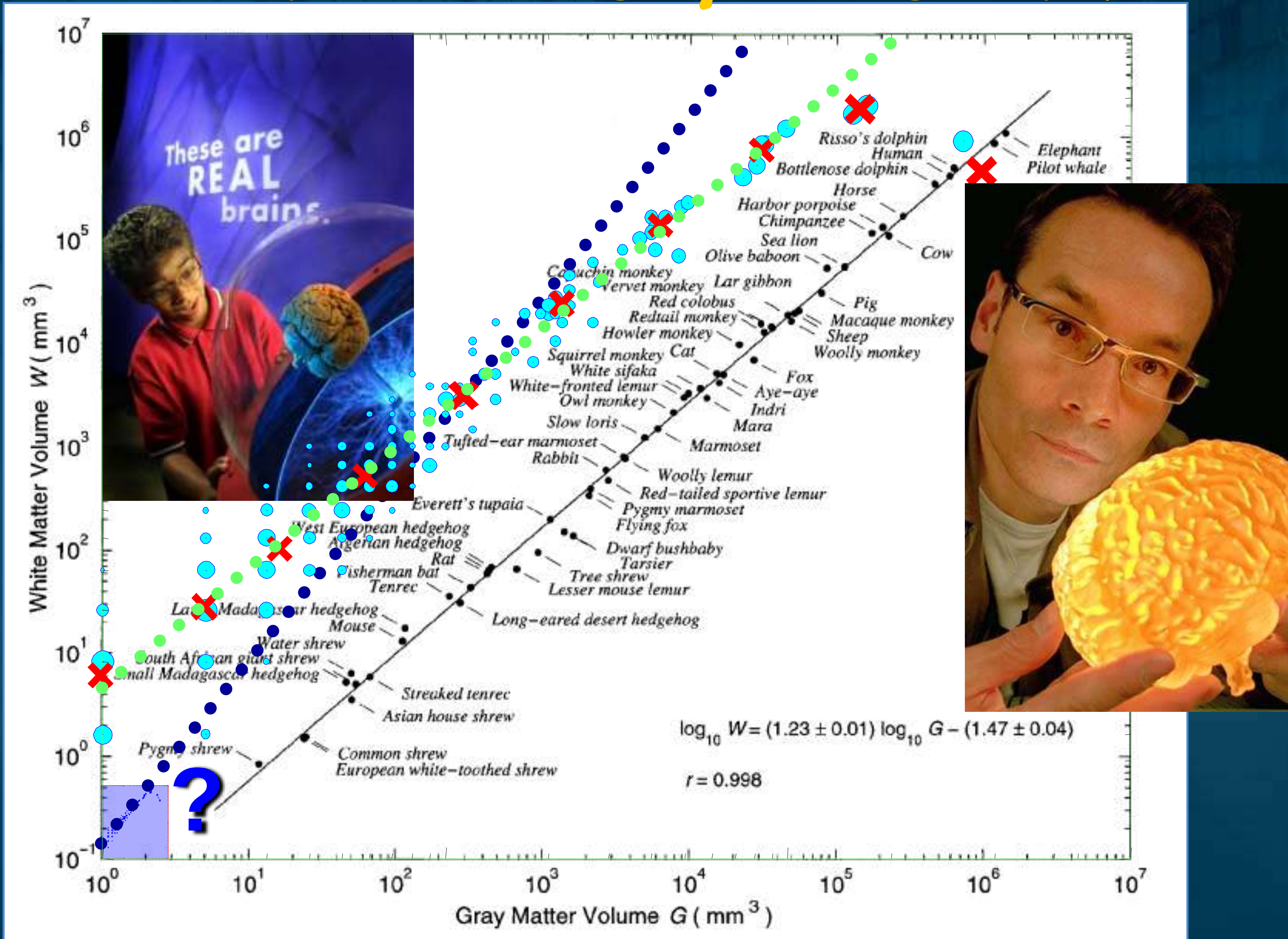
to understand the brain we have to use the computer as a model of it.

Perhaps it is time to reverse this reasoning: to understand where we should go with the computer, we should look to the brain for some clues.



Robert Noyce, IEEE Centenary "The Next 100 Years" ... 1984

A cortical (Brain's) Rent's rule?



K. Zhang & T.J. Sejnowski, *PNAS*, 2000

Hierarchical networks

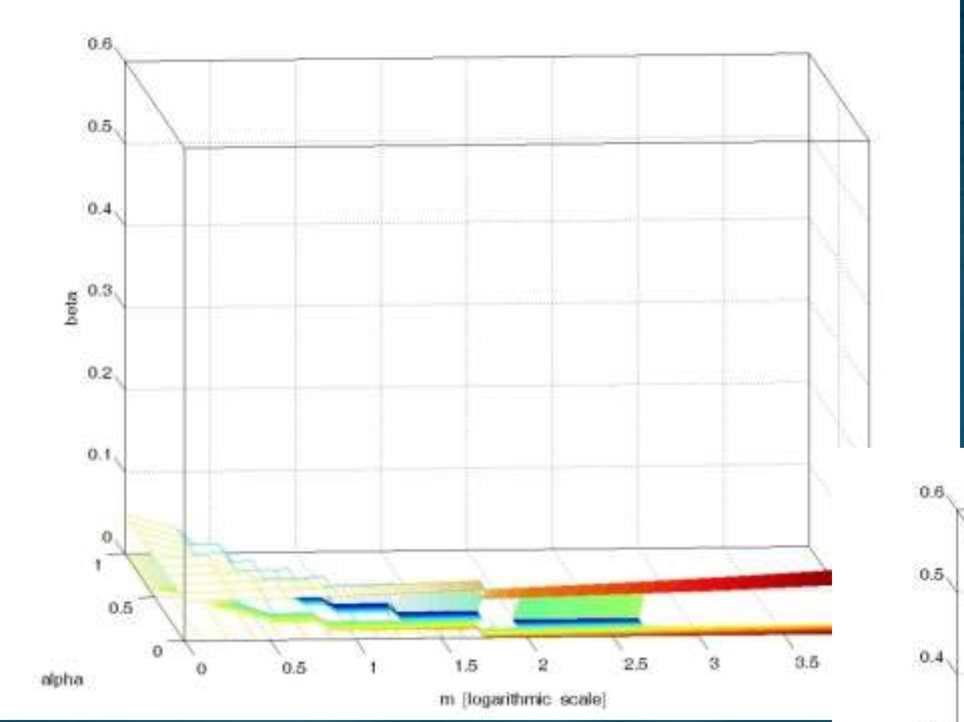
- $N_{\text{CONN}} = N_{\text{CLUSTERS}} \times N_{\text{CONN}}(\text{per cluster}) + N_{\text{CONN}}(\text{amongst clusters}) \times [1 \ldots N_{\text{PROC}}(\text{per cluster})]$

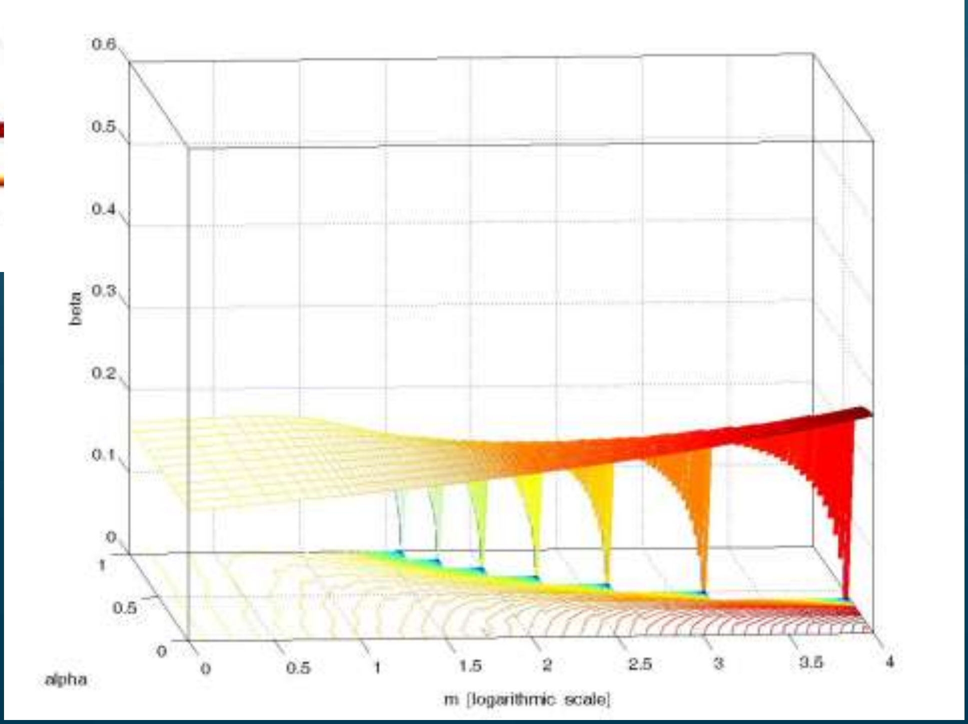
- Local network m^{1+a}
- Global network $(N/m)^{1+b}$
- Hierarchical network $N_{\text{CONN}} = Nm^a + N(N/m)^b$
- Solving for *b* gives $\log(F_{IN} - m^a)/(\log N - \log m)$

- For *fan-in* $= m^a$ the global network becomes simple ... only a few (but wider) global interconnects

V. Beiu *et al.*, *NanoArch'08* and *IC-SAMOS'08*

Fan-in = 4 and 40





V. Beiu *et al.*, *Trans. HiPEAC*, 2009

Columnar structures

- … "How do neurons connect to each other?"

  http://www.youtube.com/watch?v=ySgmZOTkQA8&list=UUIMJeVIVyGp-3_kWtspkS0Q

S.L. Hill *et al.*, *PNAS*, 2012

Brain is (at least) 3-layered

- The **first layer** is represented by the neuron itself … active membrane … solitonic transmission (cellular-like ... but not systolic/CNN/QCA, etc.)

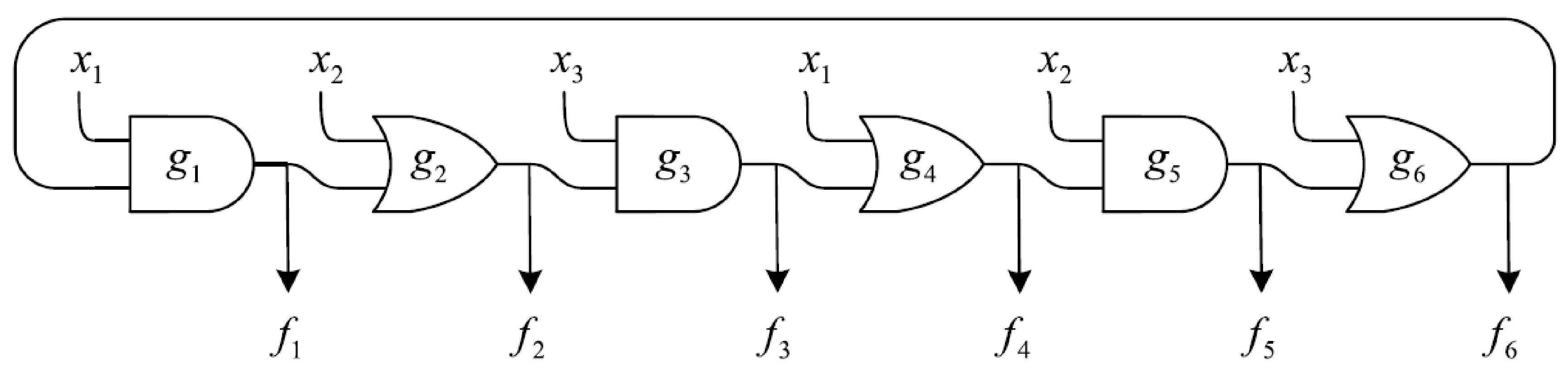
  This level is essential for reliability, while also taking care of for ultra-low power (requires both novel devices and "cellular-like" architectures)

- The **second layer** corresponds to the columnar structures. Highly connected (depending on the neurons' fan-in).

- The **third layer** corresponds to the white matter … sparsely connected using high(er)-speed "wires"

V. Beiu *et al.*, *DCIS'07*, *ISCAS'08*, *NanoArch'08*, *Nano-Net'09*, and *Trans. HiPEAC*, 2009

Columns = Circuits with feedback



$$f_1 = x_1 y$$

$$f_2 = x_2 + f_1 = x_2 + x_1 y$$

$$f_3 = x_3 f_2 = x_3(x_2 + x_1 y)$$

$$f_4 = x_1 + f_3 = x_1 + x_3(x_2 + x_1 y) = x_1 + x_2 x_3$$

$$f_5 = x_2 f_4 = x_2(x_1 + x_2 x_3) = x_2(x_1 + x_3)$$

$$f_6 = x_3 + f_5 = x_3 + x_2(x_1 + x_3) = x_3 + x_1 x_2.$$

M.D. Riedel & J. Bruck, *Discrete Applied Math.*, 2012

Many interesting points

- The view that the brain is a **very complex**, **highly interconnected** system of **unreliable analog processing units** is not an incorrect one, while unfortunately it is … **an incomplete/partial one**

❖ Behaves digitally	... at the lowest level
❖ The elementary switches behave (very much) like	... single ion devices
❖ Behaves (very much) as self-timed	... from the lowest level
❖ Can be seen as analog	... at the system level
❖ Has 10^{19} (large redundancy)	... digital switches
❖ Neurons are quite reliable	... significant redundancy

See http://www.youtube.com/watch?v=YP_P6bYvEjE

... and many misconceptions

- Brain is analog ... SET (but many)
- Brain is slow ... $10^{7\text{-}8}$ ions/sec
- Low power due to speed ... due to ion pumps
- Low power due to low V ... P is not IxV (!)
- Low power due to 3D ... Reliability due to 3D
- Low power due to spikes ... On the contrary
- Neurons are not reliable ... Yes, they are! [Synapses are not ...]
- Using learning/adaptation leads to reliable structures ... Only if "devices" are reliable "enough"

Electronics versus Brain

Electron (one carrier)	Ions (several carriers)	
Current (needs V)	Diffusion (creates V)	
Transistor	Ion channel	10^{0} SET
Wire (current)	Axon	$10^{3\text{-}4}$ SETs
IP core	Neuron	$10^{6\text{-}7}$ SETs
Processor	…	
Multi-core	Column	$10^{10\text{-}11}$ SETs
Super-computer	Area	
…	Brain	$10^{18\text{-}19}$ SETs

V. Beiu (forthcoming special issue of *Frontiers in Computational Neuroscience* -- on the metabolic cost of neural mechanisms)

Plans for the future

- HH-equivalent (model/equations) for power/energy
  - Communication between ion channels (statistically large fan-ins)
  - Logical functioning of an ion channel (probabilistic TLGs)
- HH-equivalent (model equations) for reliability
  - Statistical analysis of arrays of ion channels
  - CAD + Bayesian tools for arrays of ion channels
- Columnar structure
  - Theory of circuits with feedback (and large fan-ins)
  - CAD tools for simulations/synthesis

V. Beiu (ERC Synergy proposal in planning)

Should lead to novel and detailed ...

- Understanding of the functioning of the Brain (at the lower levels)
- Models
- CAD tools
- Architectures


THANK YOU

Oups ... forgot to do a quick review

UAEU