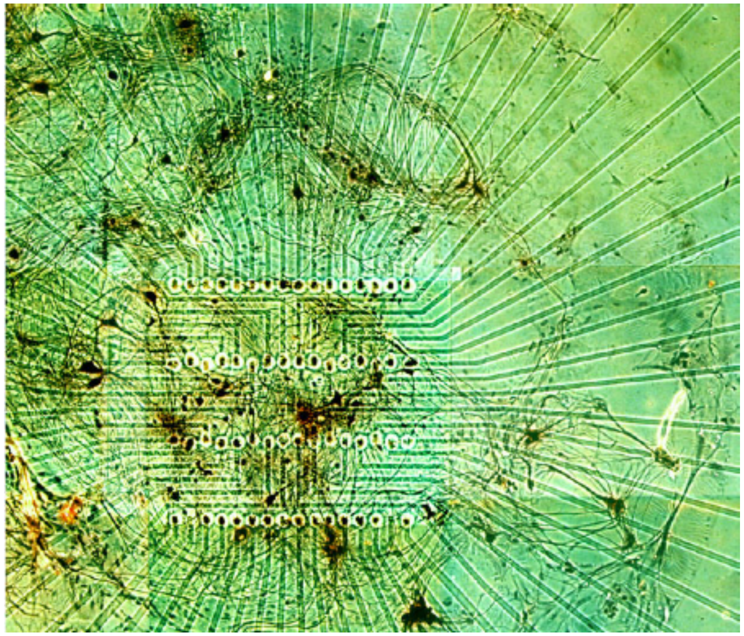
Information processing structures in neural populations in vitro



Luís M. A. Bettencourt

Los Alamos National Laboratory & Santa Fe Institute

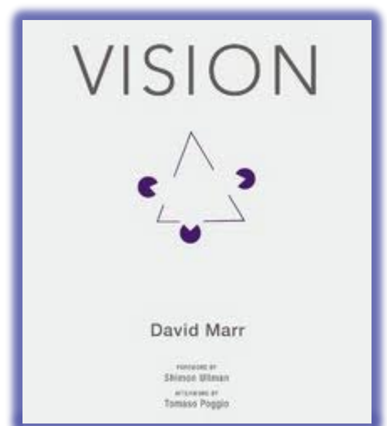
KITP-UCSB, July 19, 2011

http://math.lanl.gov/~lmbett http://www.synthetic-cognition.edu

Computational theory vs. biological detail

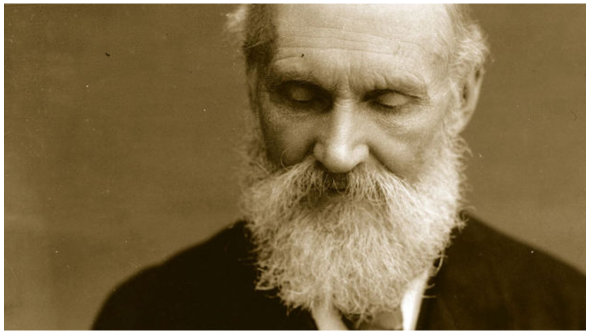
an algorithm is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied.

In a similar vein, trying to understand perception by studying only neurons is like trying to understand flight by studying only feathers: It just cannot be done...



David Marr, Vision 1982

A brief history of information



Energy conservation and efficiency in thermodynamics

von Helmholtz 1847

Entropy: $S=Q/T$ measured unavailable heat

Clausius 1865

combustion

→ dissipated heat: **lost**

Q: can we make it more efficient?

→ **work**: useful product

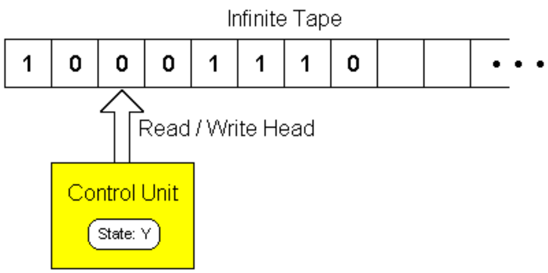
Statistical interpretation as microstates

$$S = k_B \ln W$$

Maxwell 1871, Gibbs 1878,
Boltzmann 1877

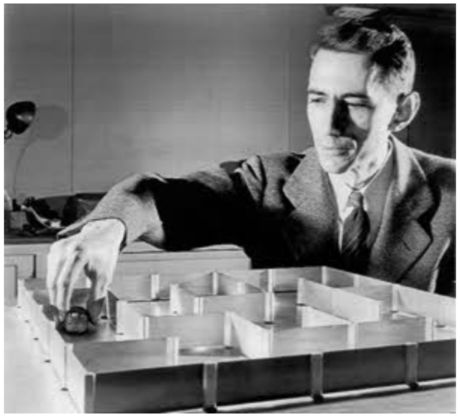
Computation is Universal

Turing 1936



Information theory & communication

Shannon 1948



$$H = -\sum_i p_i \ln_2 p_i$$

Algorithmic complexity

Solomonoff, Kolmogorov, Chaitin, 1960-69

Minimum prescription length

Information and the structure and dynamics of complex systems

genomes, brains, social organizations

now

Knowledge is 'power'

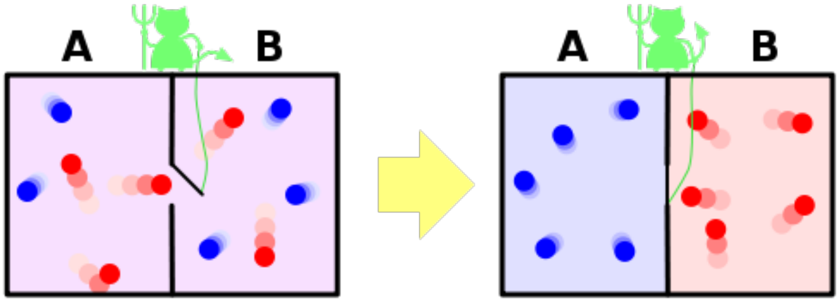
but it's not free

Entropy of an ideal gas

1. Assume that N particles are anywhere in the volume V

$$S = k_B \ln V^N = N k_B \ln V$$

2. But what if you could **know** more?



wikipedia

Maxwell's demon:

more knowledge leads to more work extracted → greater efficiency

What if you could know more about a market? A city? A firm?

Information processing in complex systems

Complex systems are ensembles that create and maintain high levels of organization locally (in space and time):

$$\Delta S_{int} < 0 \qquad \text{while} \qquad \Delta S = \Delta S_{int} + \Delta S_{ext} \geq 0$$

Complexity is the ability to store (structure) and process information (dynamics) in ways that are not independent (high T)

$$S(X,Y) \neq S(X) + S(Y)$$

… at the same time also not fully redundant (T=0)

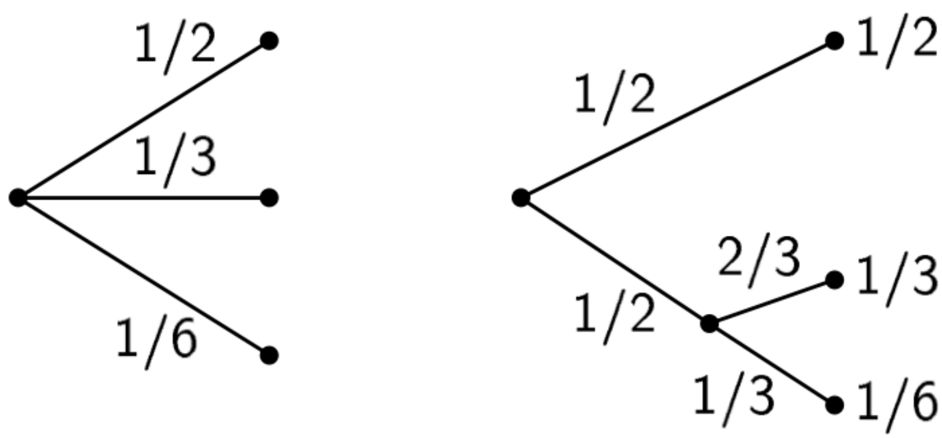
$$S(X,Y) \neq S(X) \quad \text{(for example)}$$

Parts of complex systems are organized in interdependent ways

Properties of Information
and the logarithm

Entropy H

1. Is continuous in p_i
2. If all p_i=1/N it should be a monotonically increasing function of N
3. If a choice is broken down into component choices then H should be the weighted sum of the individual H



H(1/2.1/3,1/6) =

H(1/2),1/2) +1/2 H(2/3,1/3)

Then, it follows: $$H = -K\sum_i p_i \log p_i$$

What is information?

[information theory]

Information is *increased* predictability

[or uncertainty reduction]

$X \sim P[X]$ Uncertainty in X= (Shannon) Entropy

$$H(X) = -\sum_x p(x) \ln_2 p(x)$$

If we know Y: then the (mutual) information it has on X is

$$I(X;Y) = H(X) - H(X \mid Y) \sim \frac{\Delta H(X)}{\Delta Y}$$

information and non-extensibility

Entropy is designed as an extensive property of a system

e.g. non interacting system

$$H(N) = \log \Omega(N) = \log\left[V^N C(T,P)\right] \sim A + BN$$

Information is in general sub-extensive; its properties reveal spatial, temporal and categorical relations

$$I(N) = \Delta H(N) = H(N) - H(N-1)$$

$$\frac{I(N)}{N} \xrightarrow[N \to \infty]{} 0$$

Crutchfield & Feldman 2001

Bialek, Nemenman & Tishby 2001

Information

and the structure of neural systems

Information is more than correlation

between stochastic variables

The information of 3 or more variables can be

larger	**equal**	**smaller**
synergy	**independence**	**redundancy**

to the sum of its parts.

This creates conditions for coordination and cooperation between

a) Elements in a network to store and process information
b) 'Searchers' in cooperative multi-agent optimization problems
c) Individuals seeking information may benefit from pooling it
d) message tokens in a language

A [discrete] calculus in information

measurement and information gain

Define a discrete differential calculus of entropy relative (conditional) to knowledge of other variables {Y}:

$$\frac{\Delta H(X)}{\Delta Y} \equiv H(X \mid Y) - H(X) = -I(X;Y)$$

$$\frac{\Delta^2 H(X)}{\Delta Y_1 \Delta Y_2} = \frac{\Delta}{\Delta Y_2}\left[\frac{\Delta H(X)}{\Delta Y_1}\right] = H(X \mid Y_1, Y_2) - H(X \mid Y_1) - H(X \mid Y_2) + H(X)$$

$$= I(X;Y_1) - I(X;Y_1 \mid Y_2).$$

… and where higher order variations follow via application of the chain rule.

Bettencourt et al. 2008
Bettencourt 2009

A cluster decomposition in terms of functional modules

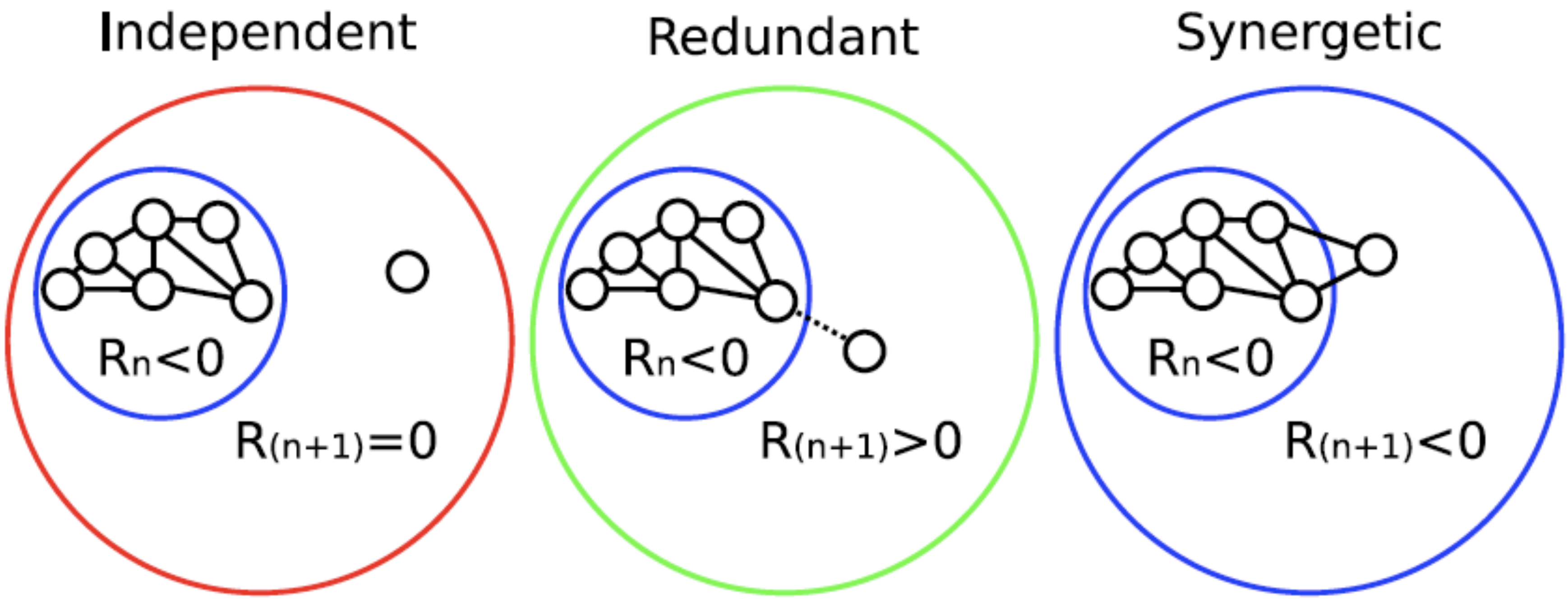
For a given set $\{Y_1, \ldots, Y_n\}$,

$$I(X, \{Y_1, Y_2, ..., Y_n\}) = S(X|Y_1, Y_2, ..., Y_n) - S(X)$$

$$= \sum_i \frac{\Delta S(X)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(X)}{\Delta Y_i \Delta Y_j} + \ldots + \frac{\Delta^n S(X)}{\Delta Y_1 \ldots \Delta Y_n}$$

- This decomposition is analogous to a Taylor series.
- Each term isolates the pairs, triplets,etc.
- Define $R_n(X, Y_{i_1}, \ldots, Y_{i_n}) \equiv \frac{\Delta^n S(X)}{\Delta Y_{i_1} \cdots \Delta Y_{i_n}} \neq R_n^S(X, Y_{i_1}, \ldots, Y_{i_n})$

R_n gives redundancy or synergy exactly to n^{th} order



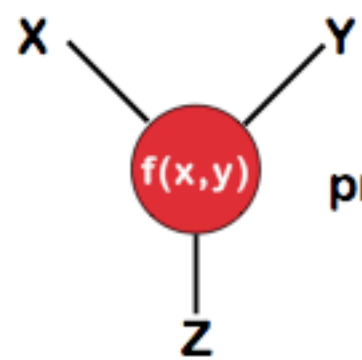
Example:

- Start with set of n nodes with $R_n < 0$
- Add one more node
- R_{n+1} gives the relationship of that node to the previous set

Independence, redundancy and synergy

The structure of larger connectivity arrangements:

"chains", "cascades"
relay information
transient memory
R>0



"circuits"
process information
R<0

$$R^{(2)}(X;Y;Z) = I(X;Y) - I(X;Y \mid Z)$$

Fully symmetric

$$= I(X;Z) + I(Y;Z) - I(\{X,Y\};Z)$$

Schneidman, Bialek et al., '03
Bettencourt et al '07

examples of circuits

Logical 'AND'

If X,Y random per unit time: H(X)=H(Y)=1

$H(Z)=2-3/4 \log_2(3)$
$I(X;Y)=0,$ $I(X;Z)=3/2-3/4 \log_2(3) = I(Y;Z)$
$I(X;Y;Z)=2-3/4 \log_2(3)$
$R= 1- 3/4 \log_2(3)$ **<0**

X	Y	Z
0	0	0
0	1	0
1	0	0
1	1	1

Logical 'OR'

$H(Z)= 2-3/4 \log_2(3)$
$I(X;Y)=0,$ $I(X;Z)=3/2-3/4 \log_2(3) = I(Y;Z)$
$I(X;Y;Z)=2-3/4 \log_2(3)$
$R= 1- 3/4 \log_2(3)$ **<0**

X	Y	Z
0	0	0
0	1	1
1	0	1
1	1	1

Synergy results from any constraint $X=f(\{Y\})$, regardless of the explicit form of f

(a)

Inputs				Synergy			Independence		Redundancy	
x	y	AND	OR	XOR	$Y \dashv X$	$X \dashv Y$	X	Y	CHAINS	
0	0	0	0	0	0	0	0	0	0	1
0	1	0	1	1	0	1	0	1		
1	0	0	1	1	1	0	1	0		
1	1	1	1	0	0	0	1	1	1	0

(b)

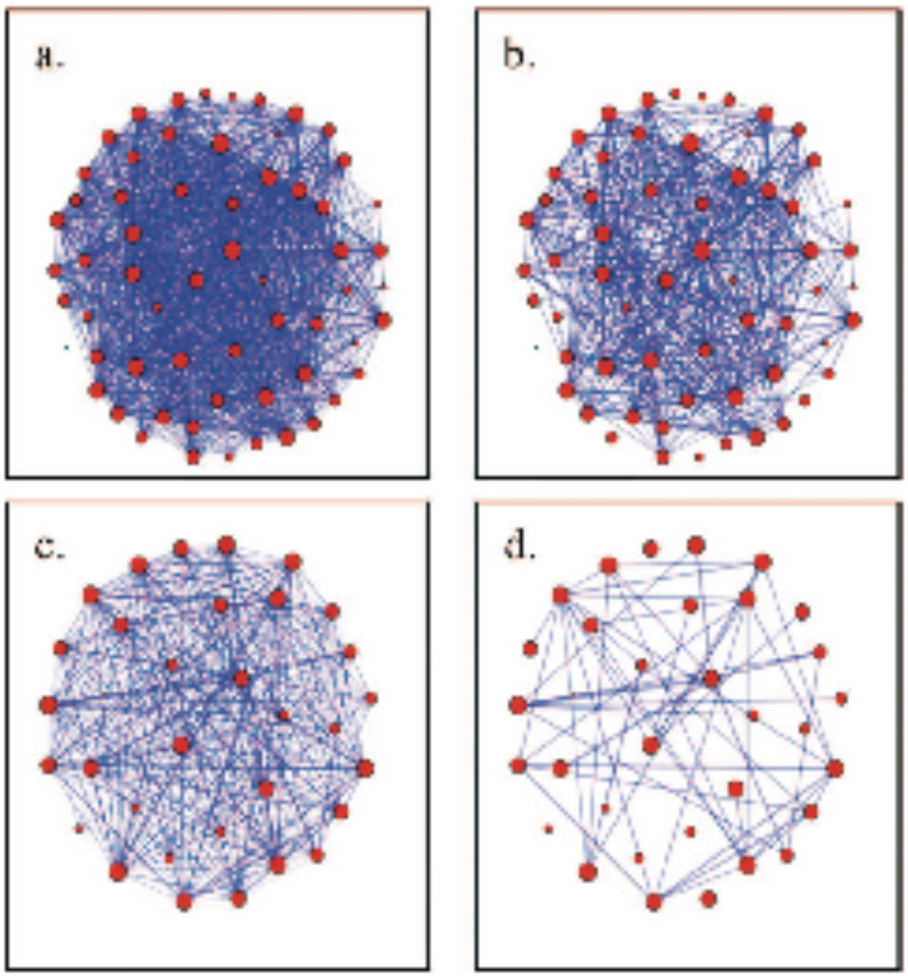
Function	$H(Z)$	$I(X;Y)$	$I(X;Z)$	$I(Y;Z)$	$I(X,Y,Z)$	R
AND	$2-\frac{3}{4}\ln_2(3)$	0	$\frac{3}{2}-\frac{3}{4}\ln_2(3)$	$I(X;Z)$	$2-\frac{3}{4}\ln_2(3)$	$1-\frac{3}{4}\ln_2(3)$
OR	$2-\frac{3}{4}\ln_2(3)$	0	$\frac{3}{2}-\frac{3}{4}\ln_2(3)$	$I(X;Z)$	$2-\frac{3}{4}\ln_2(3)$	$1-\frac{3}{4}\ln_2(3)$
XOR	1	0	0	0	1	-1
X	1	0	1	0	1	0
Y	1	0	0	1	1	0
CHAIN	1	1	1	1	2	1

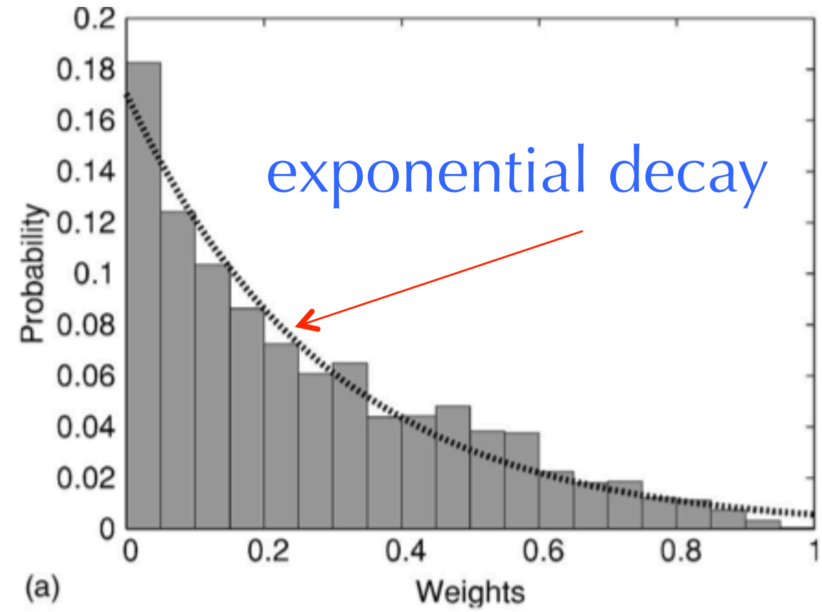
Bettencourt, Stephens, Ham and Gross
PRE 2007

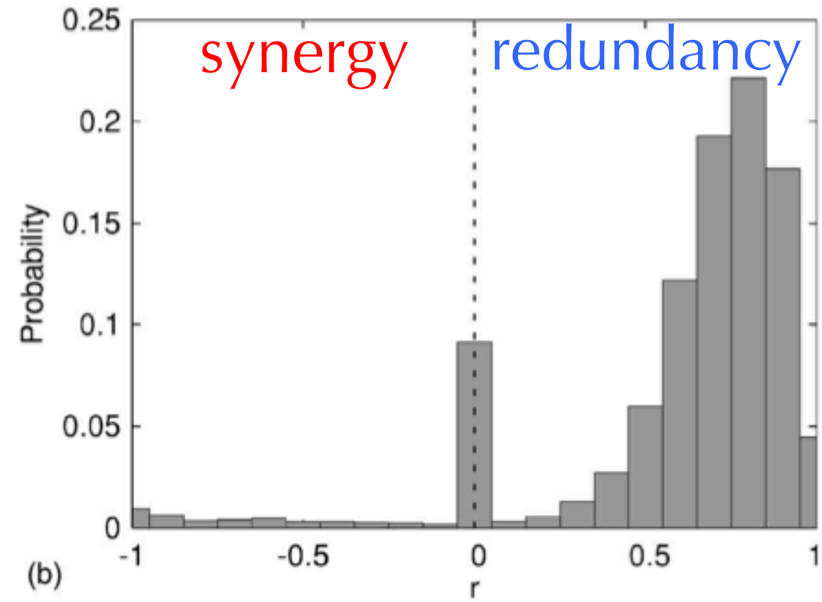
TABLE II. The joint consideration of the values of R and of the binary mutual information I leads to a classification of the connectivity arrangements between three stochastic channels. For $R>0$ only the two strongest links are adopted. For $R<0$ we adopt three links as a general representation of the specific functional interdependence between the three cells.

		No three-cell connected structure		
	No connections:	$\forall_{i,j\in\{1,2,3\}} I(X_i,X_j)=0$		
$R=0$	One connection:	$\exists!_{i,j\in\{1,2,3\}} I(X_i,X_j)\neq 0$		
$R\neq 0$		There is a three-cell connected structure		
	Two connections:		$R>0$ (redundancy)	
		$X\leftrightarrow Y\leftrightarrow Z$	$X\leftrightarrow Z\leftrightarrow Y$	$X\leftarrow Z\rightarrow Y$
		Markov:	Markov:	
		$I(X;Y)\geq I(X;Z)$	$I(X;Z)\geq I(X;Y)$	
		$I(X;Z\mid Y)=0$	$I(X;Y\mid Z)=0$	$I(X;Y\mid Z)=0$
		$R=I(X;Z)$	$R=I(X;Y)$	$R=I(X;Y)$
	Three connections:		$R<0$ (synergy)	
			$X\rightarrow Z\leftarrow Y$: $Z=f(X,Y)$	
			Each variable is a function of all others.	

Bettencourt, Stephens, Ham and Gross
PRE 2007



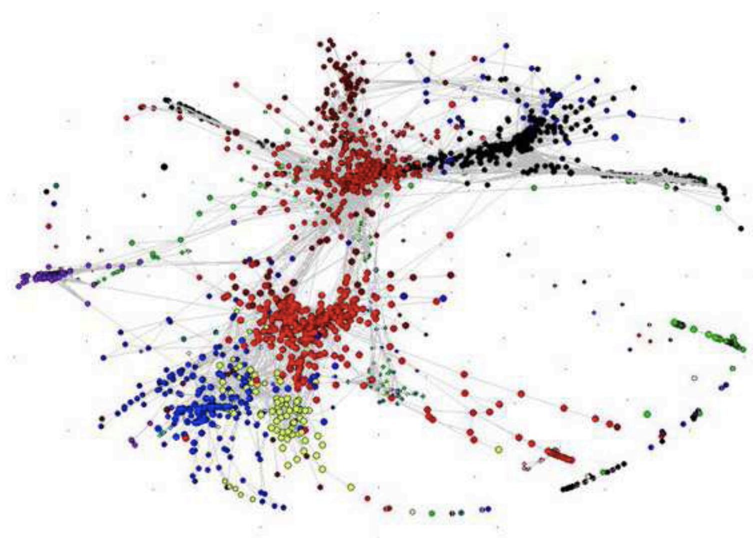




Bettencourt, Stephens, Ham and Gross
PRE 2007

Information processing in the nervous system

functional information modules in complex networks

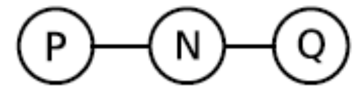


Mouse liver gene expression network from Jake Lusis Laboratory,UCLA

Functional subgraphs

Types of functional units (building blocks)

- Redundant chains

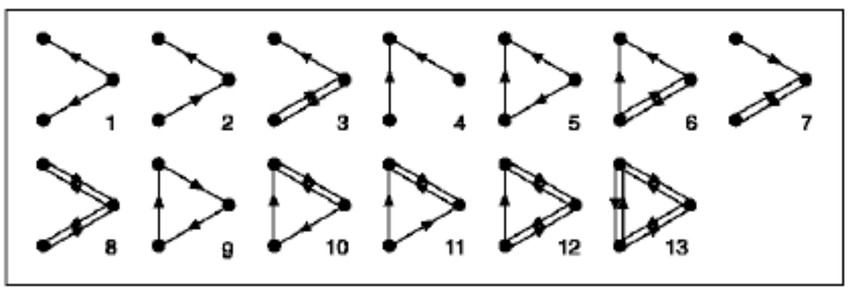


(Bettencourt et al., Schneidman et al.)

- Synergetic circuits

(Schneidman et al., Bettencourt et al., Gross et al.)

- Motifs

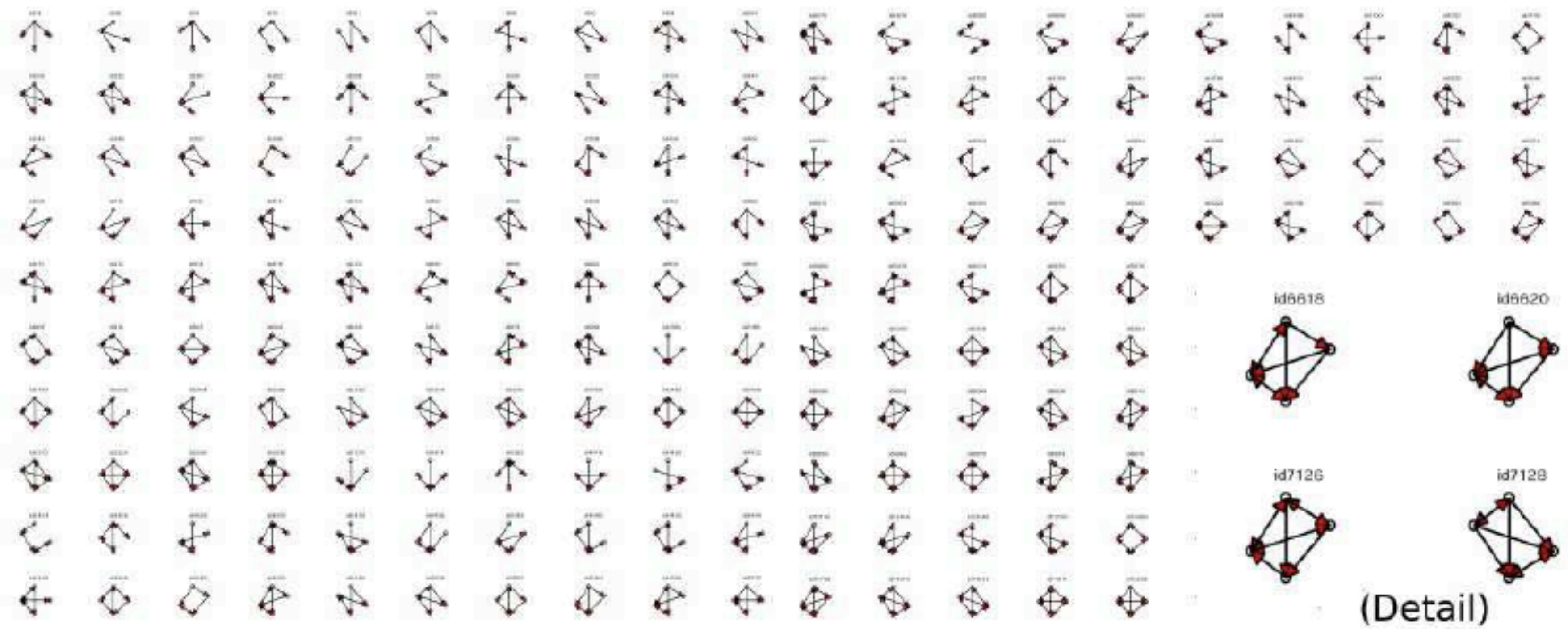


(Middendorf et al., Milo et al.)

(Image from Milo et al.)

Curse of dimensionality
motifs

There are too many!
(199 4-motifs, 9,364 5-motifs, 1,530,843 6-motifs, etc.)



(Image from Milo et al.)

Architecture and information processing in the nervous system

Frontal cortex neurons
from fetal mice
[thousands/mm^2]

Grown in vitro over a
1mm^2 microelectrode array

Disassociated Culture
spontaneously form network

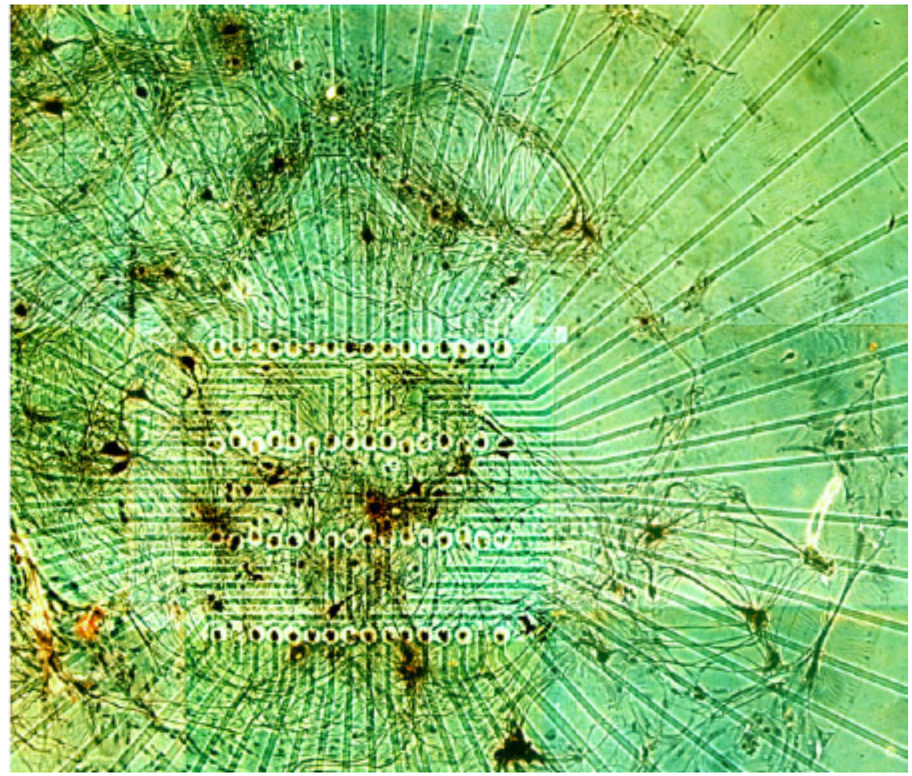
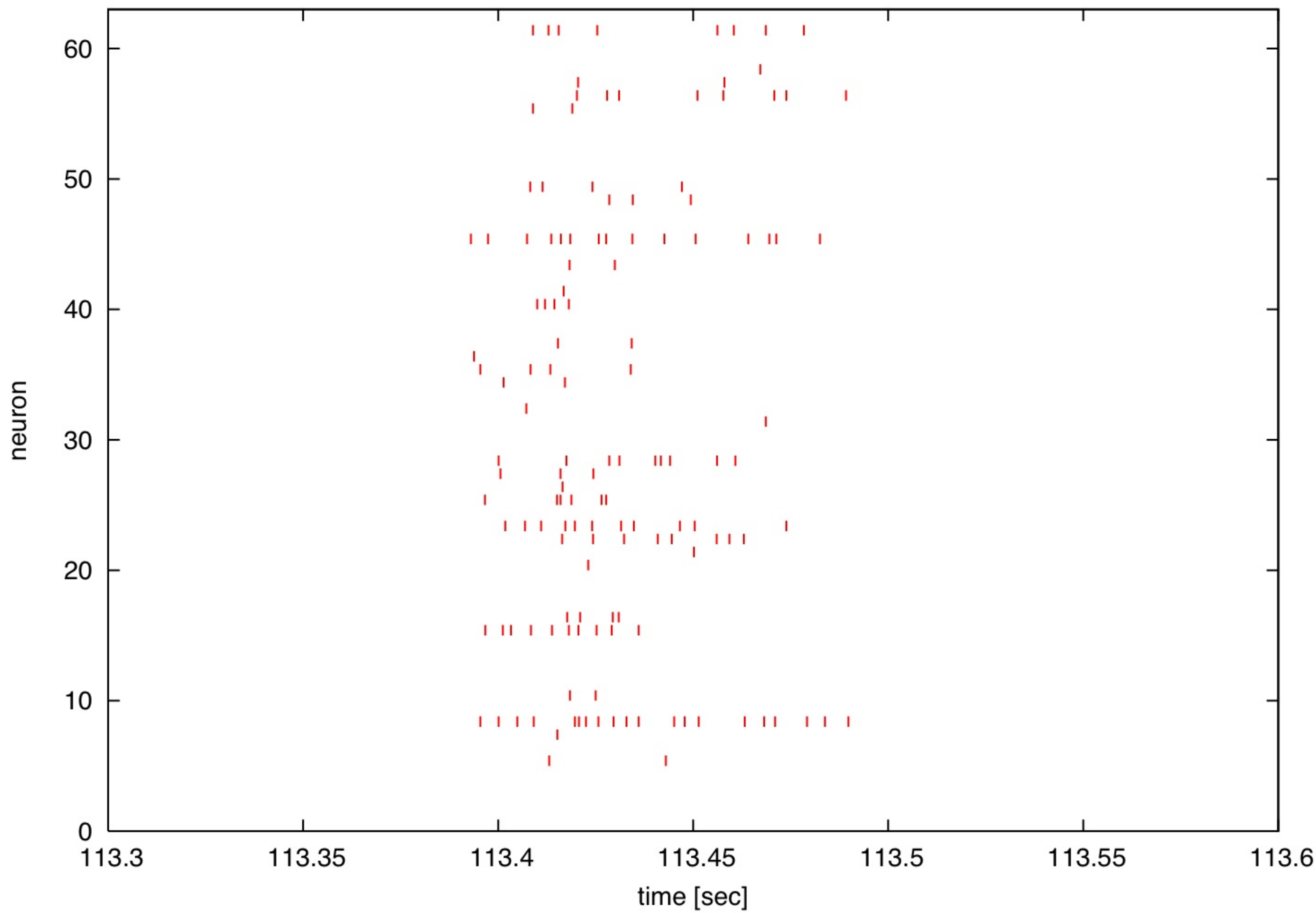


Image courtesy M. Ham and G. Gross

Cortical neural network electrophysiological activity

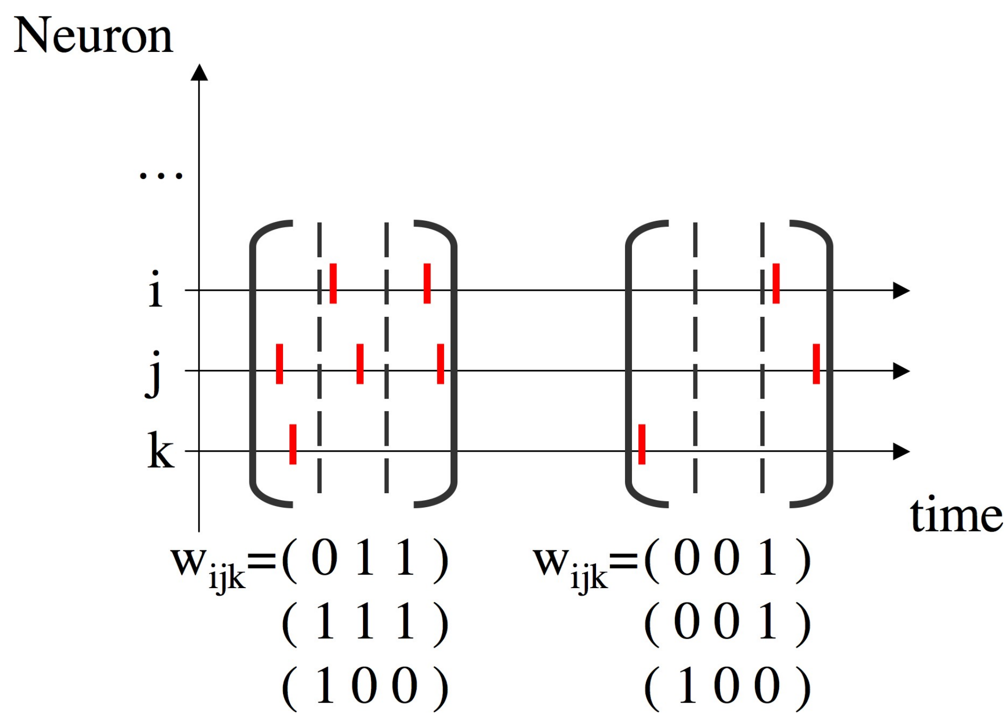


Estimation in practice

Binary 'words' from spike time series

"Spikes"

Rieke, Warland, de Ruyter van Steveninck, Bialek



And count word frequencies over time $p_w = \dfrac{n_w}{N}$

Motif search and identification
optimization in uncertainty reduction

Maximize the unique information gain R_n one order at a time. Conveniently, the expansion

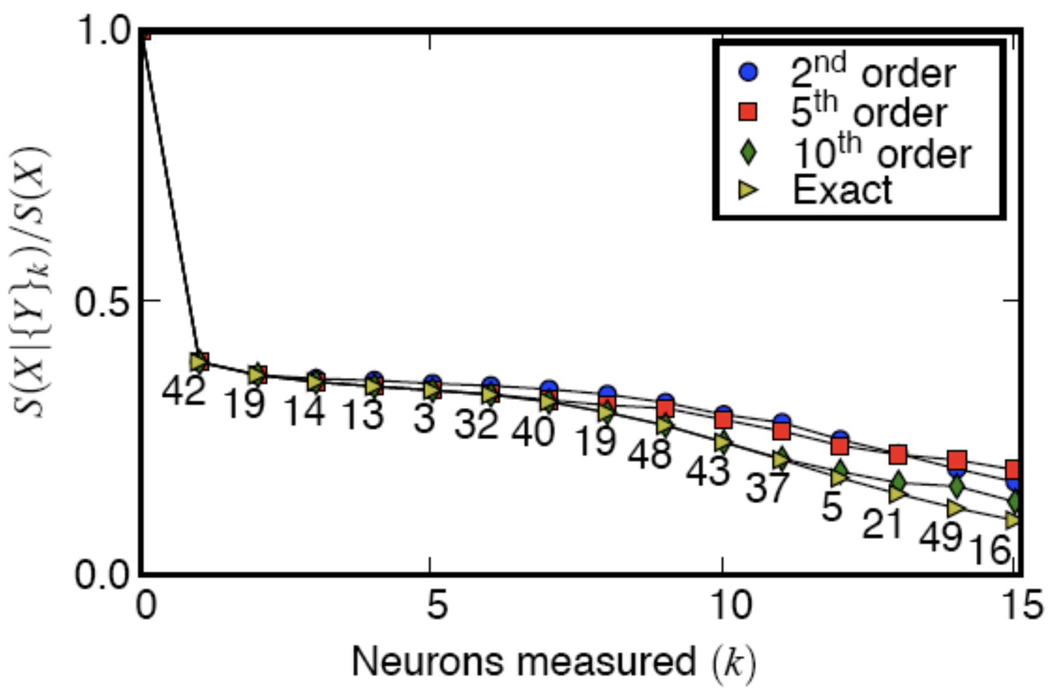
$$S(X|Y_1, Y_2, ..., Y_n) - S(X) = \sum_i \frac{\Delta S(X)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(X)}{\Delta Y_i \Delta Y_j} + \ldots + \frac{\Delta^n S(X)}{\Delta Y_1 \ldots \Delta Y_n}$$

allows us to do so order by order!

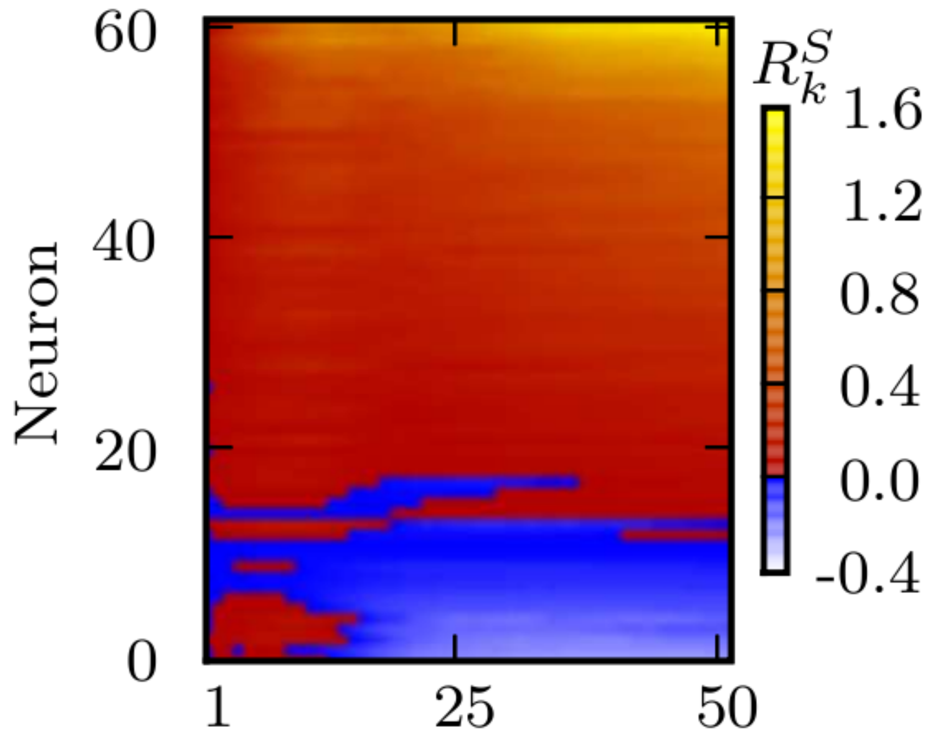
This is analogous to other optimization strategies such as the method of steepest descent.

Approximate searches

- Want to maximize $I(X;\{Y\}_n)$ by choosing best set $\{Y\}_n$
- Computationally expensive for $n > 10$
- Instead truncate expansion to $k < n$ and maximize
- How does the set $\{Y\}_n$ found using the approximation compare to the set found by using the exact expression?



global redundancy or synergy



(a) Neurons measured (k)

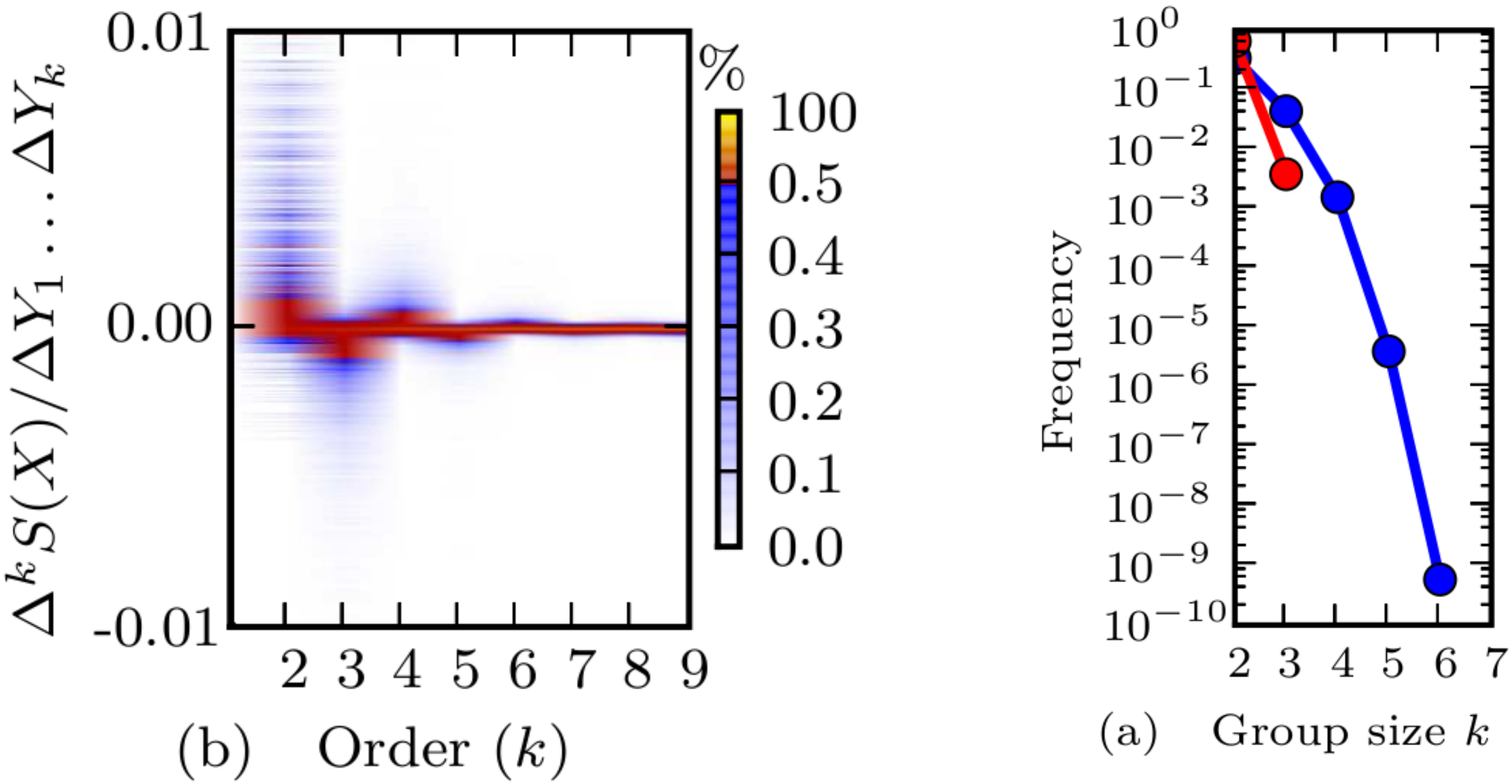
$$R^S(X;\{Y\}_k) = \sum_{i=1}^{k} I(X;Y_i) - I(X;\{Y\}_k)$$

Most neurons are part of globally redundant ensembles;

some are synergetic

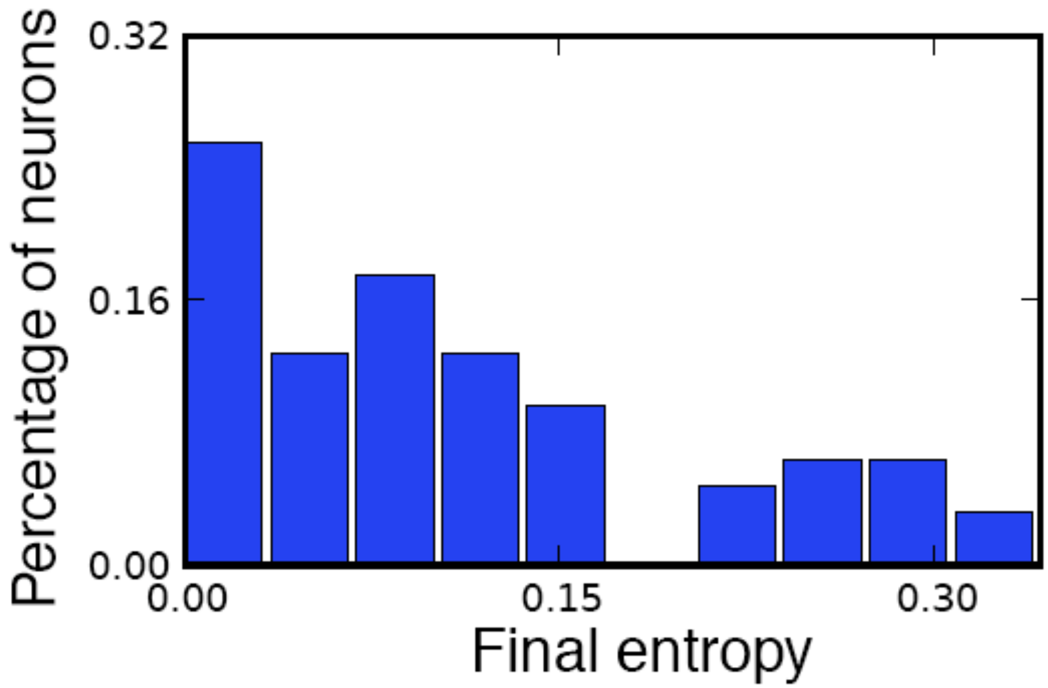
Information character of ensembles

order by order



Randomness or Structure?

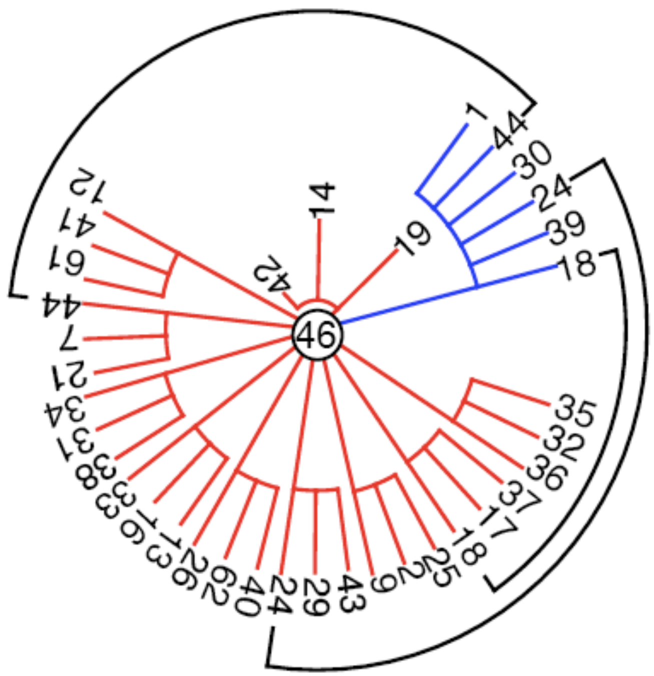
Individual uncertainty is accounted for by other nodes



After all available neurons are measured, very little $(0 - 30\%)$ of each neuron's initial entropy remains!

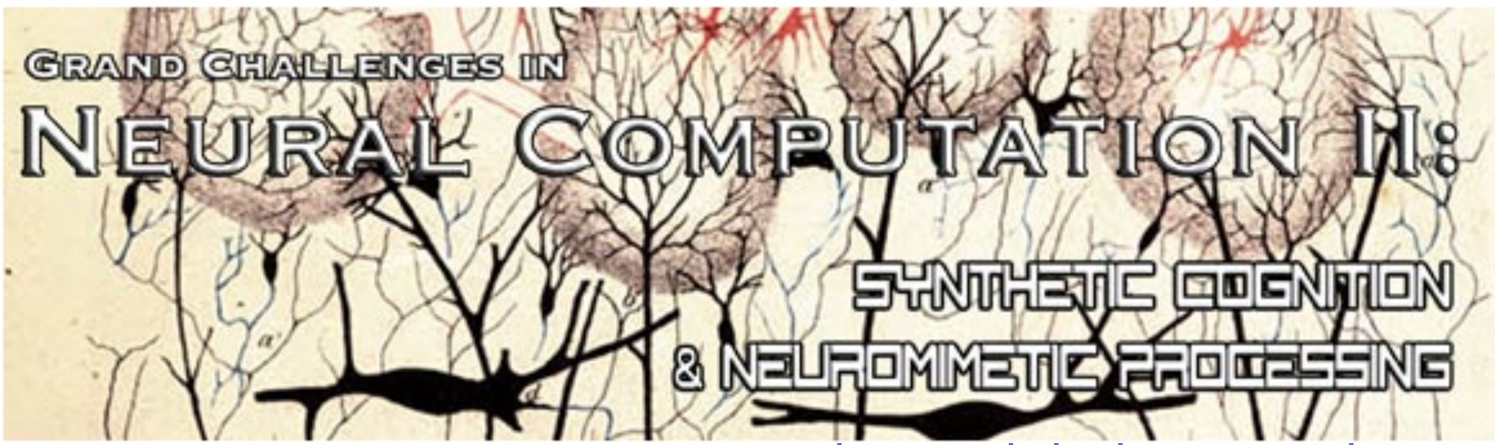
Reverse engineering network circuits

Seek purely redundant and synergetic “cores” – a set of neurons and all possible subsets that share the same functional character



Together this procedure allows us to reverse engineer all circuits associated with a target neuron

http://cnls.lanl.gov/neuralcomp



http://cnls.lanl.gov/neuralcomp2

Grand Challenge Questions

1) Are we close to understanding systems-level neural computation? Why or Why not?

2) What are the key scientific challenges or technologies for achieving an understanding of neural computation?

3) Are there applications for neuromimetic processing that can lead to better technologies immediately?

NOVAWEDNESDAYS

How Does the Brain Work?

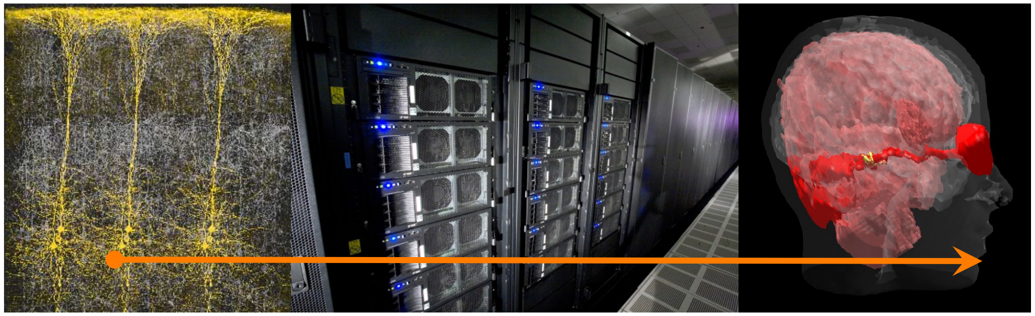
Investigate the psychology of magic tricks, magnetic wands that treat depression, artificial intelligence, and more. **Aired February 2, 2011 on PBS**

Posted 02.02.11

NOVA scienceNOW

Watch How Does the Brain Work?

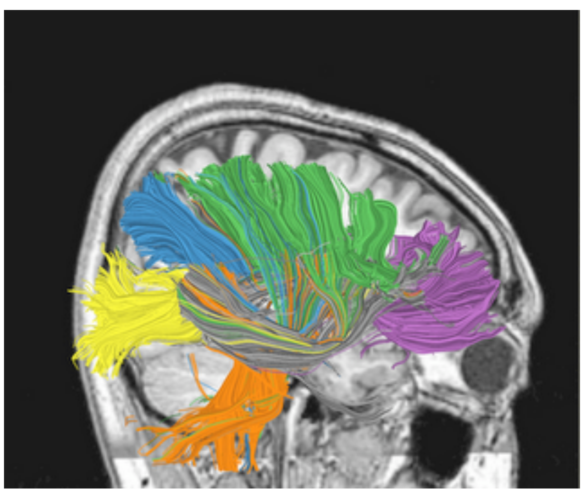
Aired February 2, 2011 on PBS



http://www.synthetic-cognition.edu

Blue Brain Project

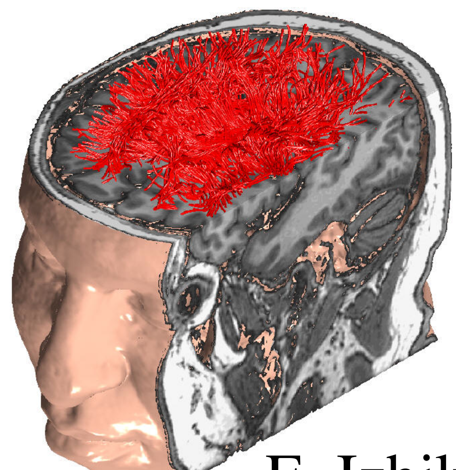
EPFL



IBM

DARPA SyNAPSE

DARPA DSO
DEFENSE SCIENCES OFFICE



E. Izhikevich 2008

The problem of scale

Units: 10^{11} neurons

10^{15} synapses

Temporal rates 10 Hz vs. GHz computer

Performance: 10 Petaflops vs. Roadrunner 1.1 Petaflops

Tianhe 2.5 Petaflops

Energy consumption 25 W vs. Roadrunner 2.345 MW

[2300 US households]

Memory 100 Tbytes vs. Library of congress 160Tb

1 second of internet traffic

Visual Experience 20Mpixel/s 30 PetaPixels/lifetime

Grand Challenges in Systems Neuroscience

1) The problem of **function** [information processing]

What's the system computing?

2) The problem of **representation** [memory]

How is information represented (disentangled!) in the brain?

3) The problem of **learning** [self-organization]

How does the brain self-assemble?
How do representations change with experience?

Work supported by
LANL LDRD program
National Science Foundation
DARPA

References

- L. M. A. Bettencourt, G. J. Stephens, M. I. Ham and G. W. Gross
  PHYSICAL REVIEW E 75, 021915 (2007).
- L. M. A. Bettencourt, V. Gintautas, and M. I. Ham
  PHYSICAL REVIEW LETTERS 100, 238701 (2008).
- M. I. Ham, L. M. A. Bettencourt, G. W. Gross, and F. D. McDaniel.
  Journal of Computational Neuroscience, 24, 346-357 (2008).
- M. I. Ham et al.
  Biological Cybernetics 102 (2010)
- L. M. A. Bettencourt
  Topics in Cognitive Science 1 (2009) 598–620

See also

http://synthetic-cognition.lanl.gov