XXIV$^{ème}$ Colloque Gretsi
4 septembre 2013

# Codes sur graphes et mémoire cérébrale

Claude Berrou,

Vincent Gripon, Olivier Dufor, Xiaoran Jiang et Behrooz Kamary Aliabadi

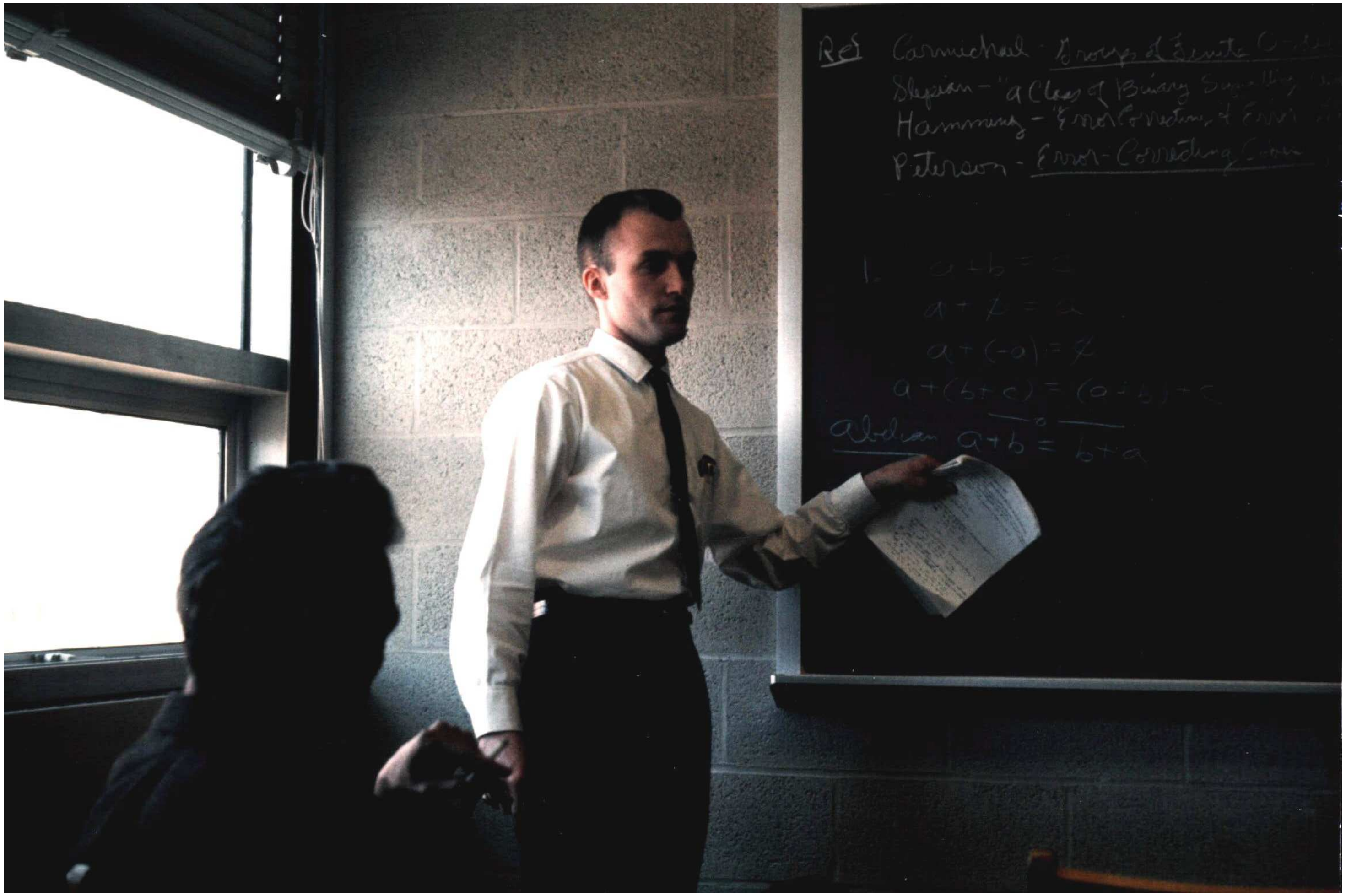
INSTITUT Mines-Télécom

TELECOM Bretagne

Lab-STICC

CNRS

European Research Council
Established by the European Commission




Jim Massey (1934-2013)

# Communication model

Source → Source coding → Channel coding → destination

**Parsimony** (remove useless redundancy)

**Robustness** (add smart redundancy)

relevant to:

- telecommunication systems
- storage devices

? cerebral memory

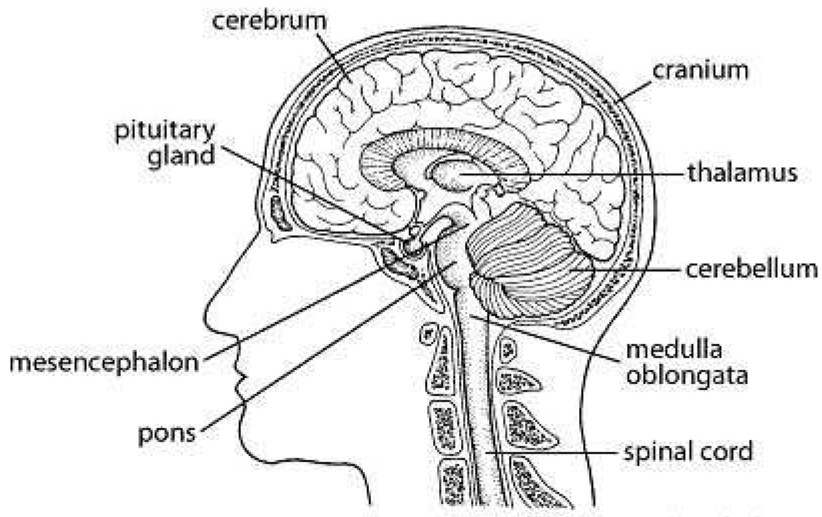
# Communication model

Nervous information | Mental information

Physical world → Source coding → Channel coding → Mental world

richly detailed

fleeting




parsimonious

durable

Mental information is robust and durable, therefore must be redundantly memorized.

We are interested in

« Pure mental information »

02 29 00 13 06

9x8 = 72

« It so happens I am sick of being a man… »

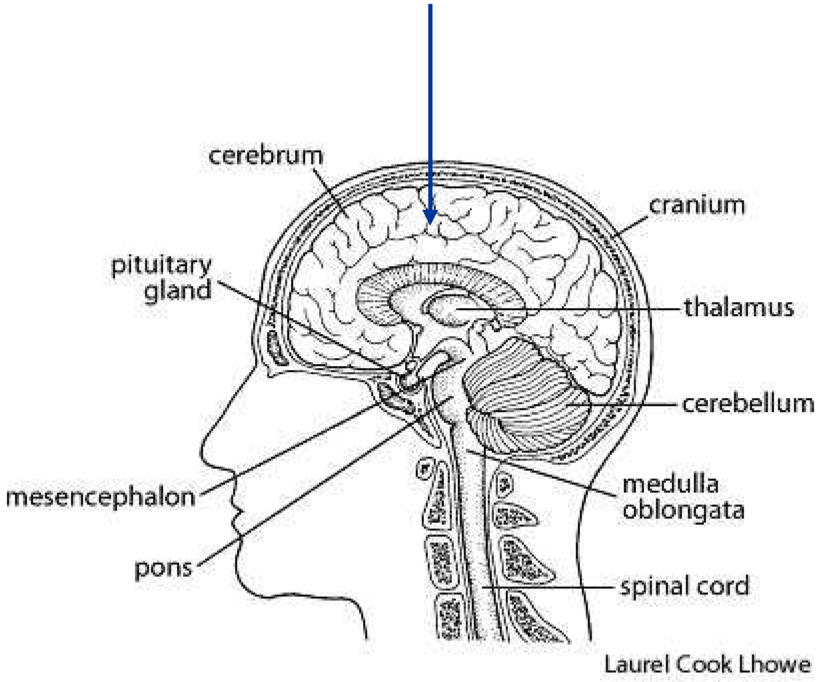
$$H = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

## Invaluable redundancy!

**Intelligence**

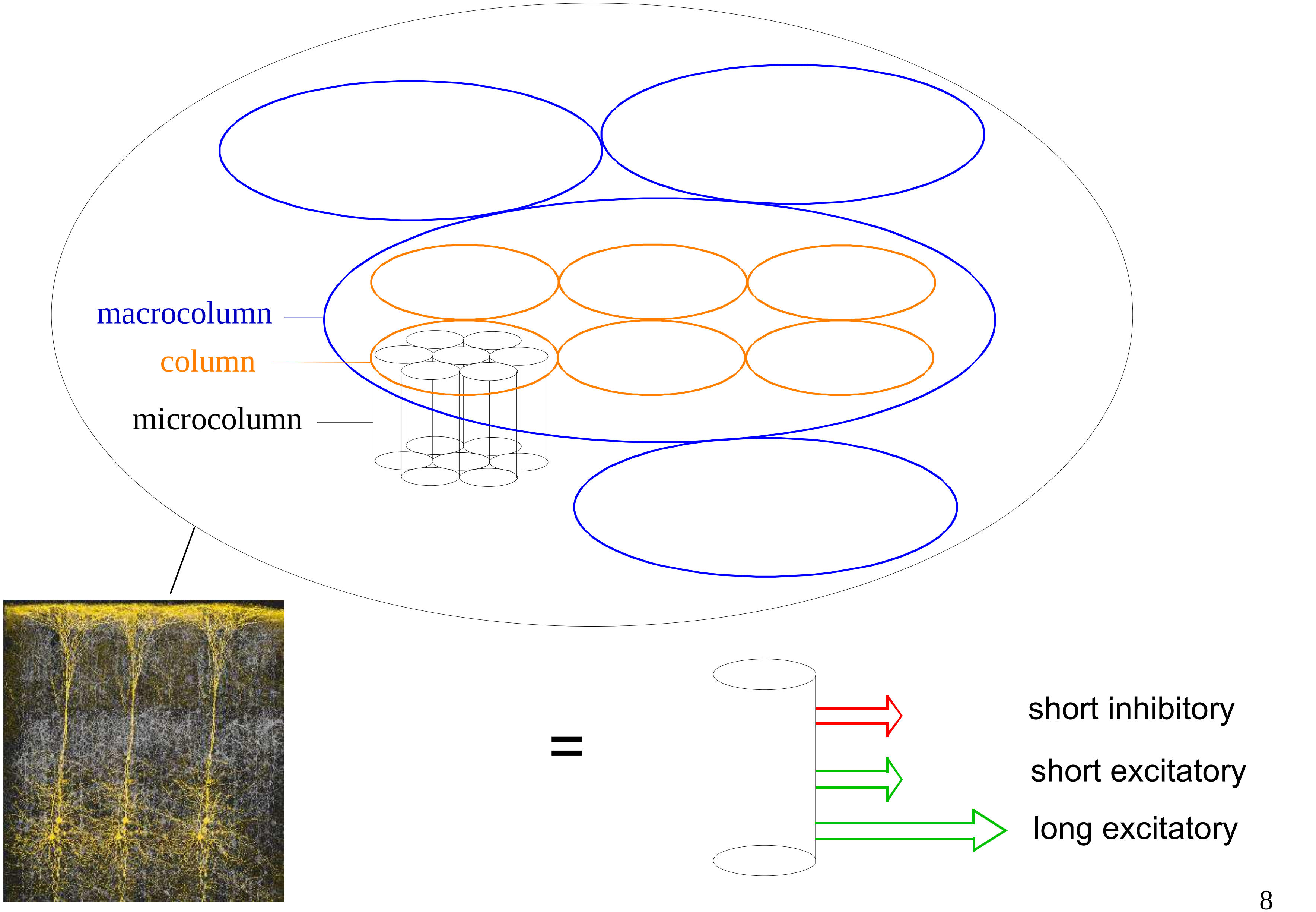
brain

Contrary to ancestral sensory and motor feed-forward circuits, the neocortex can be essentially regarded as a very recurrent organized graph




The self-repeating unit (node) in the graph is the so-called microcolumn (~ 100 neurons)

# Functional area of the cerebral cortex



So, distribution, quasi-random graph, redundancy, message passing, etc.

$\rightarrow$ the neocortex behaves like a distributed decoder!

Binary signalization: (0 or 1) $\longleftrightarrow$ (Neuron inactive or firing)

(inhibitory signals are only for control)

Astronomic number of combinations

Fixed point decoding $\longleftrightarrow$ Non confused, single thought

Large minimum distances $\longleftrightarrow$ Easily separable thoughts

Resilience

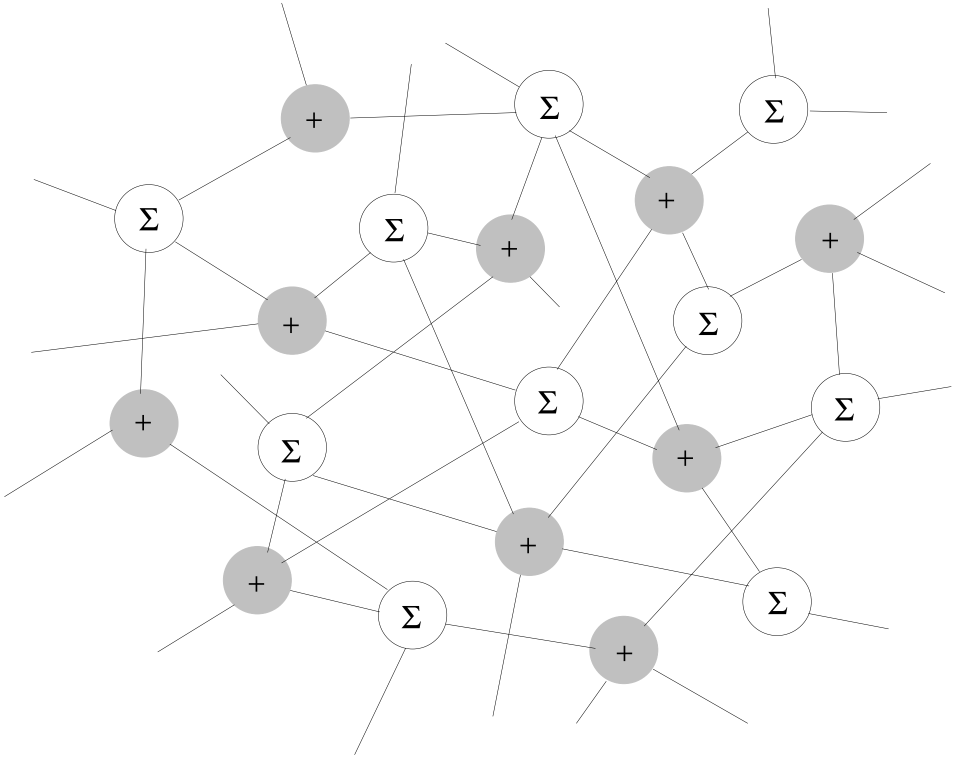
Linearity $\longleftrightarrow$ Nonlinearity

Importance of cycles

Importance of correlation

# The neocortex behaves like a distributed decoder!
# Which one?

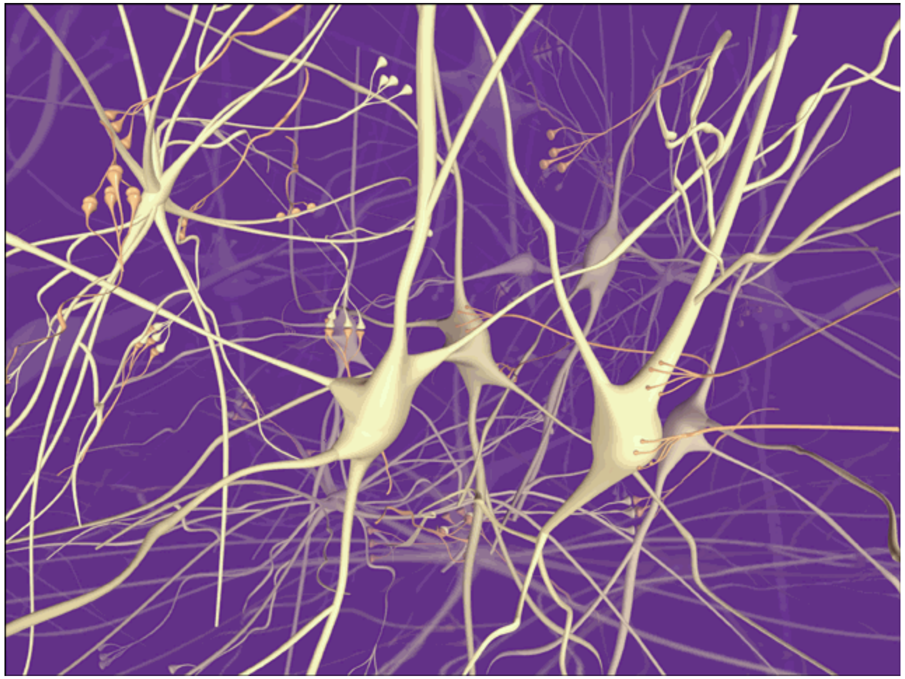
LDPC decoder



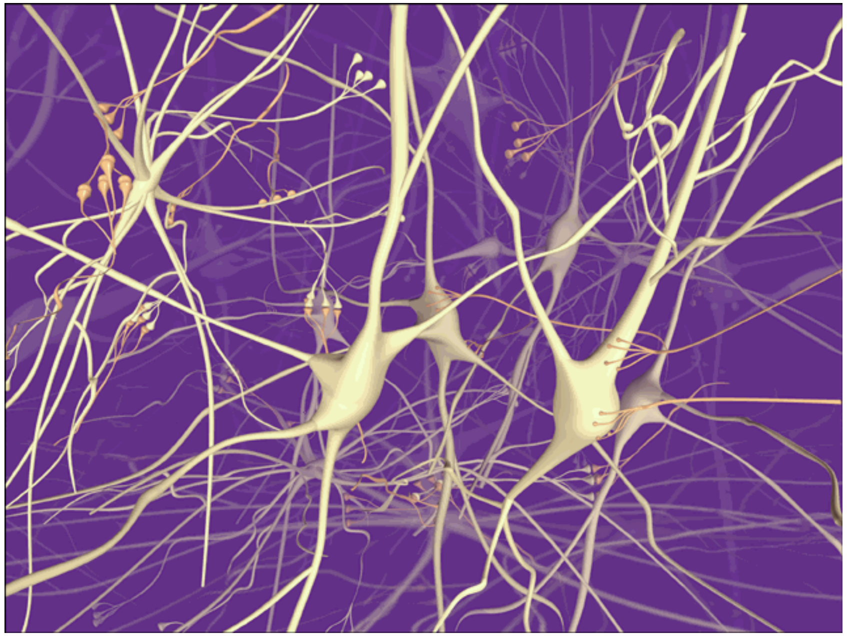

+ : parity processor

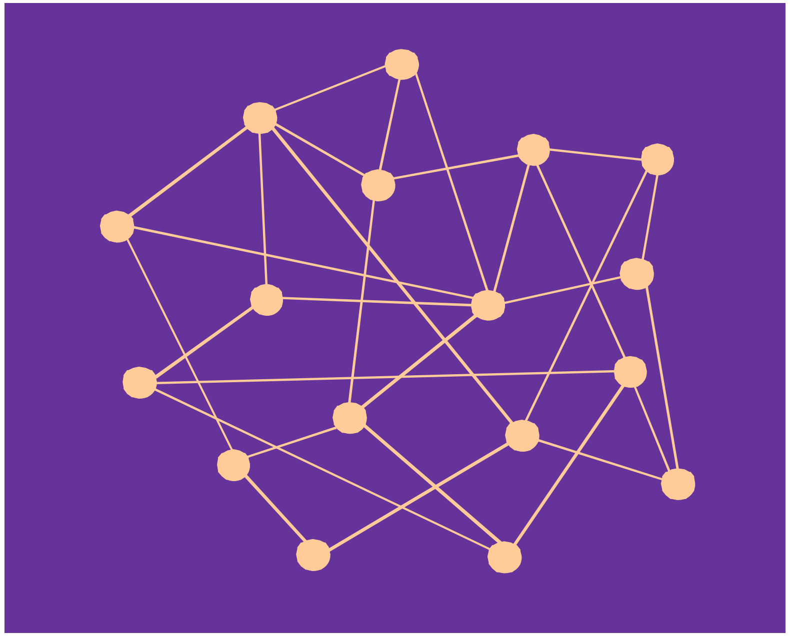
Σ : variable processor

Cortical decoder



# What is the code?

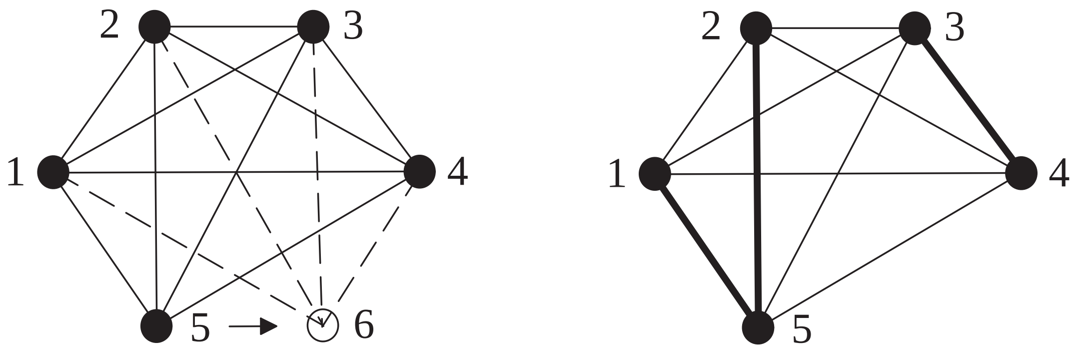




A redundant, distributed, graphical code !

# What is the code?

## The fundamental brick: the clique




(a) (b)

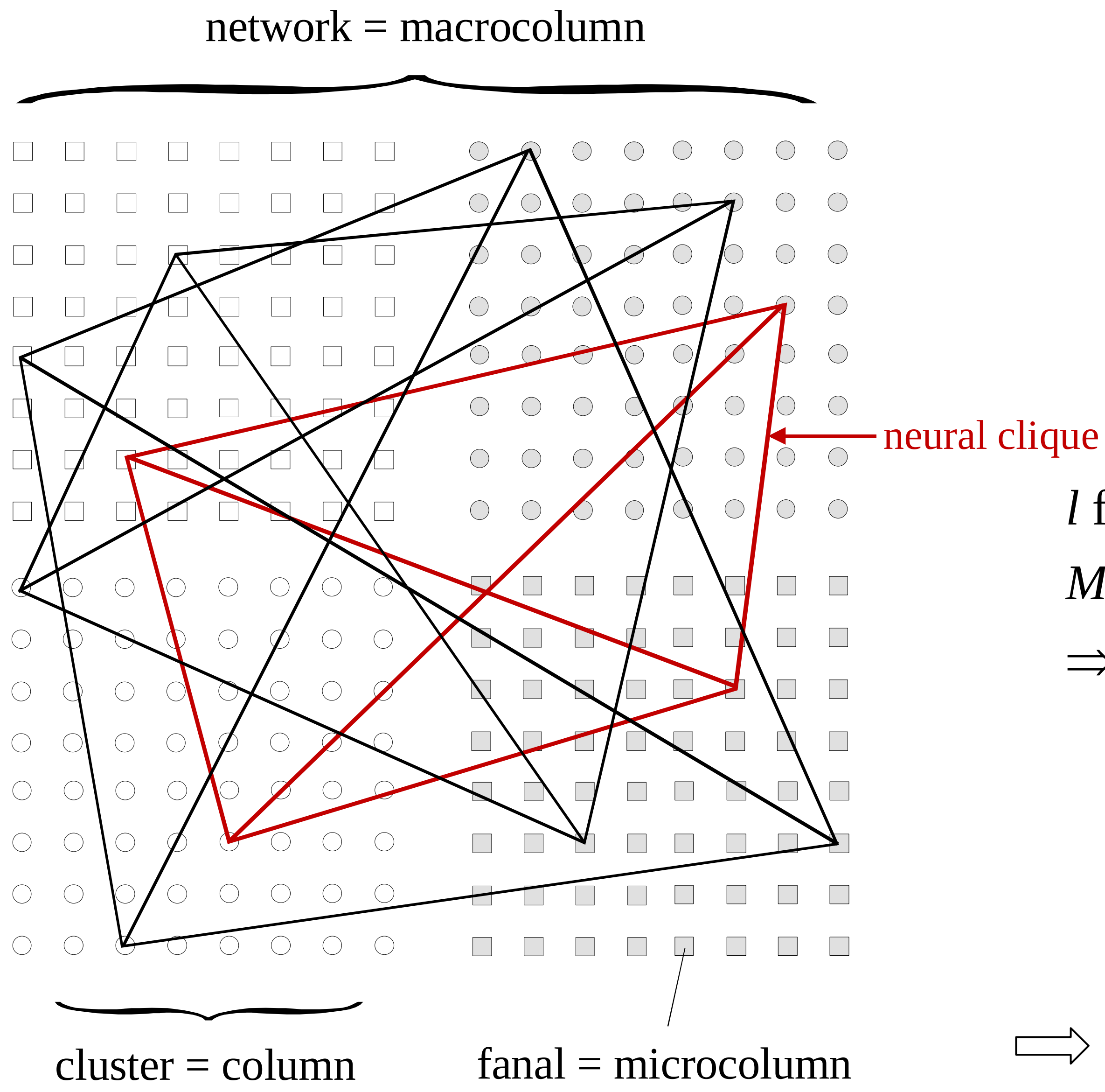
$c$ vertices: $d_{\min} = 2(c-1)$ $R = \frac{\left\lfloor \frac{c+1}{2} \right\rfloor}{\frac{c(c-1)}{2}} = \frac{1}{c-1}$ (for $c$ even)

$$F = Rd_{\min} = 2$$

V. Gripon and C. Berrou, “Sparse neural networks with large learning diversity”, *IEEE trans. on Neural Networks*, vol. 22, n° 7, pp. 1087-1096, July 2011

V. Gripon, V. Skachek, W. J. Gross and M. Rabbat, “Random clique codes”, *ISTC’12*, Gothenburg, Sweden, 2012

# In order to control the cliques, the graph is structured according to the neocortical architecture




$l$ fanals per column,

$M$ messages:

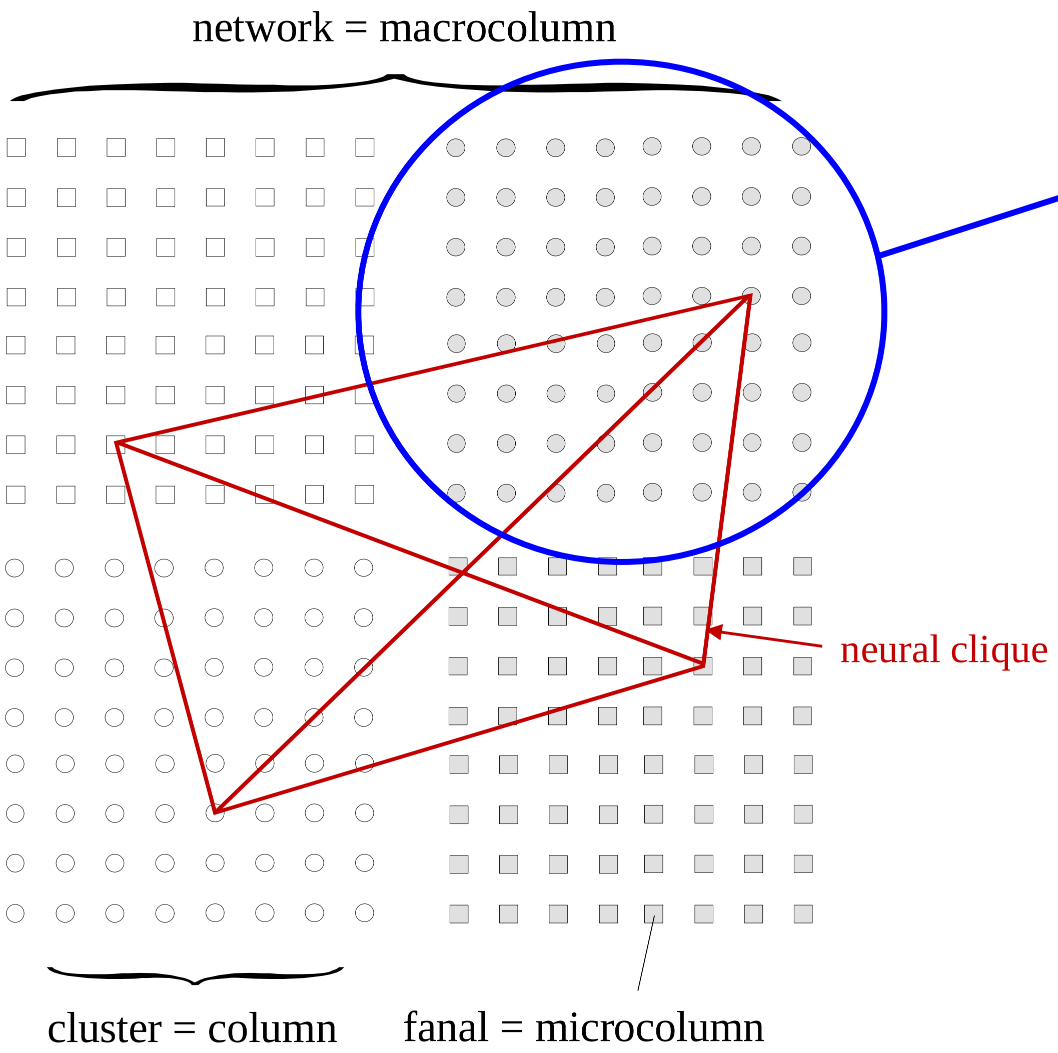
$\Rightarrow$ density $d$:

$$d = 1 - \left(1 - \frac{1}{l^2}\right)^M$$

$$\approx \frac{M}{l^2} \text{ if } M << l^2$$

$$\Longrightarrow \quad M \approx dl^2$$

# Concatenation of simple and thrifty codes




a constant-weight code(*) with length $l$ and weight $w = 1$

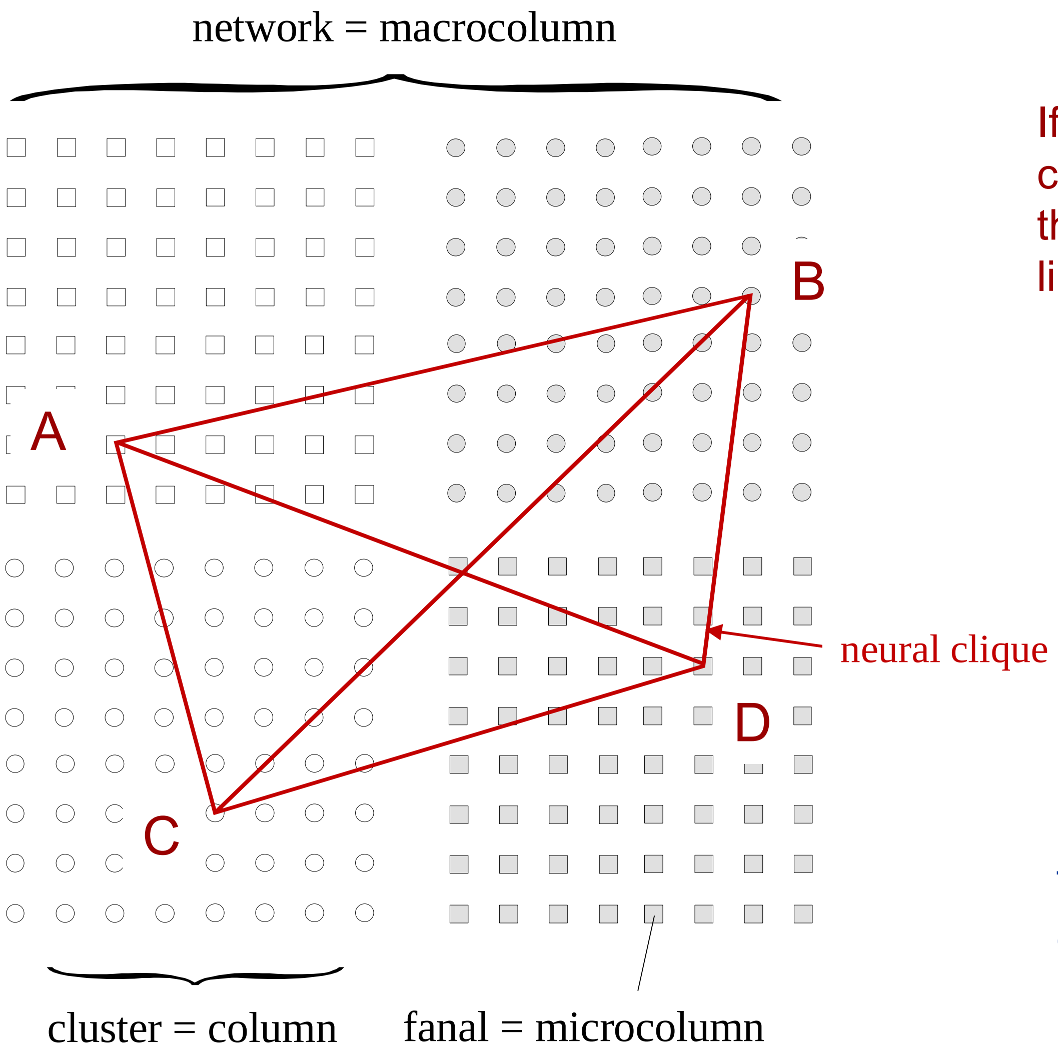
$k = \log_2(l)$ bits $\Rightarrow$

$R = \log_2(l)/l$

$d_{min} = 2$ only but easy to decode according to the *winner-take-all* (WTA) rule (max function)

(*) F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*, pp. 526-527, North-Holland, 1979.

# Decoding: relying on correlation!!!




If A and B are both connected to C and D, then A and B are very likely to be connected.

Iterate:

- message passing through established connections
- local winner-take-all

# Application to associative memory

$c$ = 8 clusters, $l$ = 256 fanals

Messages of $8 \times \log_2(256) = 64$ bits

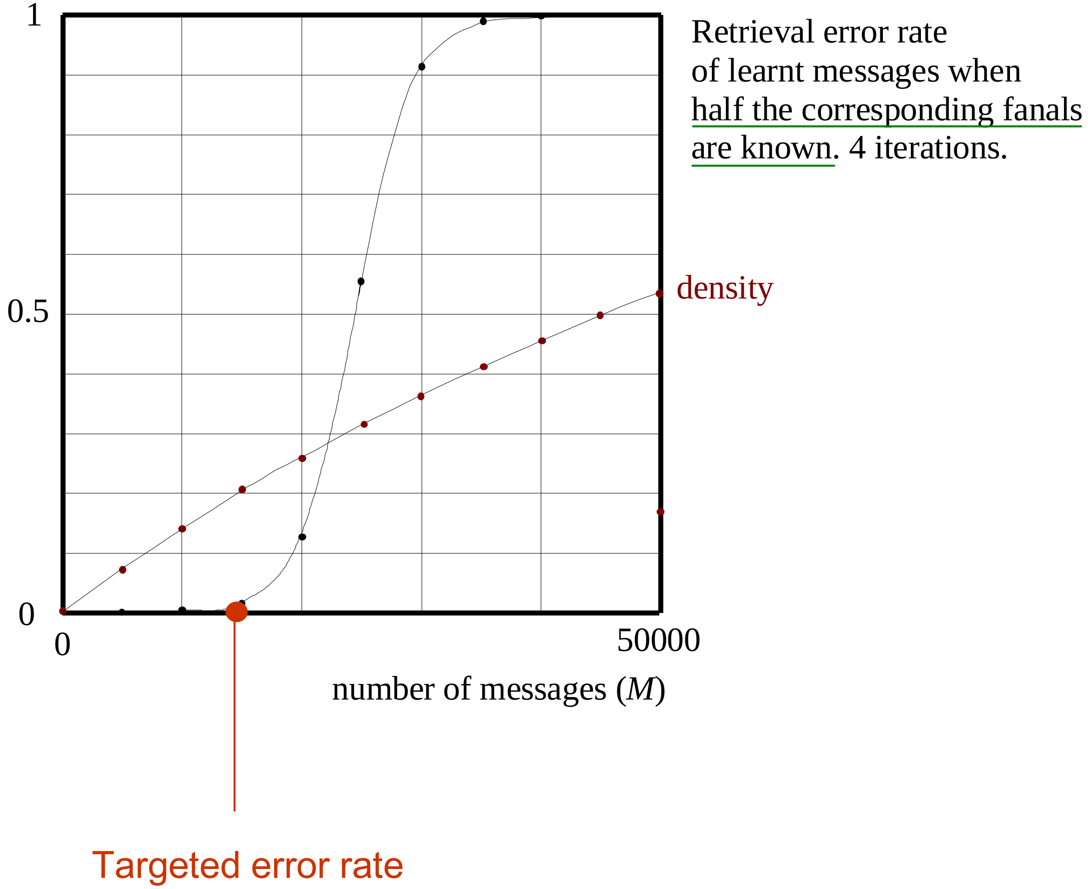
Gains compared to Hopfield networks (with the same amount of memory used):

diversity: 250

capacity: 20

efficiency: 20

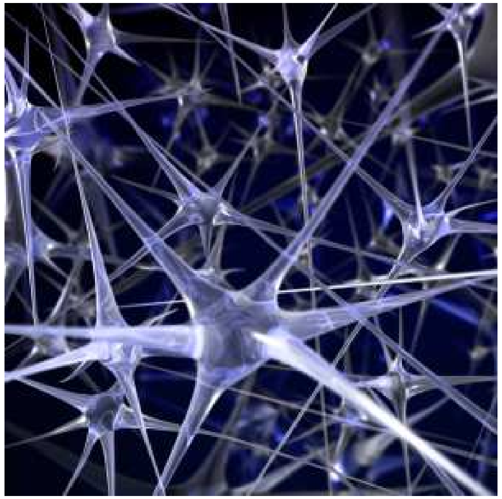
(52%  instead of 2.6%)



# Analog versus digital

Telecommunications

Neural networks



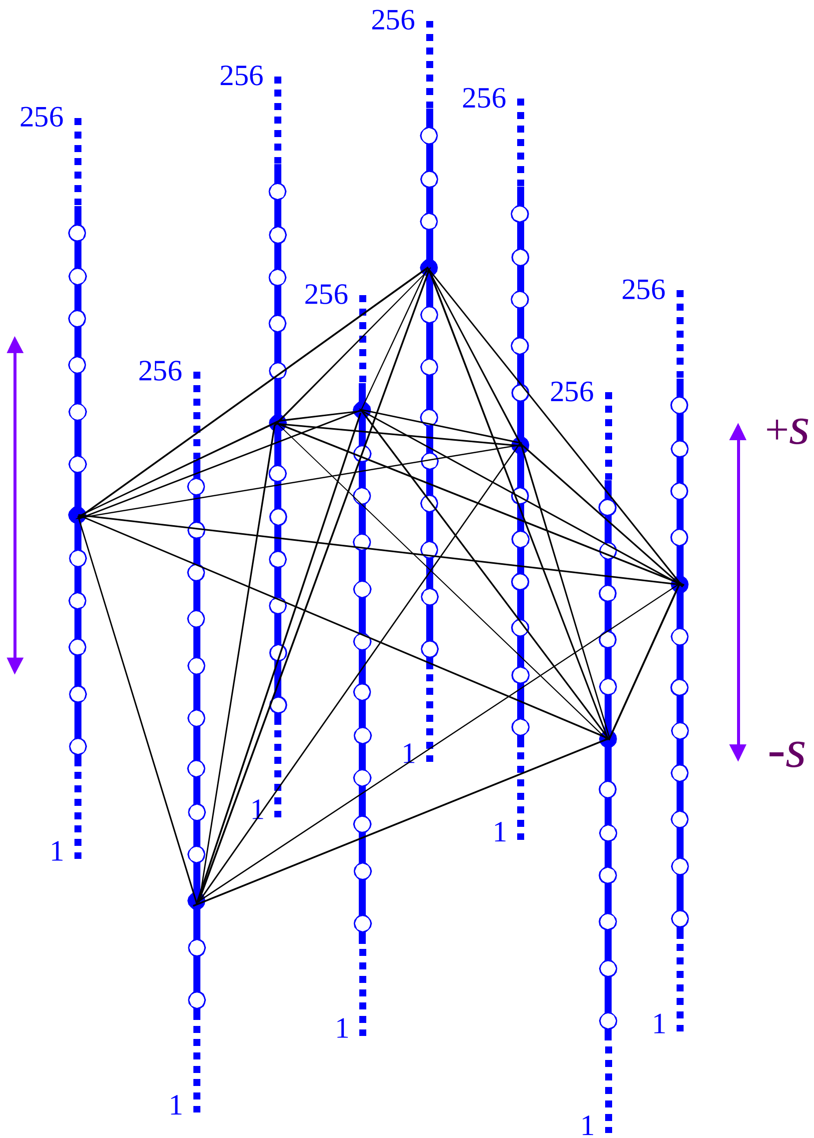
Substituting digital for analog: considerable gains in capacity

So, this natural question arises:

Isn’t our long term memory digital?

(compare with DNA)

# Associative memory with blurred stimuli




$c$ = 8 clusters, $l$ = 256 fanals

Messages of $8\text{x}\log_2(256)$ = 64 bits

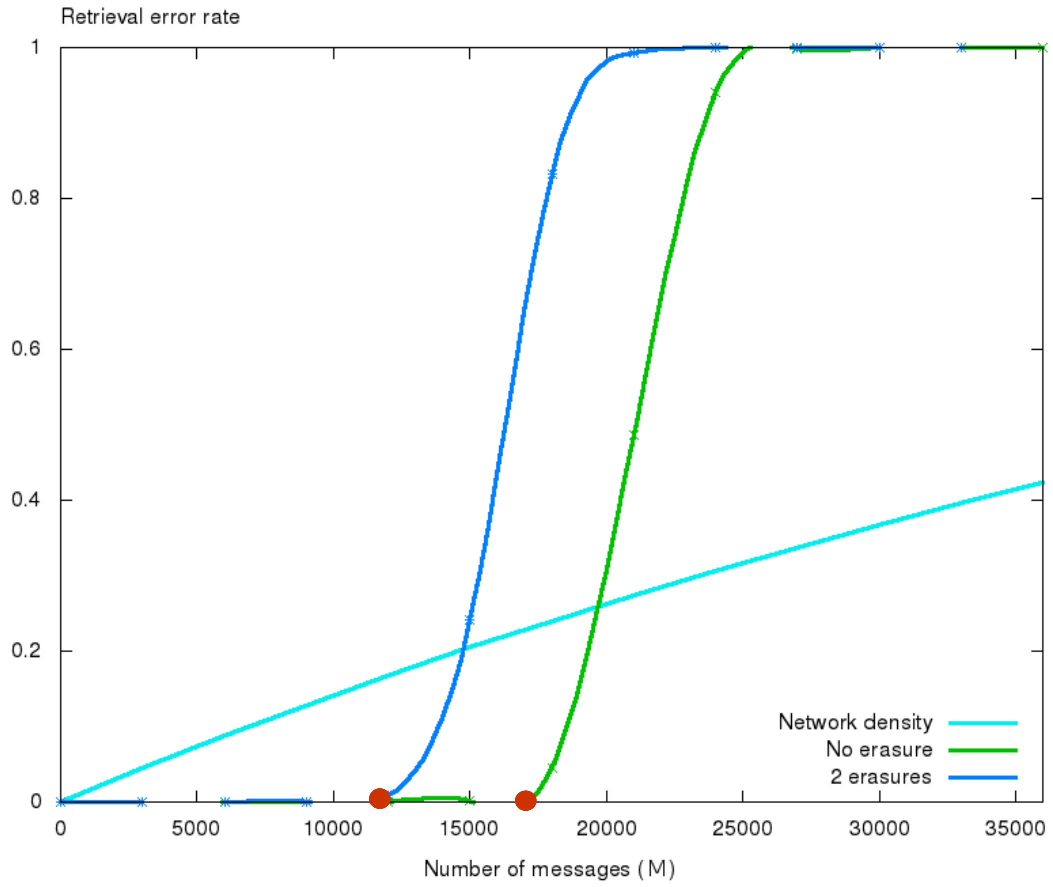
Fanals are approximately known, in a certain vicinity [-$s$,+$s$].

## Associative memory with blurred stimuli

$c$ = 8 clusters, $l$ = 256 fanals

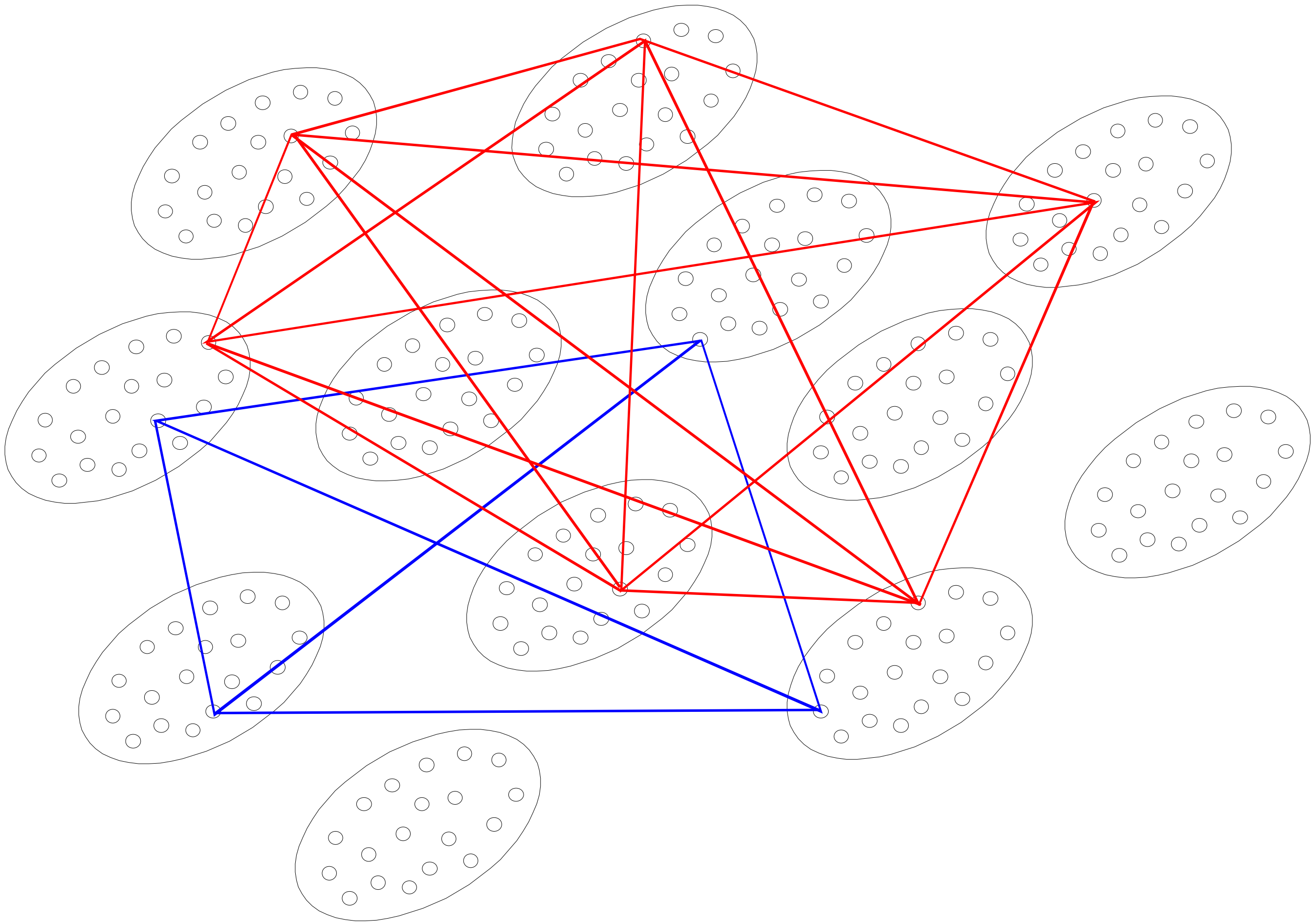
Messages of $8 \times \log_2(256) = 64$ bits

$s = 5$




4 iterations

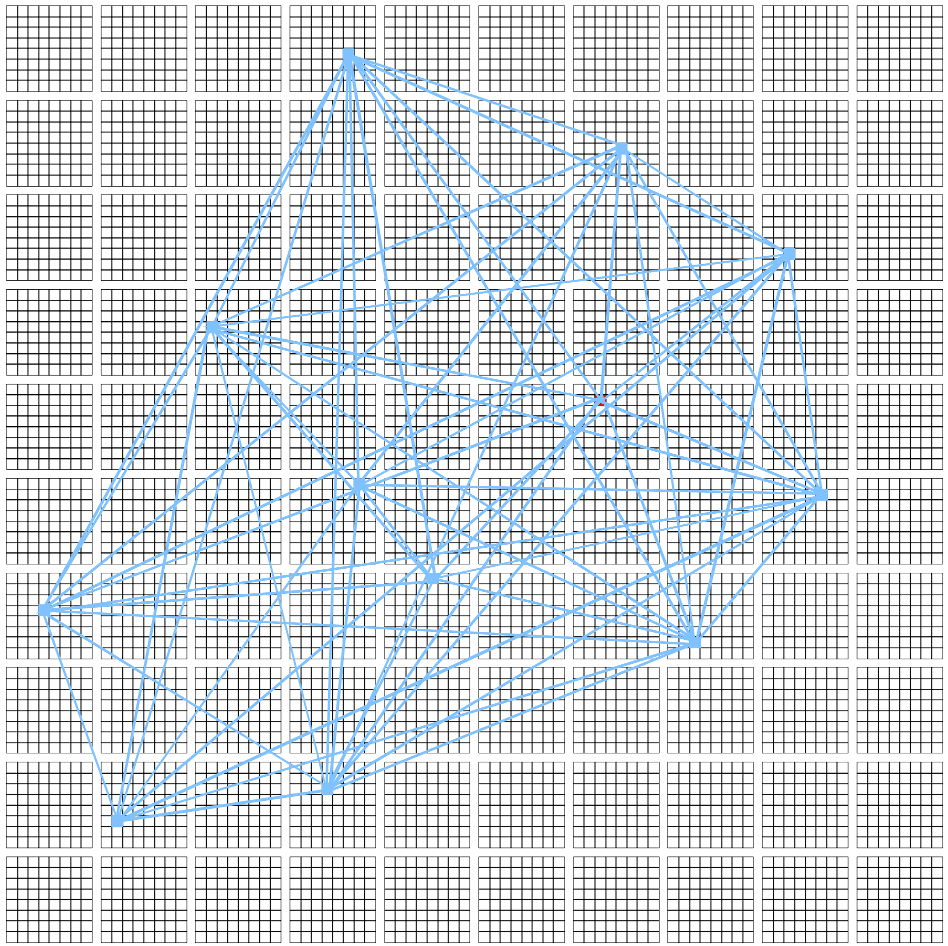
# Sparse messages



*M* proportional to $n^2$

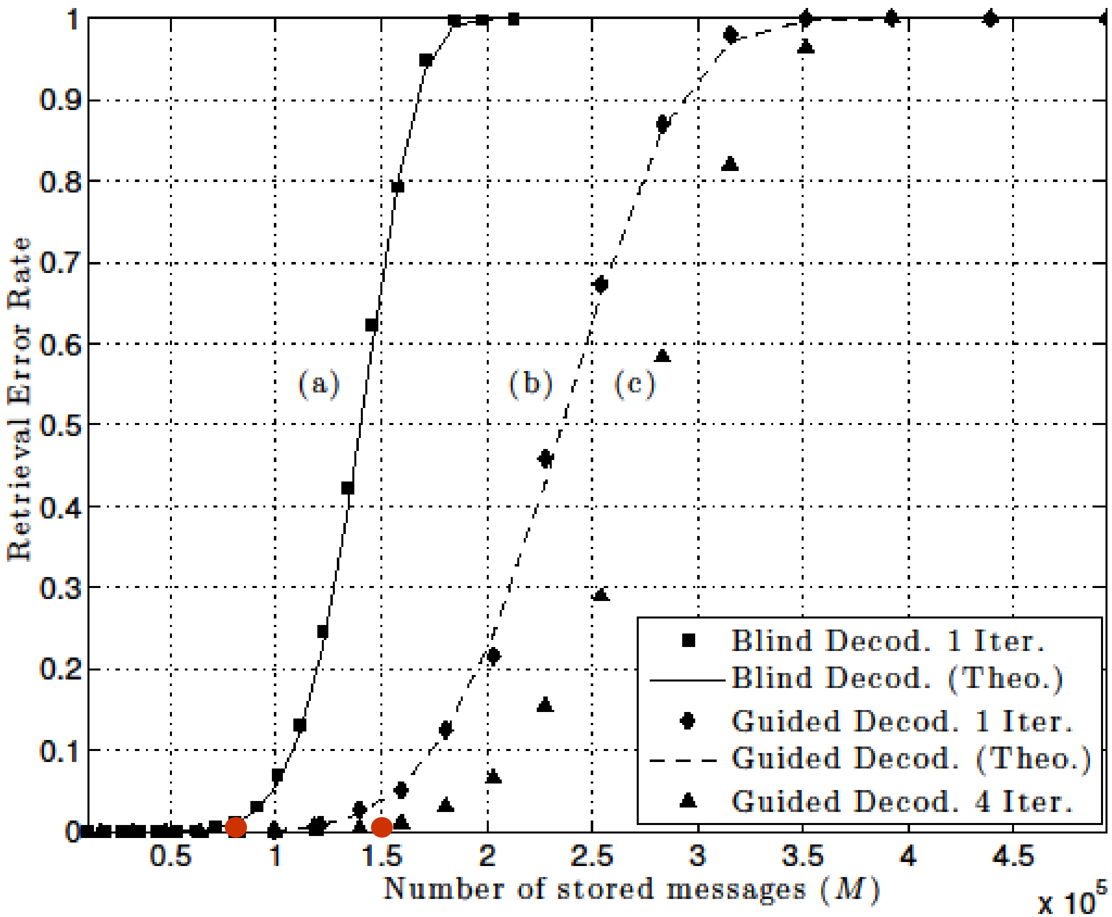
B. Kamary Aliabadi, C. Berrou, V. Gripon and X. Jiang, “Storing sparse messages in networks of neural cliques”, to appear in *IEEE trans. on Neural Networks*

# Sparse messages



100 clusters of 64 fanals (microcolums) each : about $10^{-5}$ x human cortex

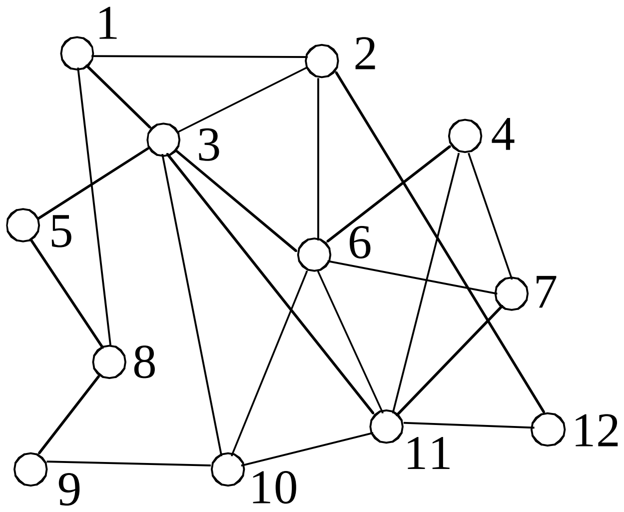
Cliques with $c$ = 12 vertices



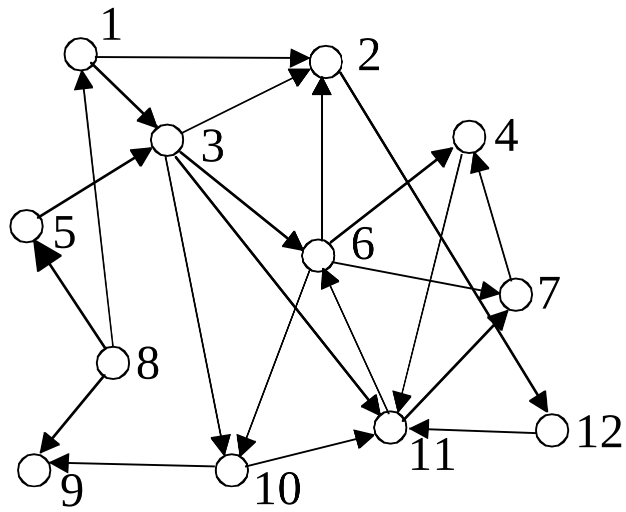

100 clusters of 64 fanals (microcolums) each : about $10^{-5}$ x human cortex

Cliques with $c$ = 12 vertices, $c_e$ = 3 vertices are not known

## To store sequences instead of atemporal messages: replace cliques with tournaments



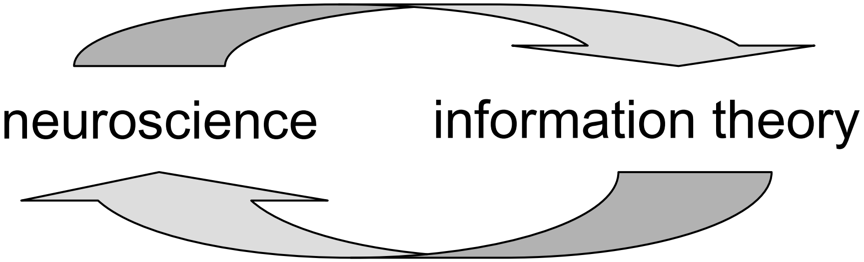

(a)




(b)

X. Jiang, V. Gripon and C. Berrou, "Learning long sequences in binary neural networks," *Proc. of Cognitive 2012*, Nice, France, July 2012

# Conclusion




a very promising cross-fertilization

Physical world → Source coding → Channel coding → Mental world

| Machine learning | Networks of cliques and tournaments |
|---|---|