# Ultra-Sparse Representations in Neural Networks: Biological Inspiration for Artificial Intelligence?

Simon Thorpe

Director of the Brain and Cognition Research Centre (CerCo)

Director of the Toulouse Mind & Brain Institute

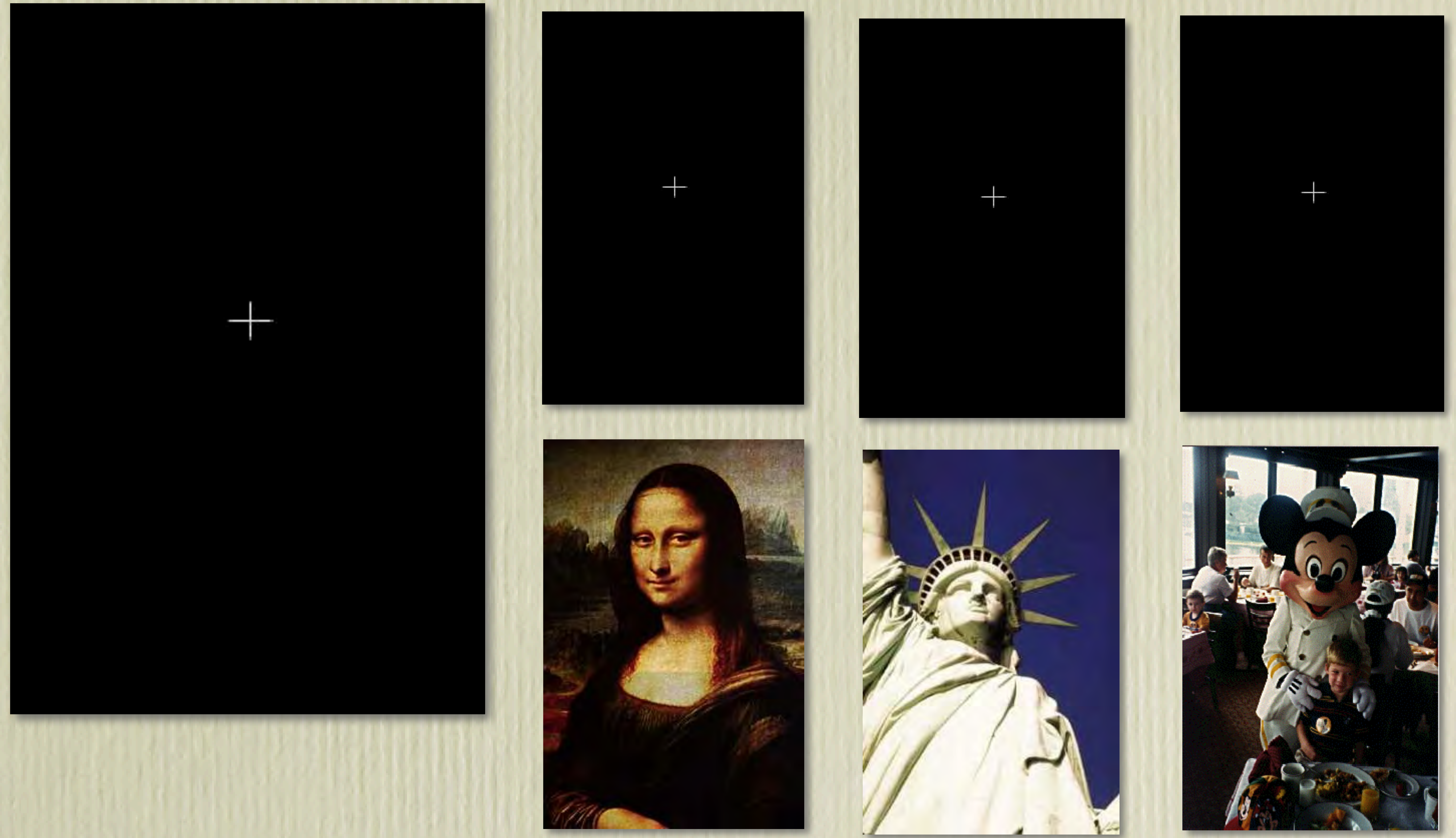
Toulouse, France

# Overview

- Fast Visual Processing
  - Biological Constraints
- The State of the Art
  - Human level performance with deep networks
- What's missing?
  - Spikes
    - Neurons don't send floating point numbers
    - Order of firing across populations - coding with one spike per neuron
  - Ultra-sparse processing
    - Grandmother Cells!
    - Neocortical Dark Matter!
- Towards a new computational paradigm
  - Fusion of Neuroscience and Computation

# Fast Visual Processing

- Rapid Serial Visual Presentation (RSVP)
- 1970s
  - Molly Potter
- RSVP at 10 fps



# Ultra Rapid Scene Categorisation

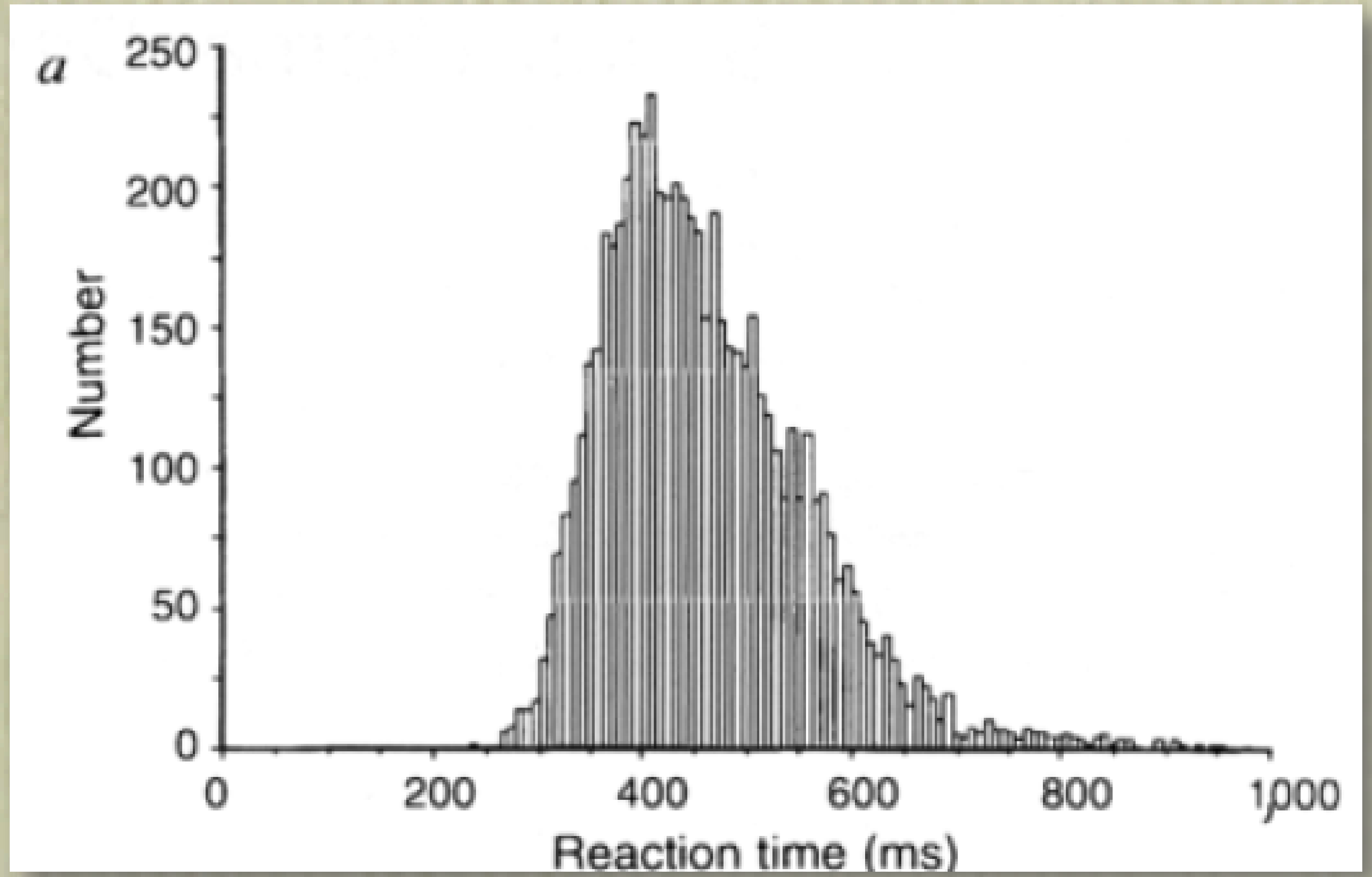
**Speed of processing in the human visual system**

**Simon Thorpe, Denis Fize & Catherine Marlot**

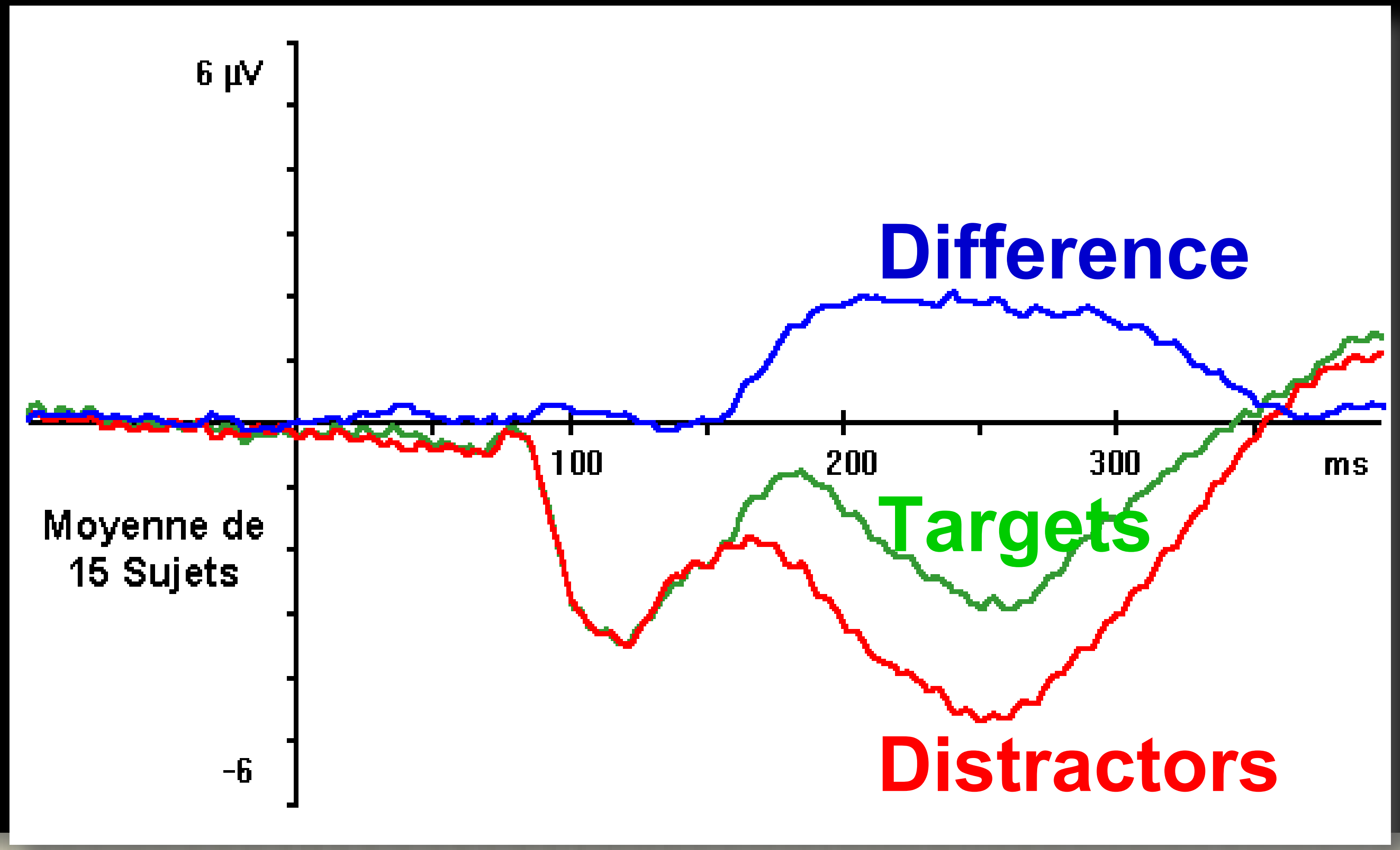
Centre de Recherche Cerveau & Cognition, UMR 5549, 31062 Toulouse, France

NATURE · VOL 381 · 6 JUNE 1996

Behavioural Reaction Times

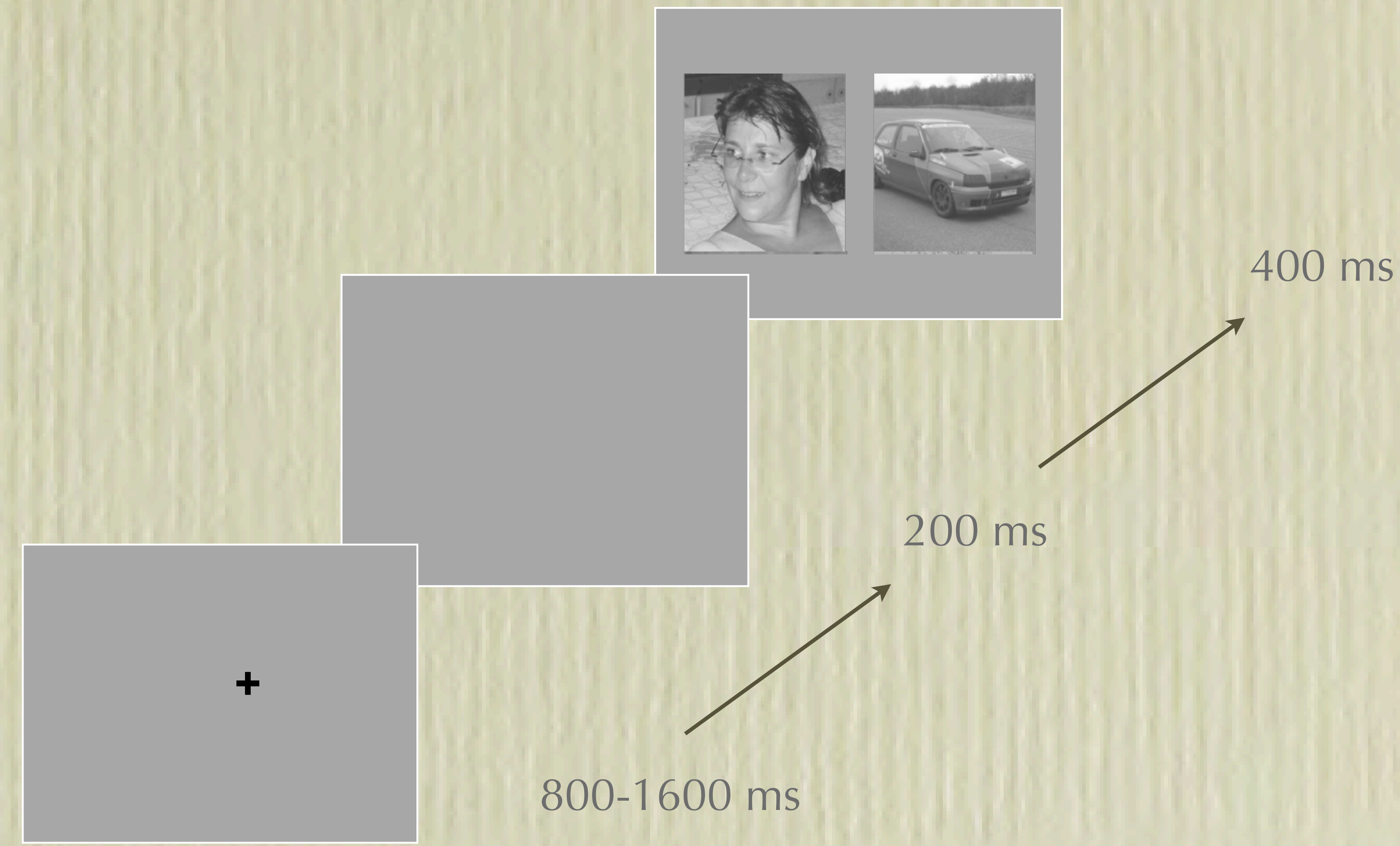



Event Related Potentials

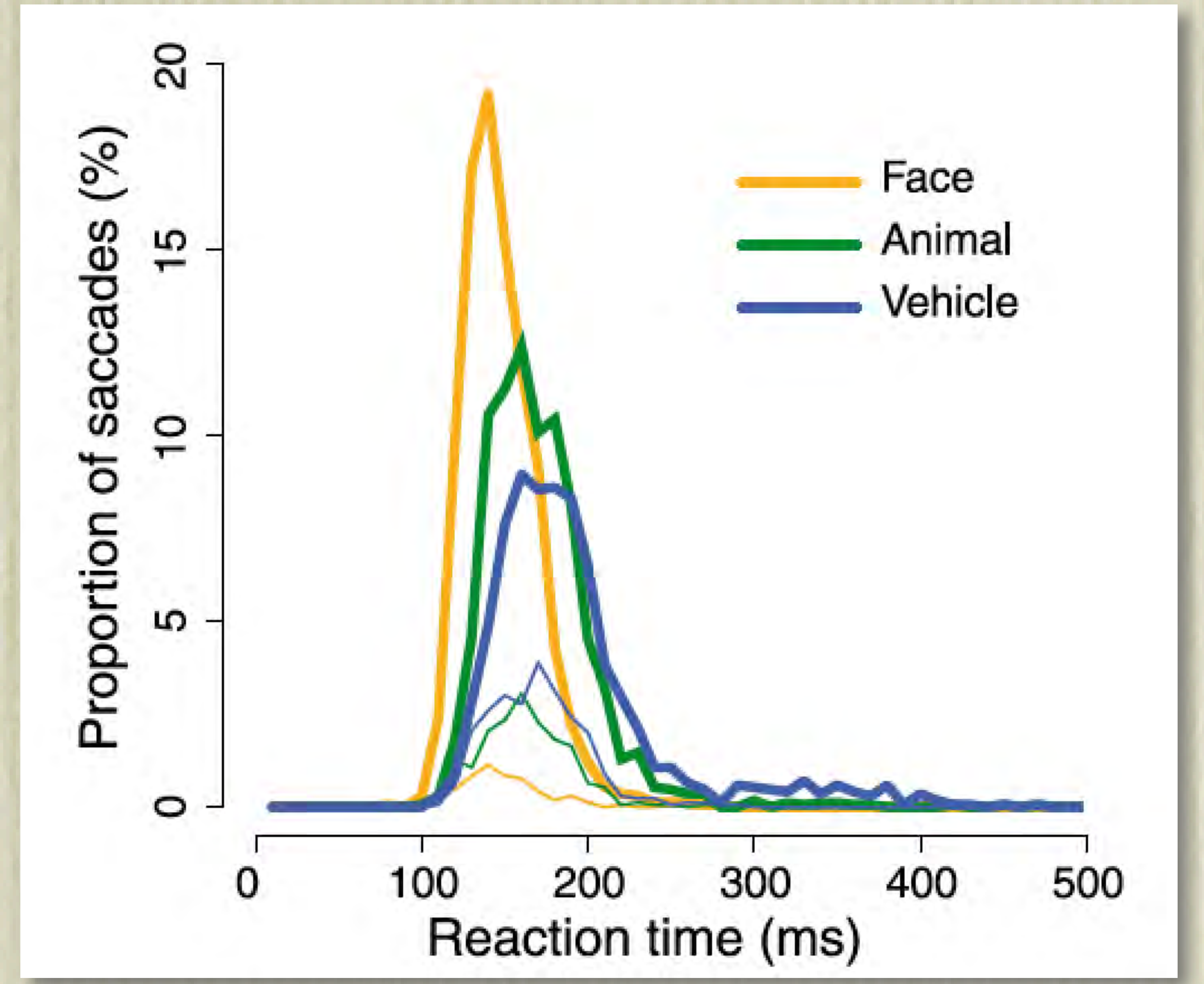



Scene Processing in 150 ms

# Saccadic Choice Task




Crouzet, Kirchner & Thorpe, 2010




- Saccades towards faces in 100 ms!
- Minimum for animals : 120 ms
- Minimum for vehicles : 140 ms

# Face Zapping Task

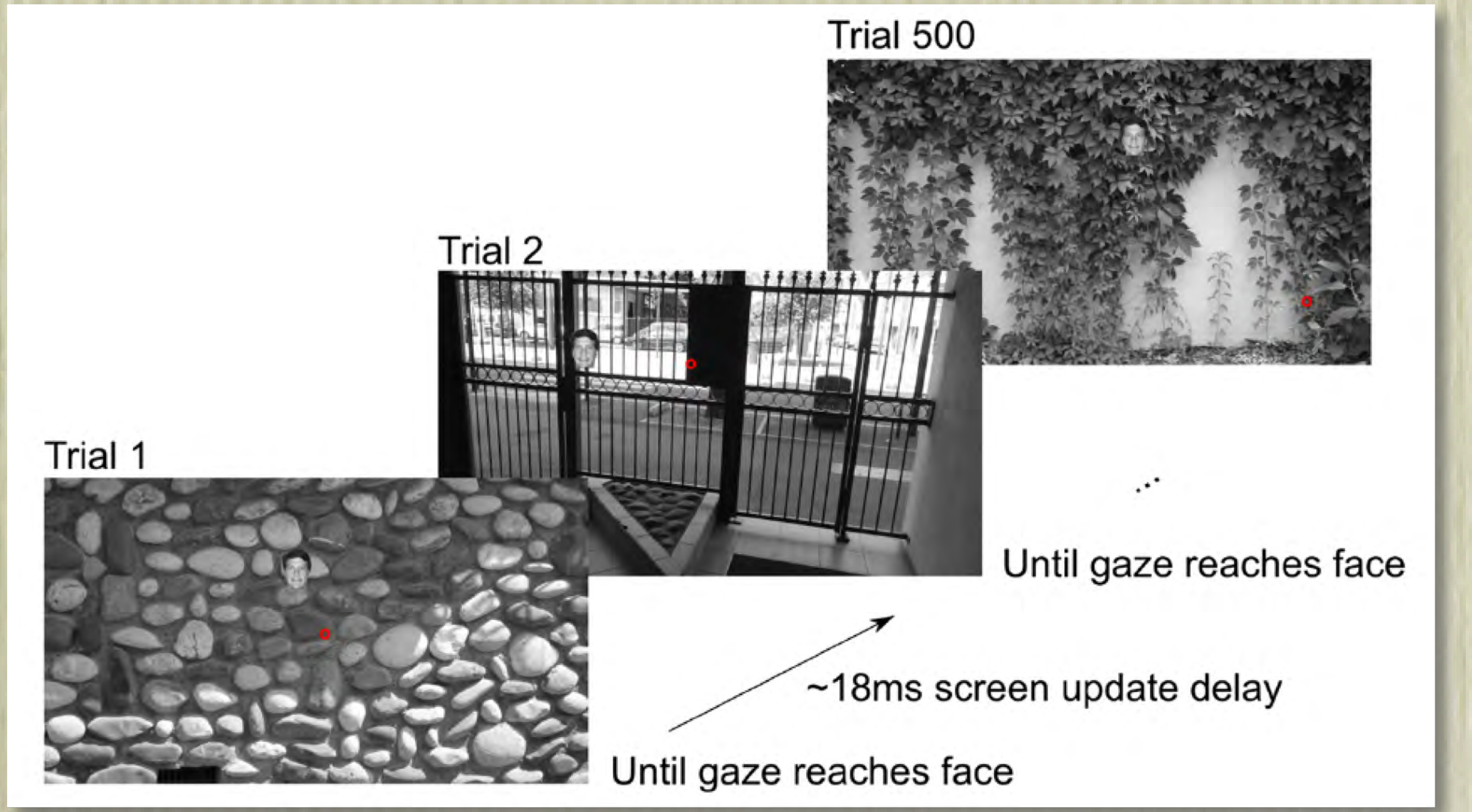

SCIENTIFIC REPORTS

OPEN

**Zapping 500 faces in less than 100 seconds: Evidence for extremely fast and sustained continuous visual search**

Jacob G. Martin[1,2], Charles E. Davis[1], Maximilian Riesenhuber[2] & Simon J. Thorpe[1]

Received: 17 November 2017
Accepted: 25 July 2018
Published online: 20 August 2018

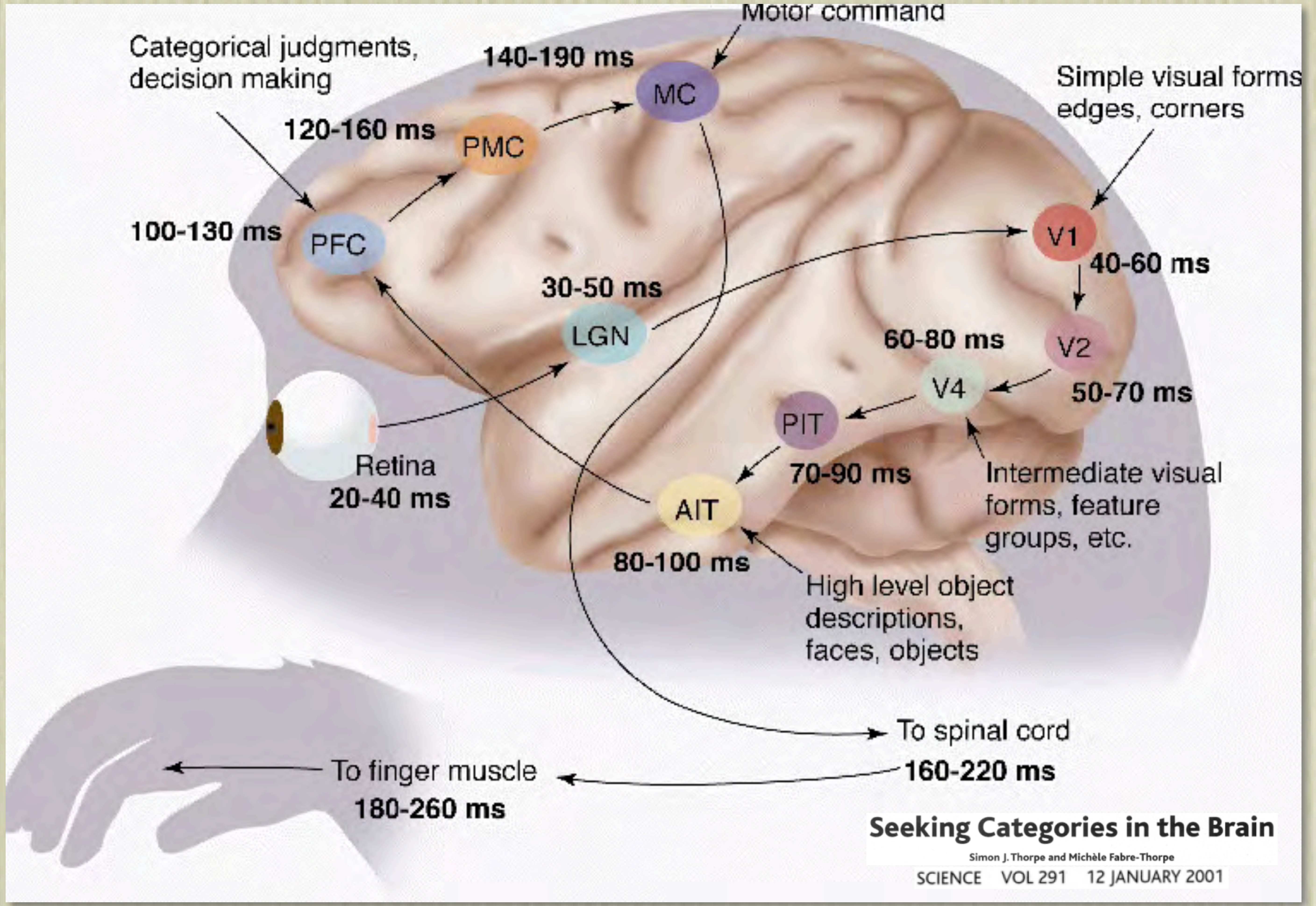



- Up to 6 saccades a second!

# Face Zapping Task

# Ultra-Rapid Visual Processing

- Animal ERP difference at 150 ms in humans
- Saccades to animals in 120 ms
- Saccades to faces in 100 ms
- 5 saccades a second

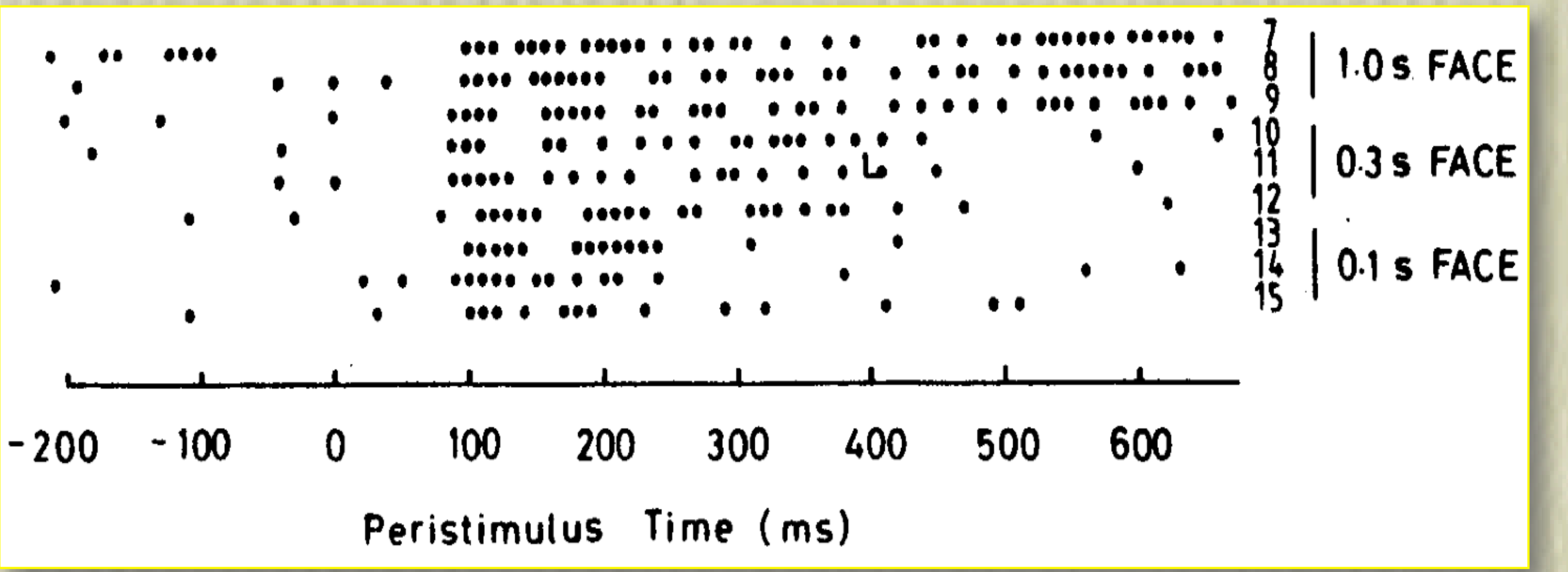


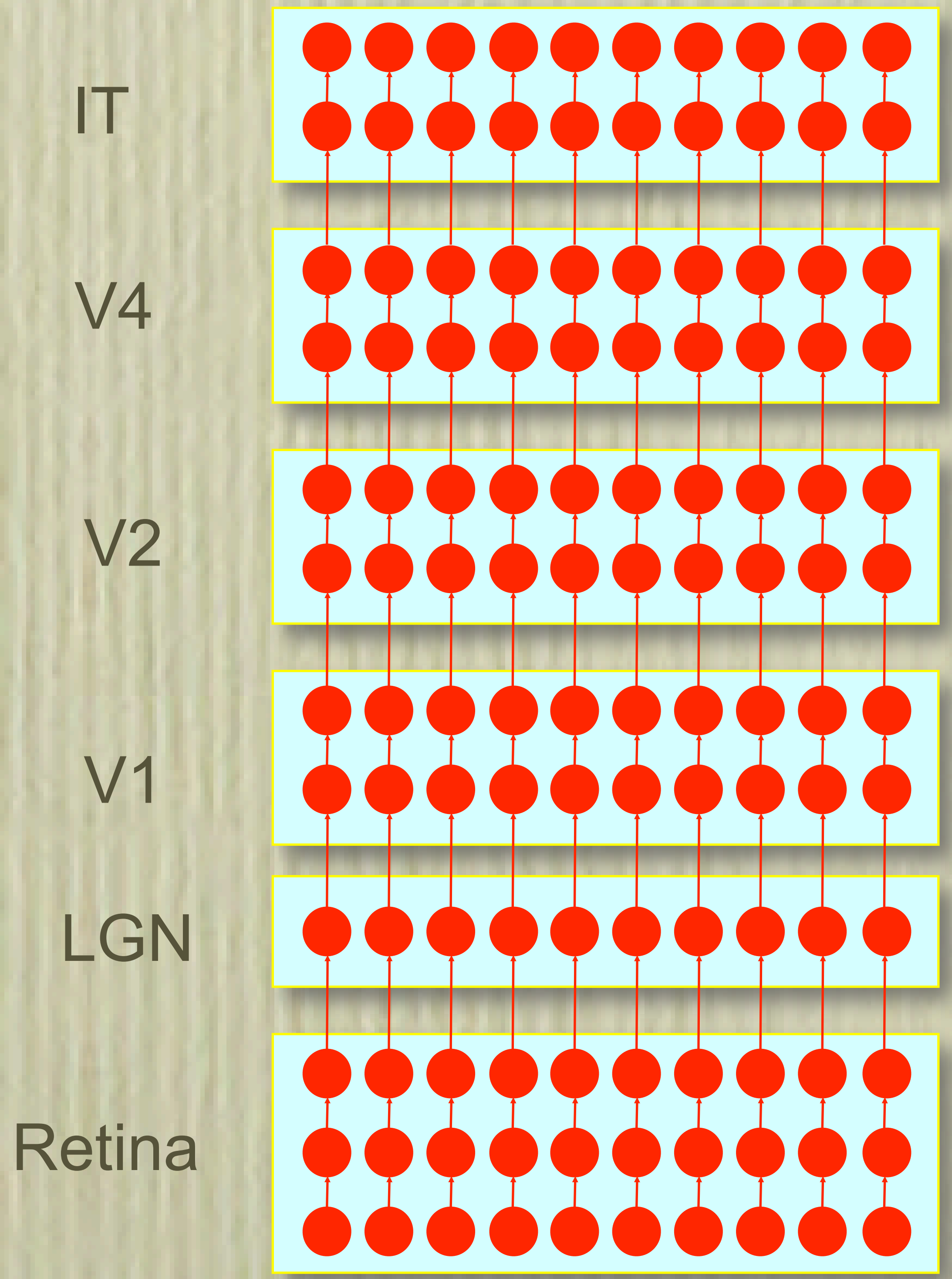

- Feedforward processing
- Only a few milliseconds per processing step
- Processing without context based help

- Could an artificial system do the same thing?

# Temporal Contraints - 1989

Face selectivity at 100 ms (Perrett Rolls & Caan, 1982)




## Argument

- Roughly 10 layers
- 10 ms per layer
- Firing rates 0-100 Hz

## Therefore

- Essentially feedforward (?!)
- One spike per neuron (?!)
- Rate coding impossible

BIOLOGICAL CONSTRAINTS ON CONNECTIONIST MODELLING

Simon J. Thorpe and Michel Imbert,

**A.** The visual system is arranged as a massively parallel multilayer feedforward net with at least 10 processing layers.

**B.** Considerable visual analysis is possible with a single forward pass through the network.

**C.** In many situations, each unit can only emit 1 spike before the units in the next layer have to respond.

**D.** Firing rate per se cannot be used during visual processing to code analog values with any real precision.

**E.** Coding of analog values could however be achieved by making use of the arrival times of spikes from different sources - the earliest arriving signals could be given priority.

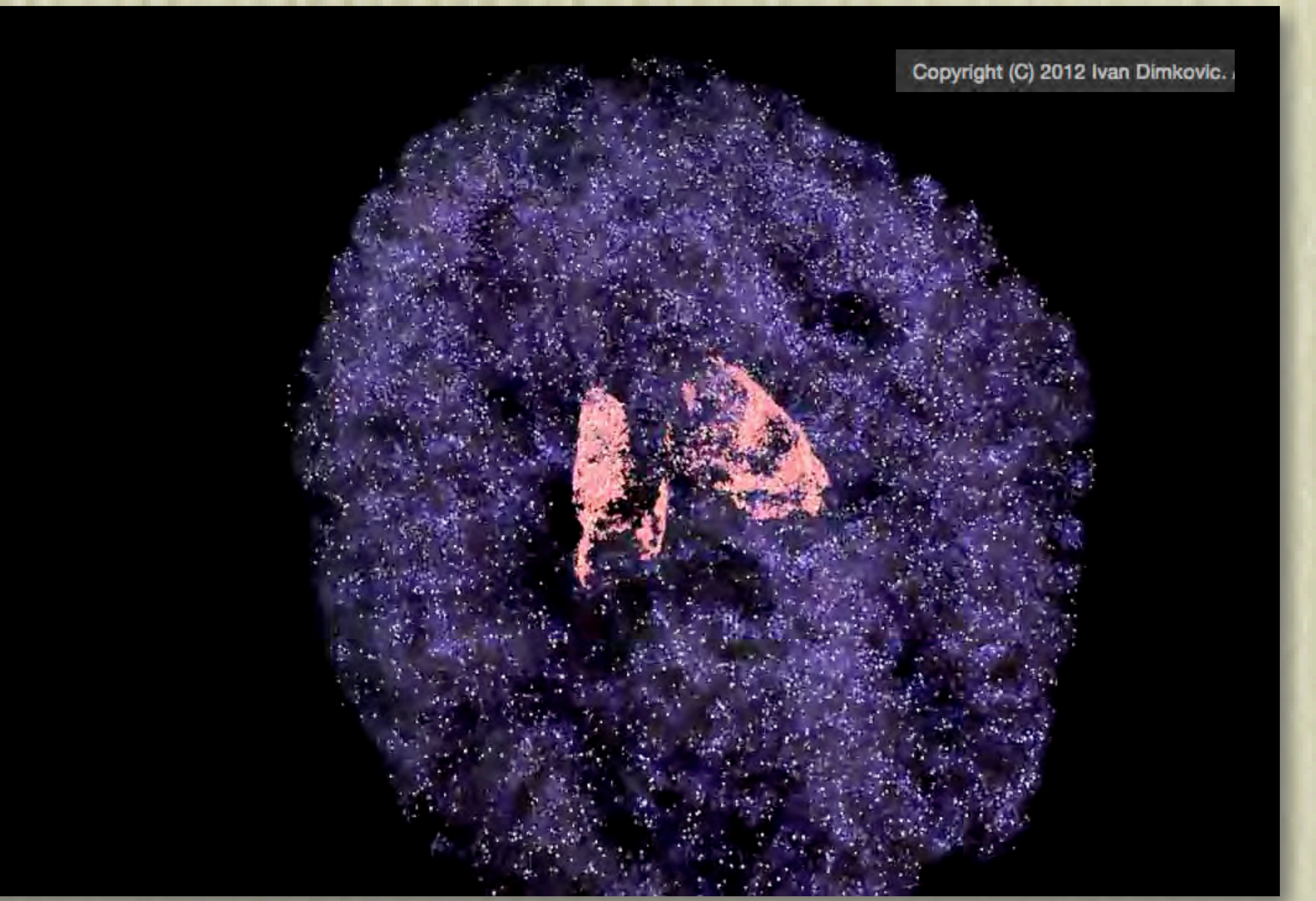
**F.** Sophisticated dendritic processing could mean that each unit could be doing more than simply calculating the sum of all the input activations - logical "and", "and not" functions could well make the system highly non-linear.

**G.** Although feedback pathways between different layers are present, there may not be time to use them during normal visual processing. They could however play a role in the effects of context, imagery, attention, resolution of ambiguous stimuli and learning.

**H.** The use of iterative loops is kept to an absolute minimum and perhaps even eliminated by the use of massive parallelism.

# Biological vs Computer Hardware

## Brain




- 86 billion neurons
- 16 billion in the cortex
- 4 billion in the visual system
- 1 KHz
- 1-2 m/s conduction velocity
- 20 watts

## Computer

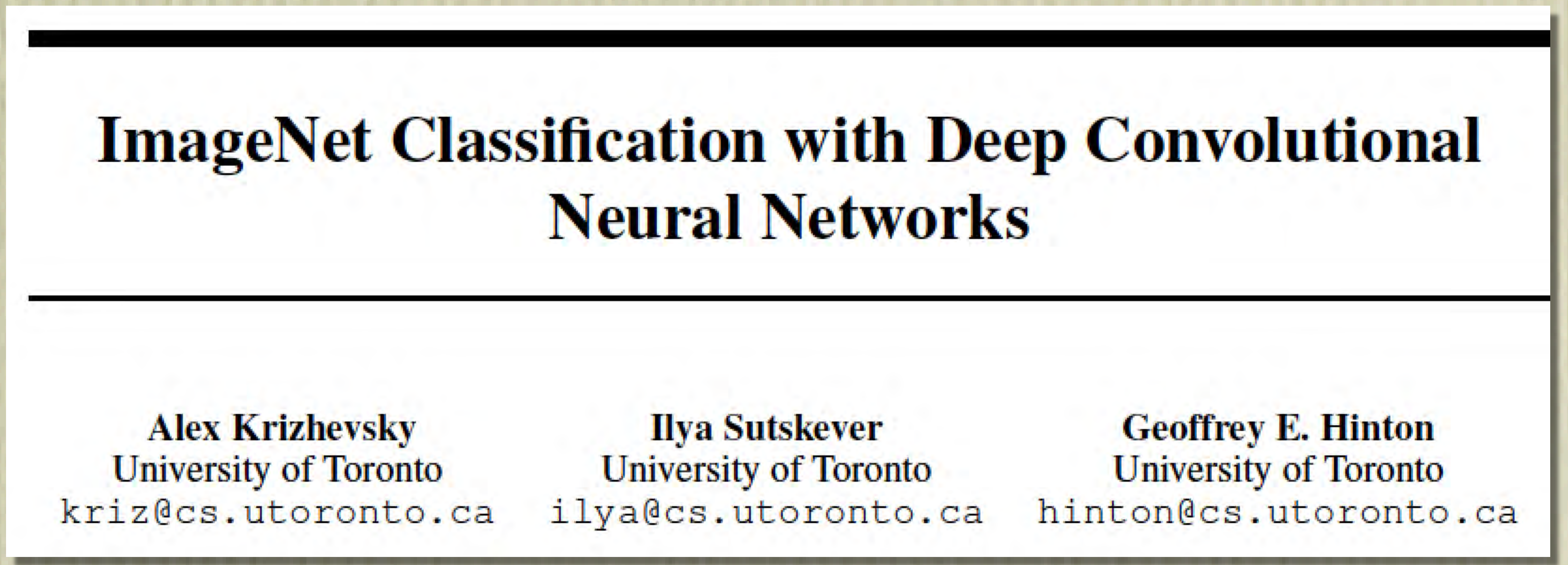
- Nvidia GTX Titan X
  - 11 TeraFlops!
  - 3854 cores
  - 12 billion transistors
  - 480 Gbytes/sec Memory bus
  - 250 watts
  - $1200


Is that enough to reproduce human performance?

# The ImageNet Challenge

- 10,000,000 training images
- 10,000+ labels
- Systems tested on new images, with 1000 possible labels
- ECCV 2012 Firenze
  - The state of the art was beaten by a “simple” feedforward convolutional neural network trained with Back-Propagation

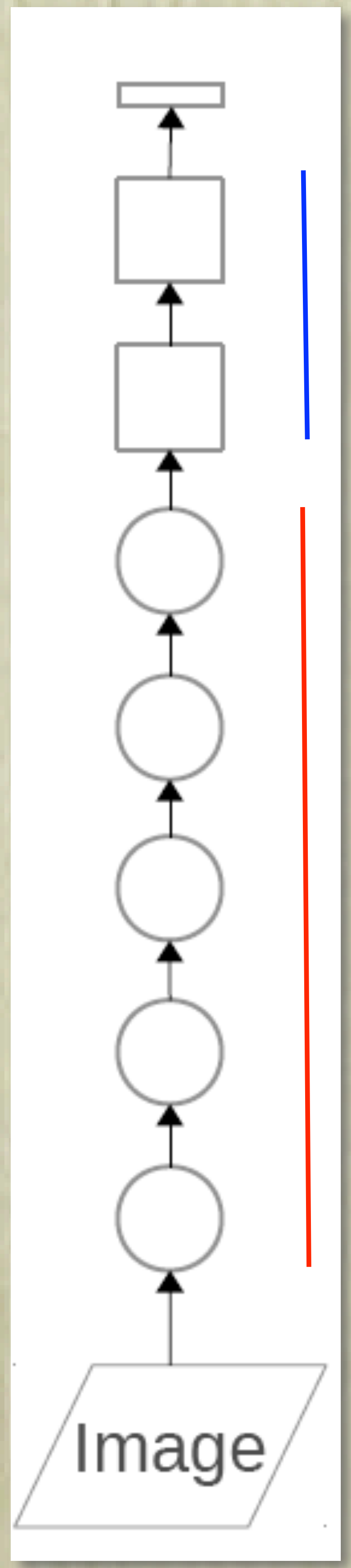
# AlexNet

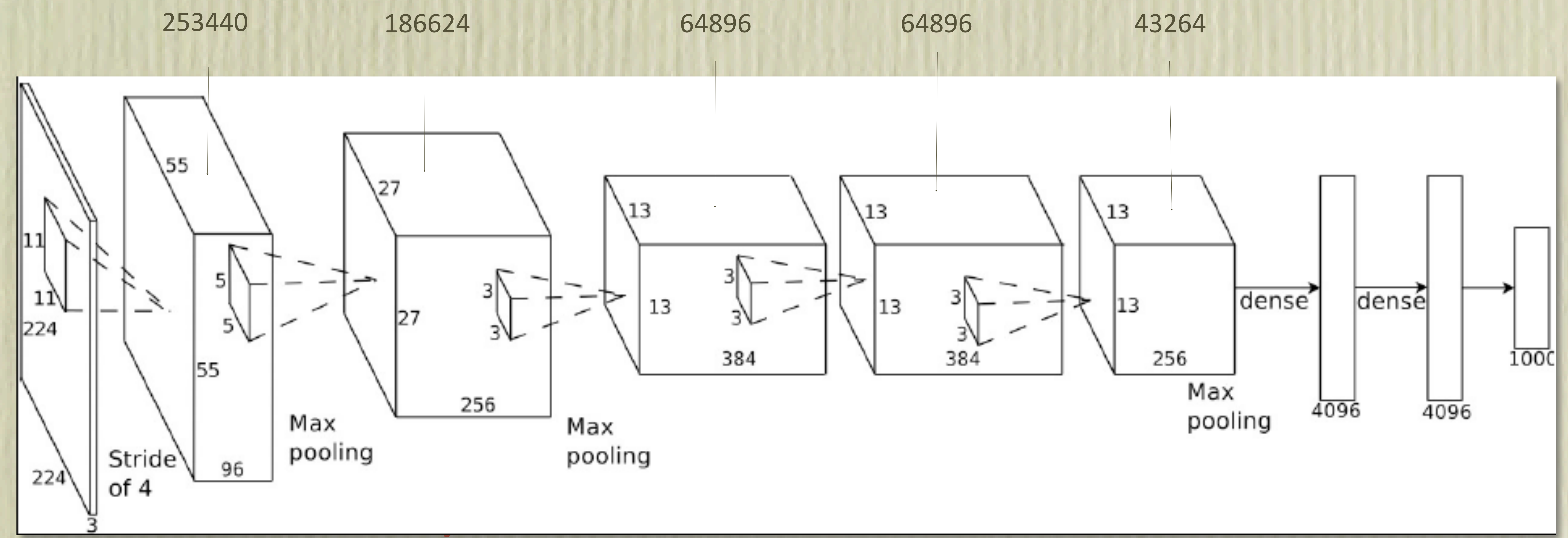

ImageNet Classification with Deep Convolutional Neural Networks

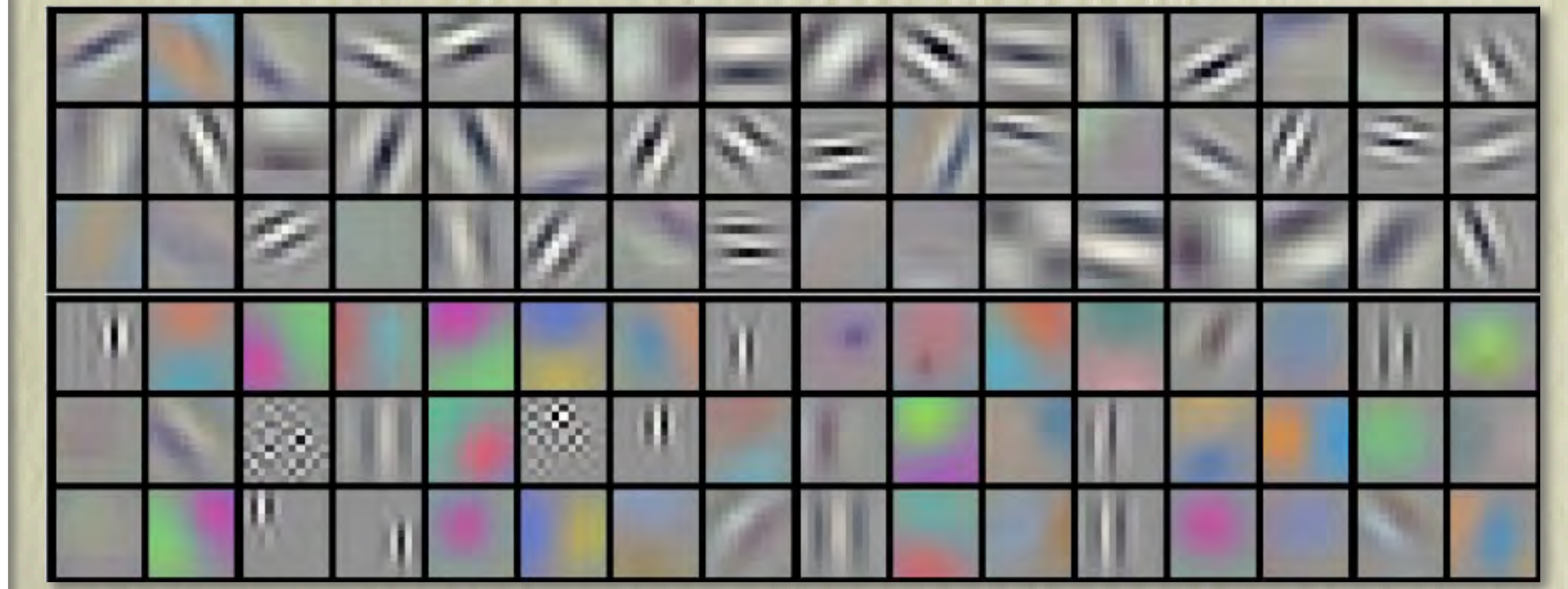
Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

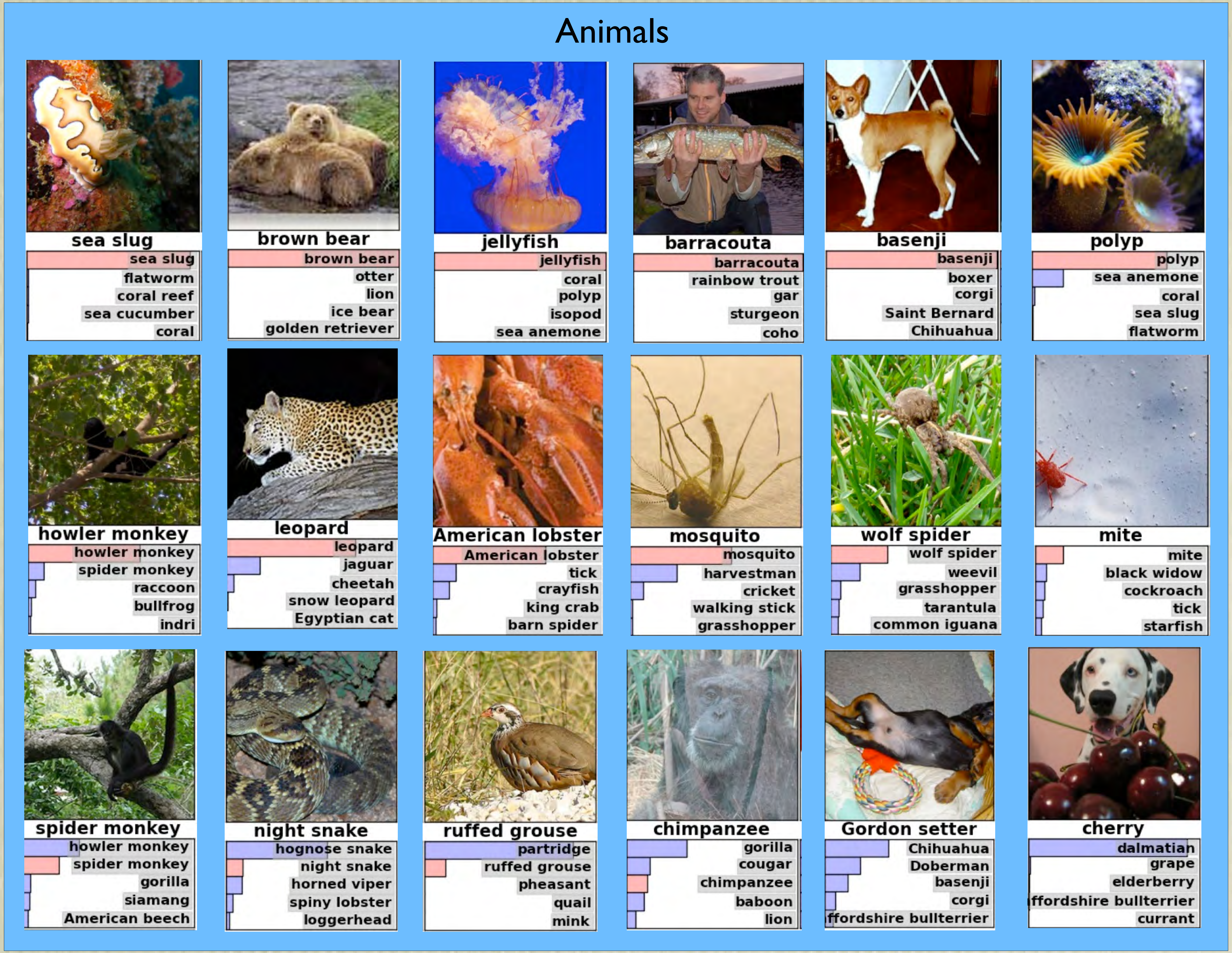


- 650,000 "neurons"
- 60,000,000 parameters
- 630,000,000 "synapses"

# AlexNet



Animals
sea slug
sea slug
flatworm
coral reef
sea cucumber
coral
brown bear
brown bear
otter
lion
ice bear
golden retriever
jellyfish
jellyfish
coral
polyp
isopod
sea anemone
barracouta
barracouta
rainbow trout
gar
sturgeon
coho
basenji
basenji
boxer
corgi
Saint Bernard
Chihuahua
polyp
polyp
sea anemone
coral
sea slug
flatworm
howler monkey
howler monkey
spider monkey
raccoon
bullfrog
indri
leopard
leopard
jaguar
cheetah
snow leopard
Egyptian cat
American lobster
American lobster
tick
crayfish
king crab
barn spider
mosquito
mosquito
harvestman
cricket
walking stick
grasshopper
wolf spider
wolf spider
weevil
grasshopper
tarantula
common iguana
mite
mite
black widow
cockroach
tick
starfish
spider monkey
howler monkey
spider monkey
gorilla
siamang
American beech
night snake
hognose snake
night snake
horned viper
spiny lobster
loggerhead
ruffed grouse
partridge
ruffed grouse
pheasant
quail
mink
chimpanzee
gorilla
cougar
chimpanzee
baboon
lion
Gordon setter
Chihuahua
Doberman
basenji
corgi
ffordshire bullterrier
cherry
dalmatian
grape
elderberry
ffordshire bullterrier
currant

# And then….

- Geoff Hinton and his two students launched a start-up (DNNresearch)
- bought by Google…

WIRED GEAR SCIENCE ENTERTAINMENT BUSINESS SECURITY DESIGN OPINION VIDEO

Meet the Man Google Hired to Make AI a Reality

BY DANIELA HERNANDEZ 01.16.14 6:30 AM

Geoffrey Hinton (right) Alex Krizhevsky, and Ilya Sutskever (left) will do machine learning work at Google. *Photo: U of T*

- Yann LeCun, a pioneer of feed-forward convolutional networks since the end of the 1980s
- Hired by Facebook…

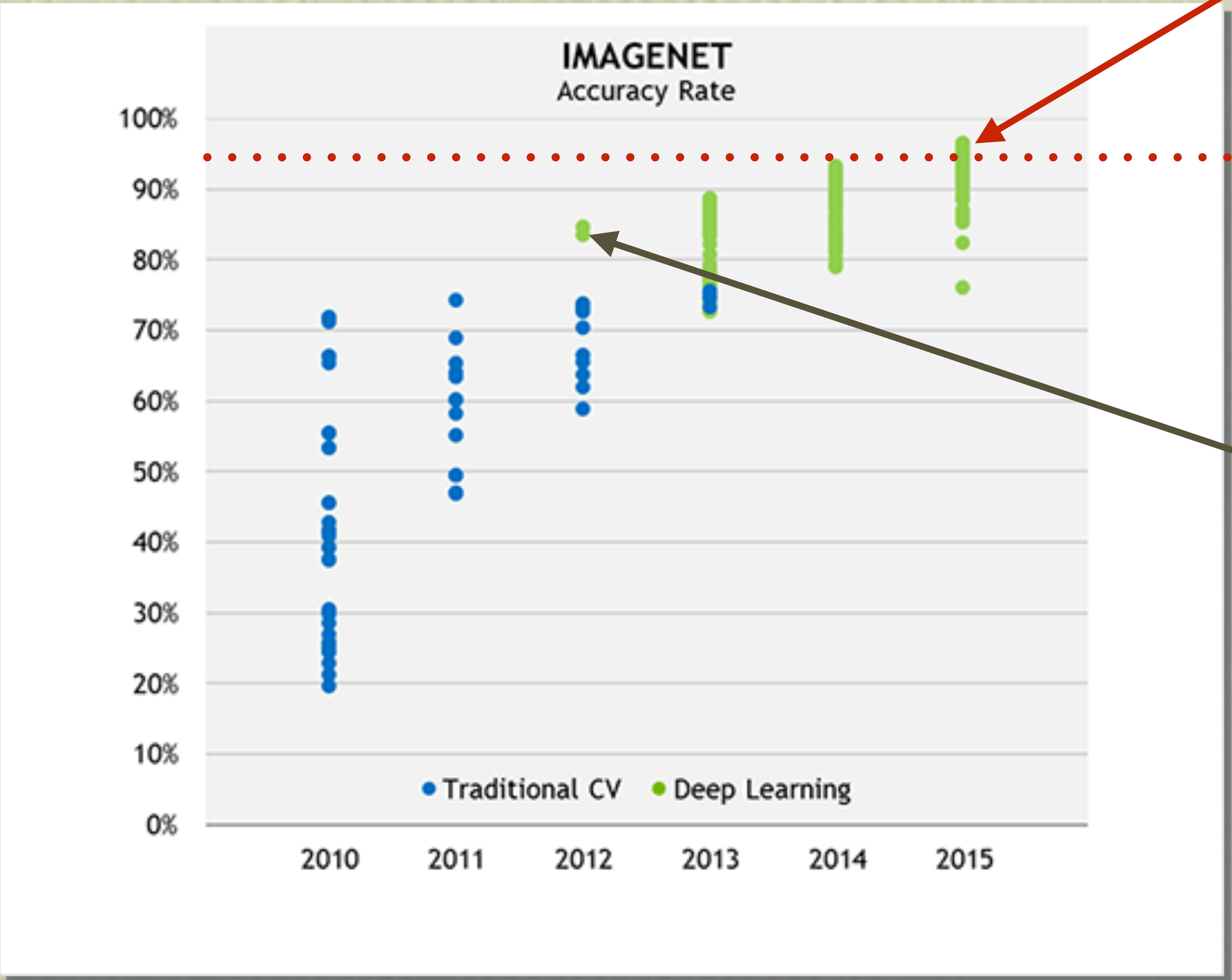
GIGAOM EVENTS RESEARCH JOBS

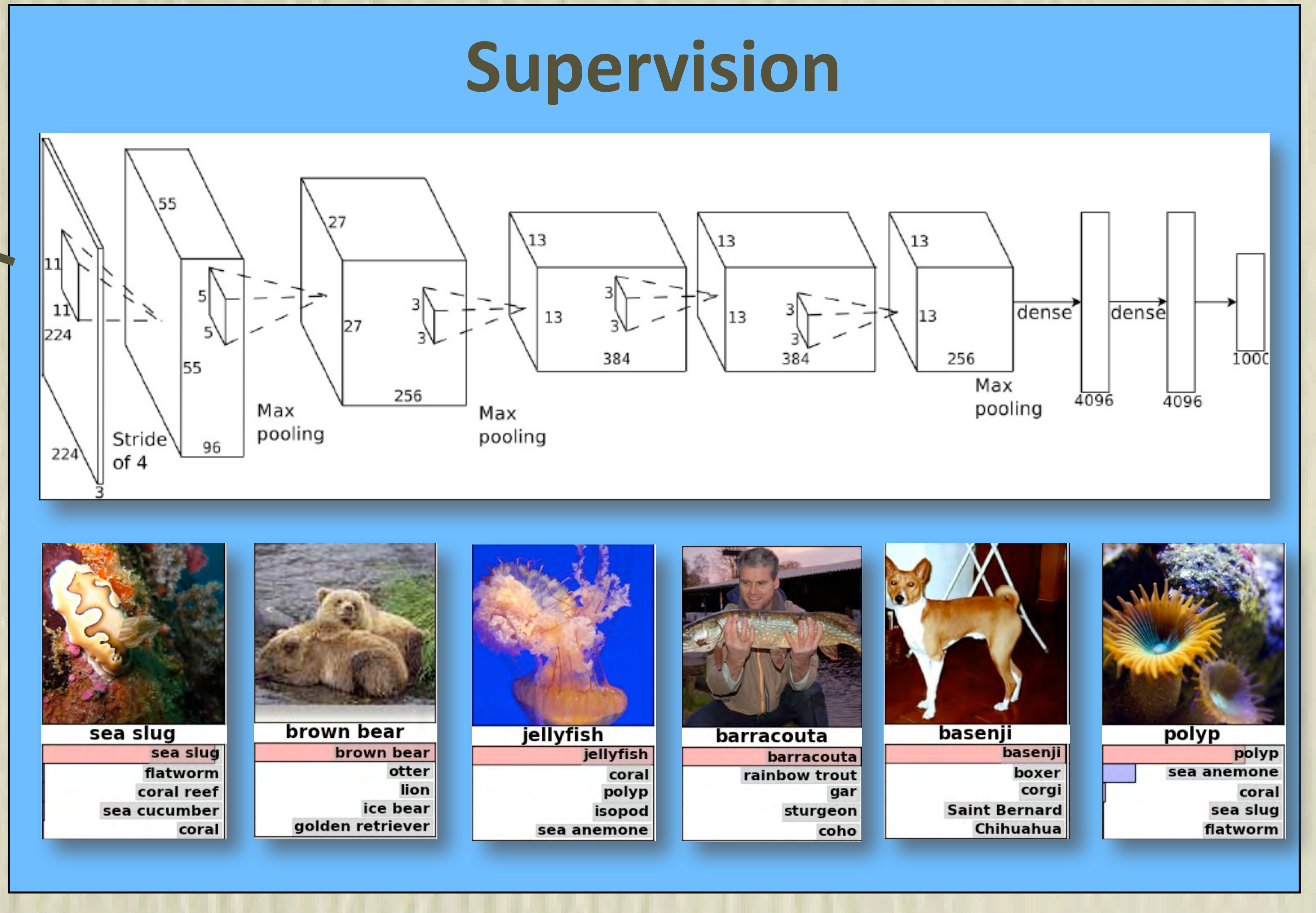
Facebook hires NYU deep learning expert to run its new AI lab

by Derrick Harris DEC. 9, 2013 - 10:38 AM PDT

SUMMARY: *Facebook has hired deep learning expert Yann Lecun from New York University to head up its new artificial intelligence lab. It's part of a bigger push along with — and against — companies like Google and Microsoft to advance research while improving their platforms.*

# ImageNet Performance

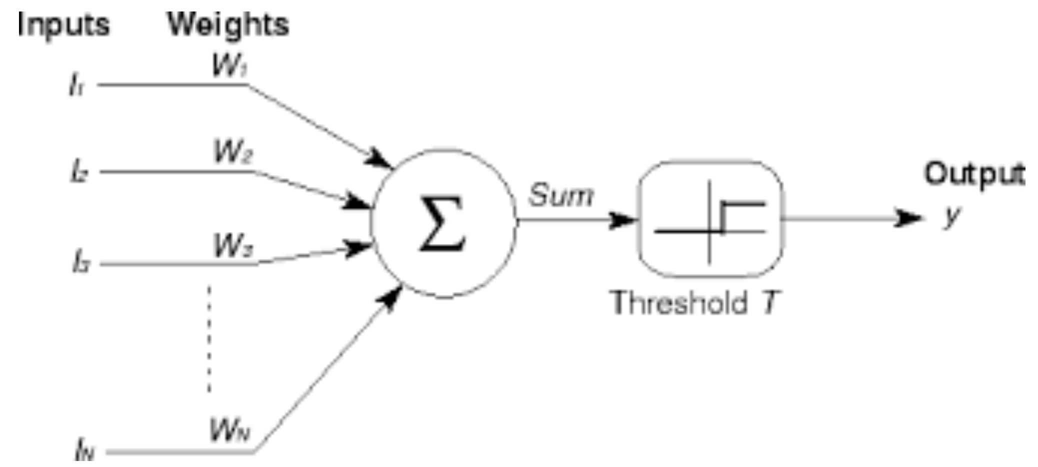


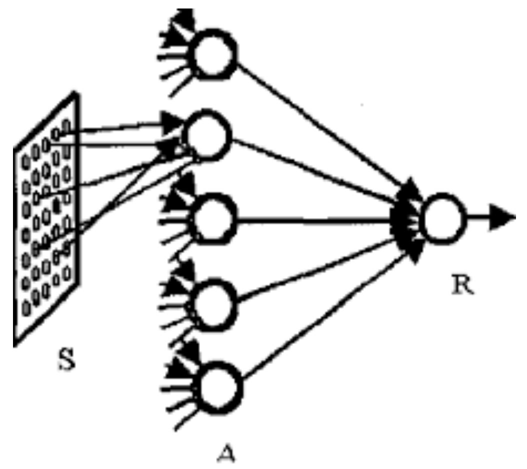



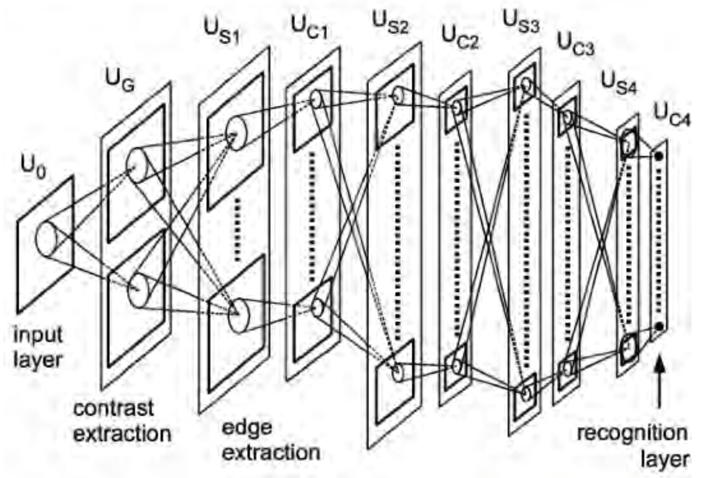

- Feedforward architectures really can be very powerful
- But has vision been “solved”?

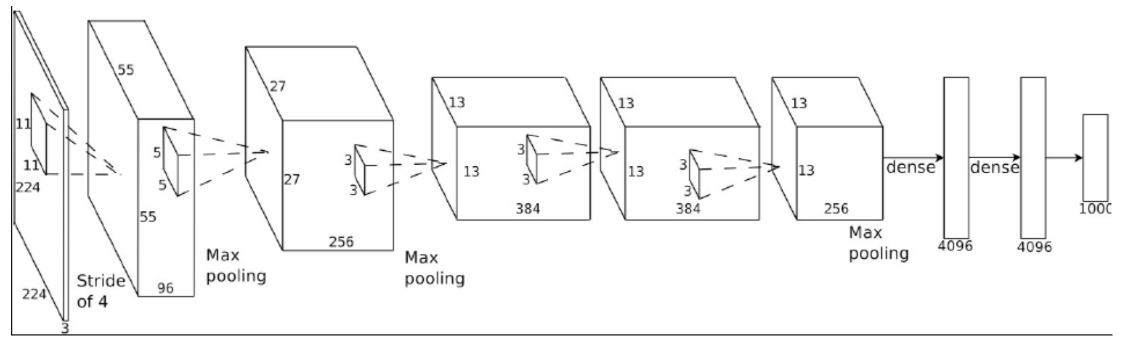
# History of Neural Networks

- 1940
- Warren McCulloch & Walter Pitts (1943)
- 1950
- Frank Rosenblatt's Perceptron (1958)
- 1960
- Rosenblatt's Multi Layer Perceptron (1962)
- 1970
- Kunihiko Fukushima's Neocognitron (1980)
- 1980
- Rumelhart, Hinton & Williams - Backpropagation (1986)
- 1990
- 2000
- 2010
- Alex Krizhevsky, Ilya Sutskever & Geoff Hinton (2012)
- Now



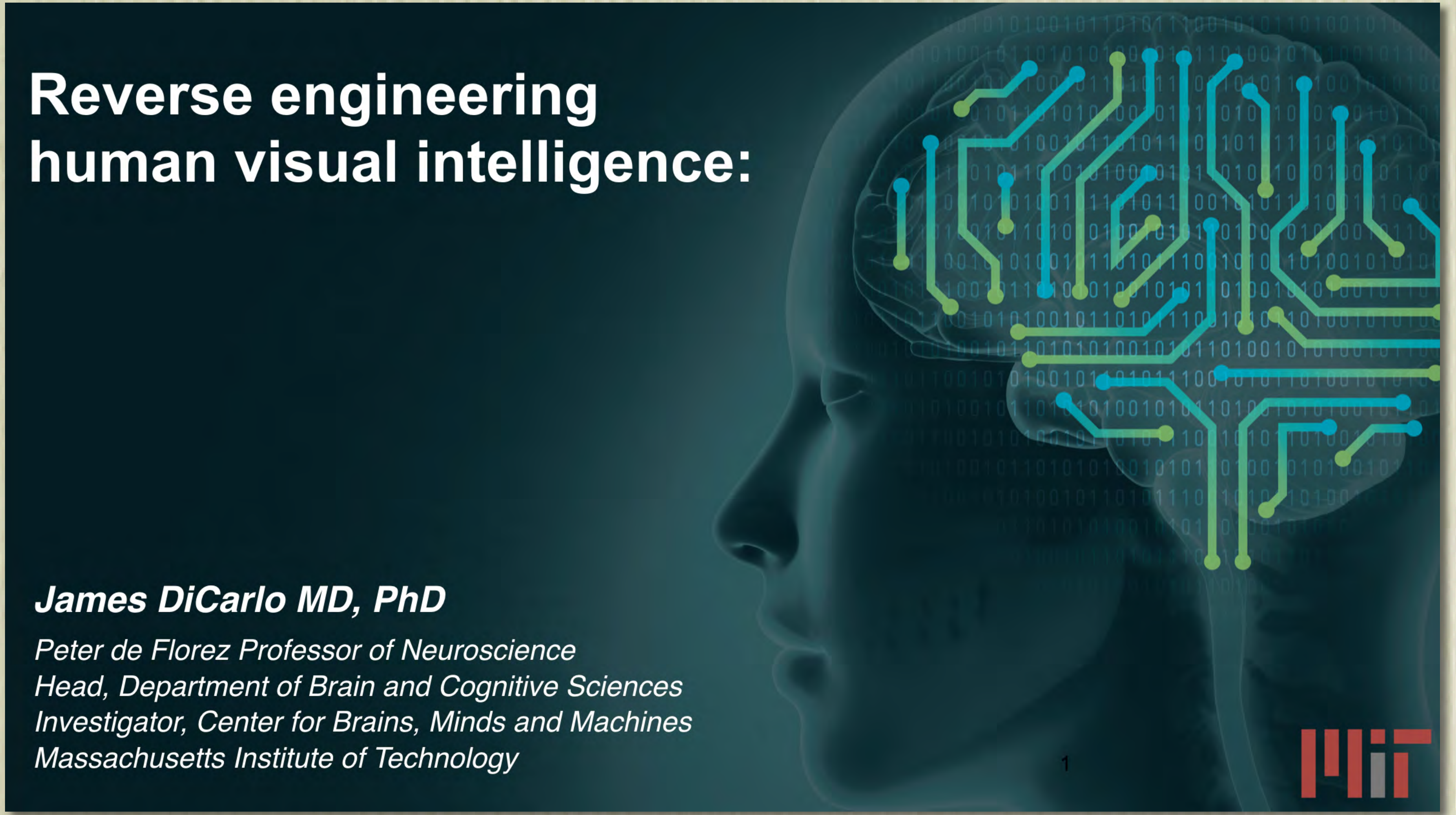




- Old ideas!
- Why the change?
  - Development of GPU gaming hardware
  - Massive quantities of labelled training data

# Comparing Neurons and Deep Networks

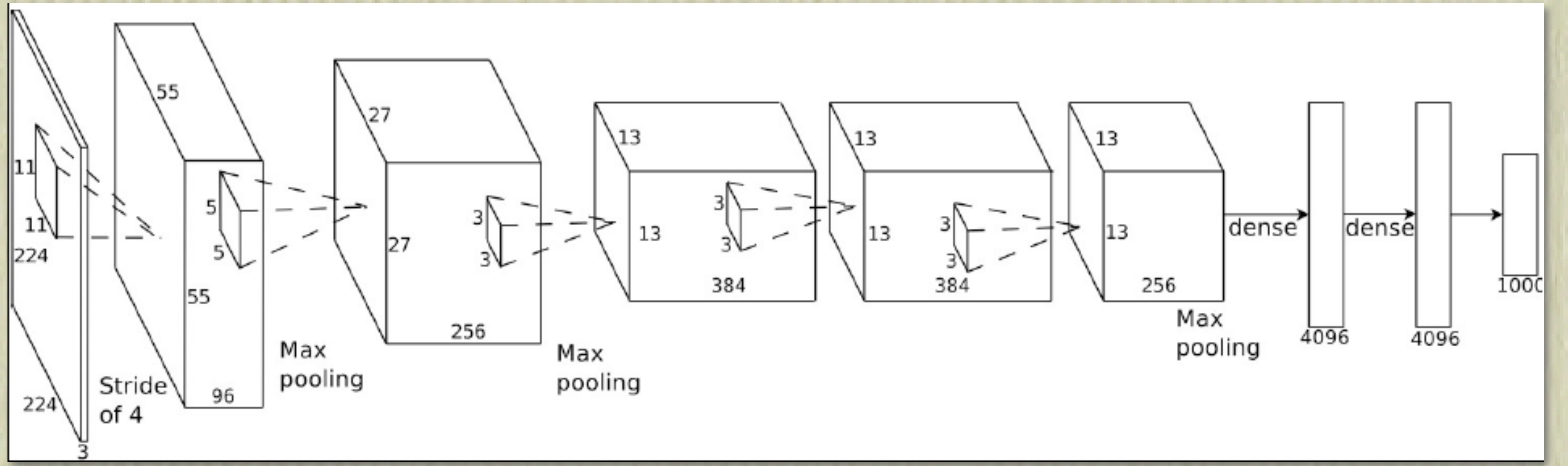
- Jim DiCarlo

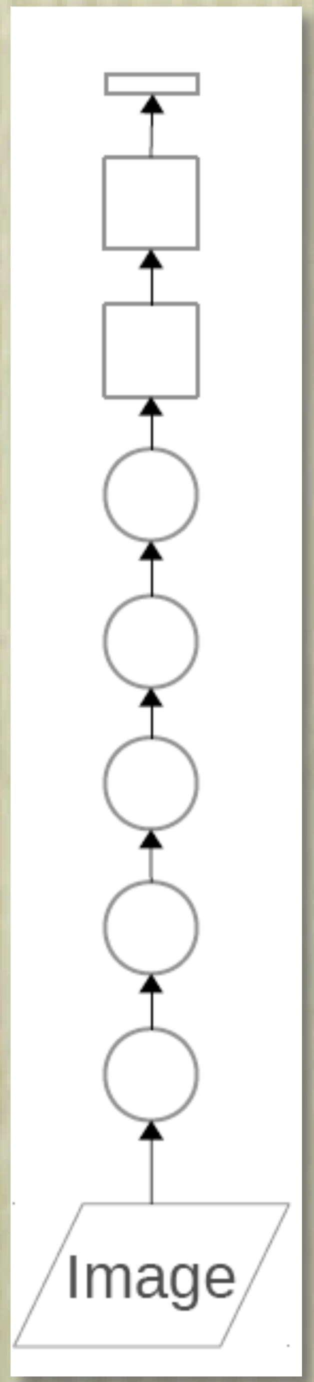


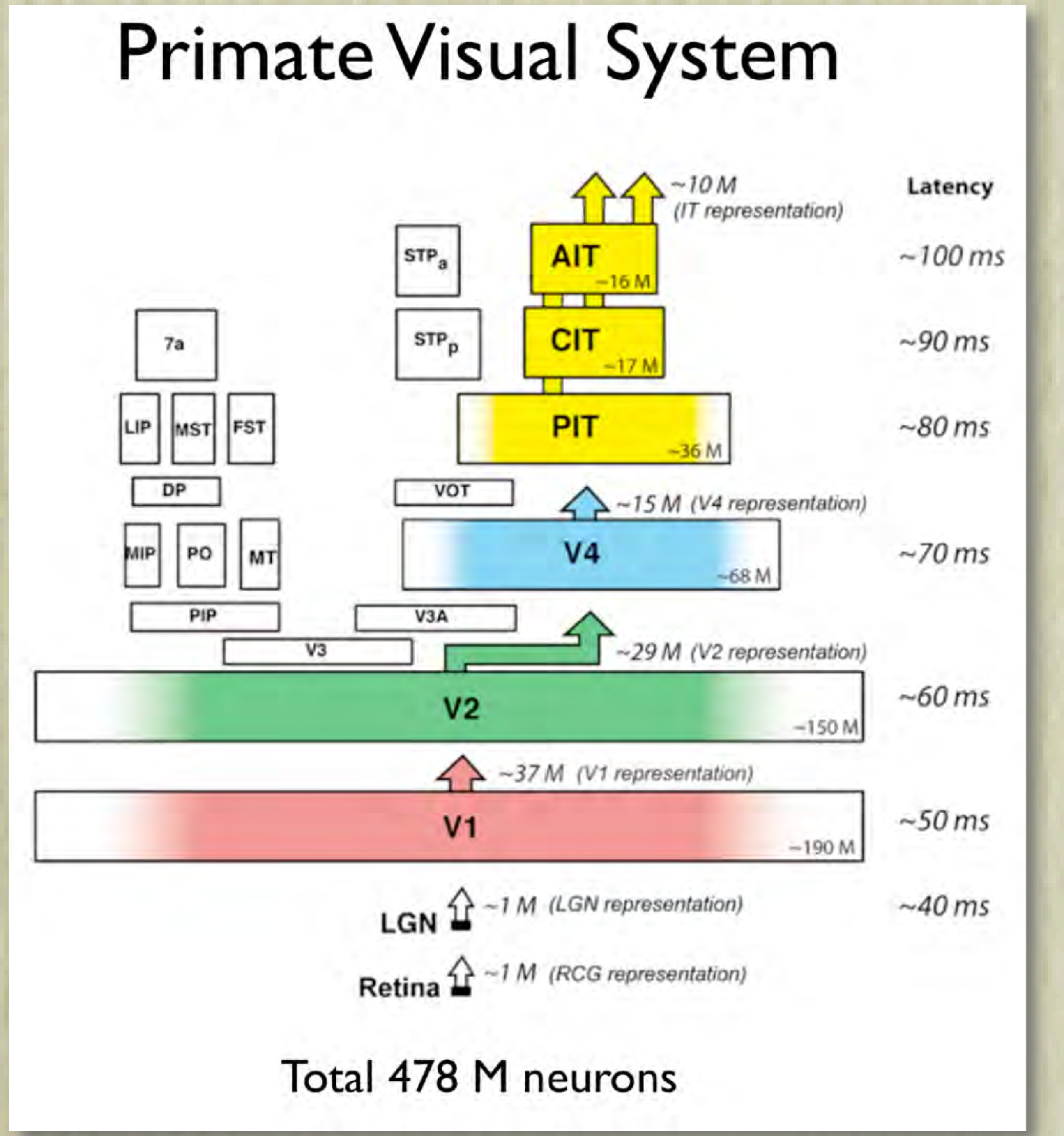
*Dan Yamins, Hong, Solomon, Seibert and Jim DiCarlo* ***NIPS (2013), PNAS (2014)***

**A specific deep ANN** (evolved to try to solve core recognition)

# High Level Vision with neurally plausible architectures!



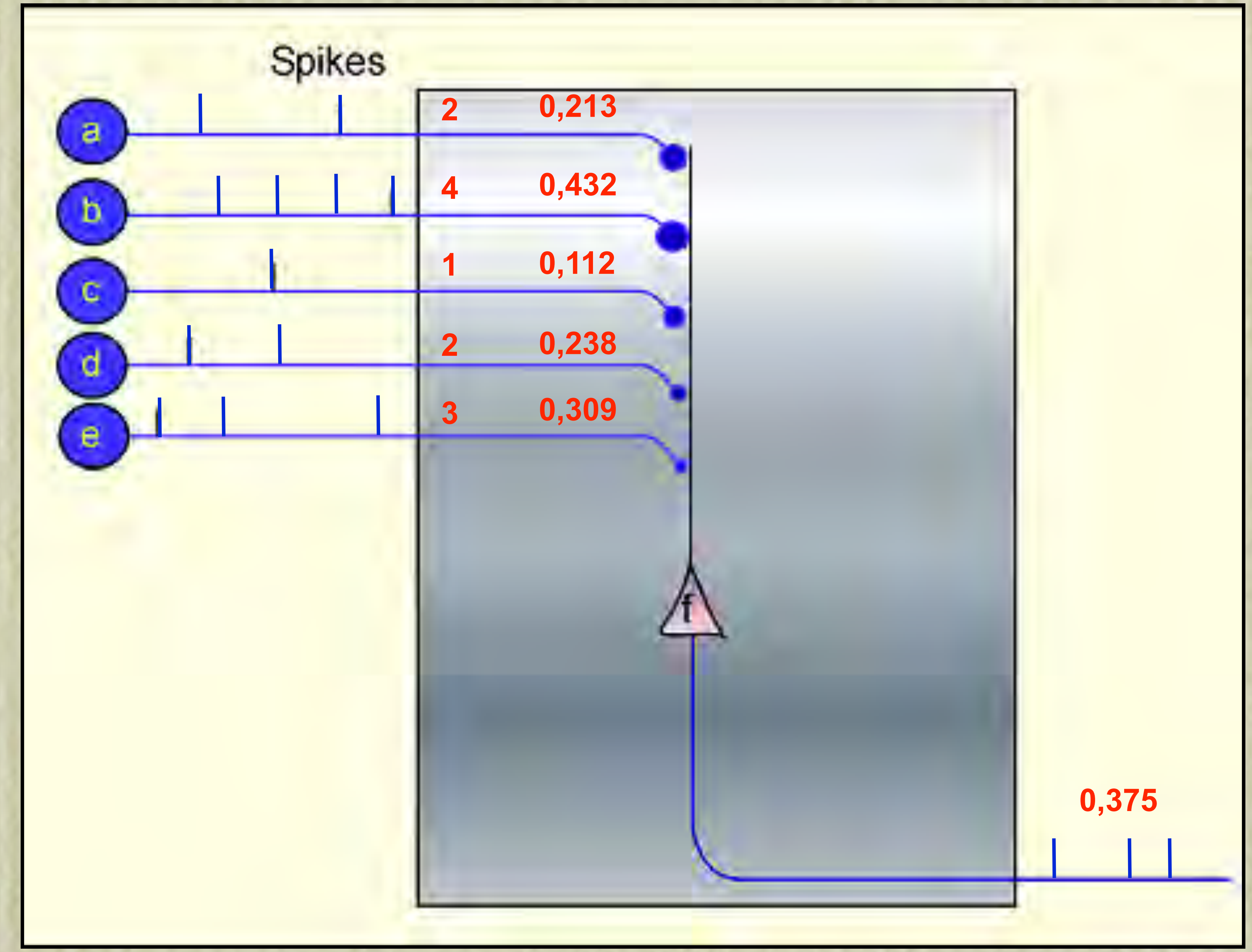




## What's missing?

- No recurrent connections
- No horizontal connections
- No dendrites
- No synaptic dynamics
- No Memory
- No Attention
- No Binding
- No Oscillations
- No Spikes!

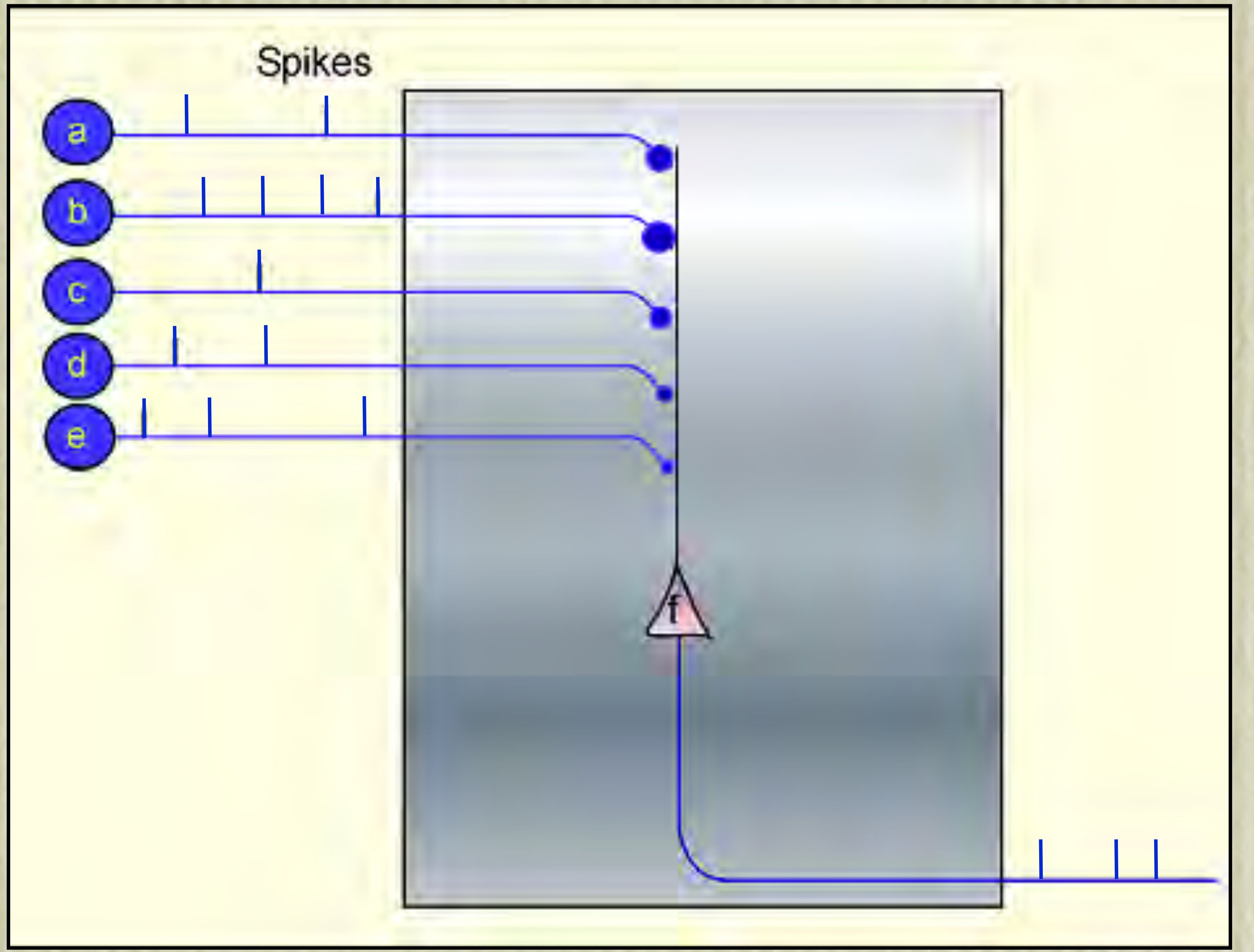
# Coding by Neurons : The Classic View




- Spikes don't really matter
- Neurons send floating point numbers
- The floating point numbers are transformed into spikes trains using a Poisson process
- God plays dice with spike generation!

- View reinforced by the success of Deep Learning and Convolutional Neural Networks

# Temporal Coding Option

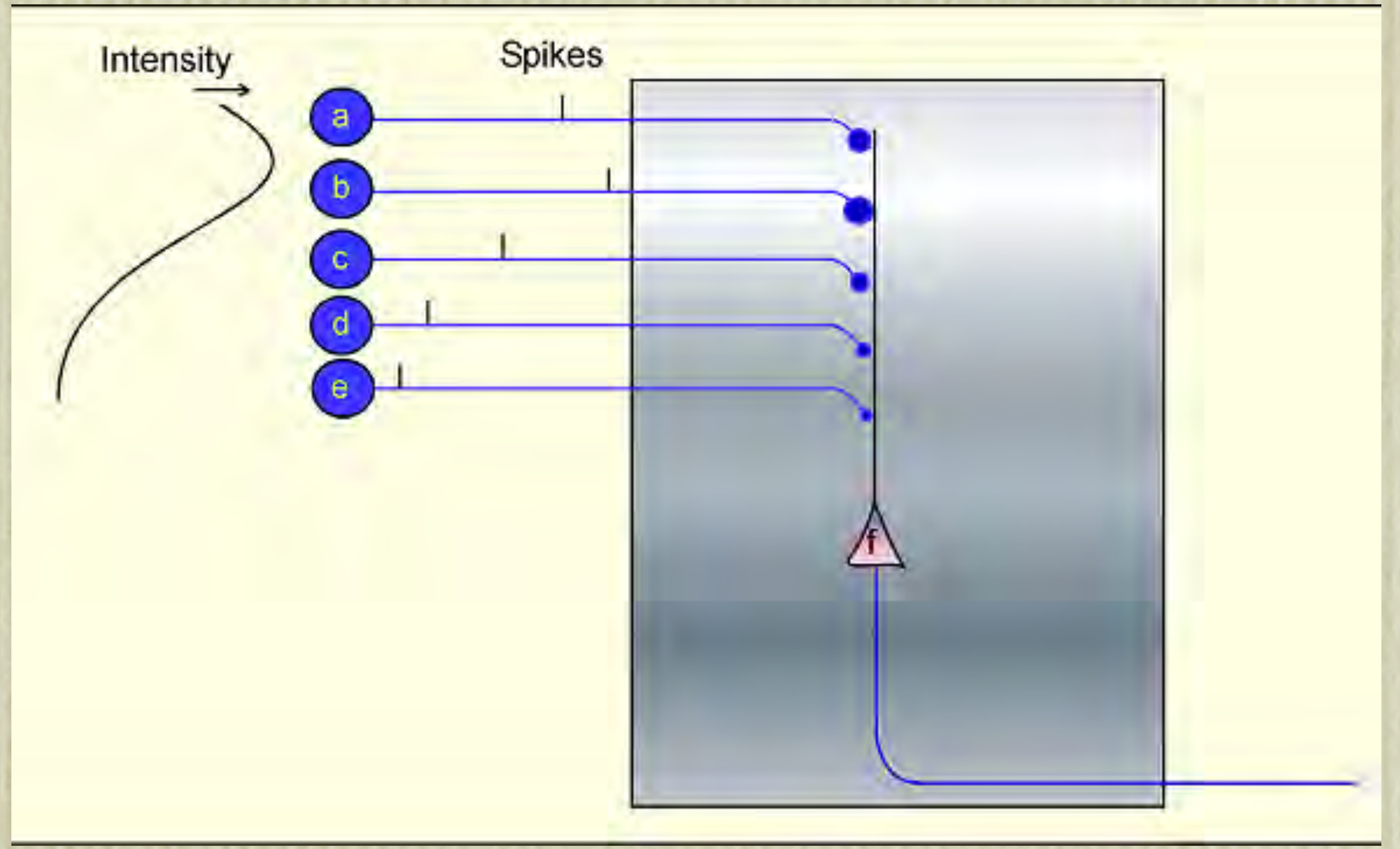



- Spikes do really matter

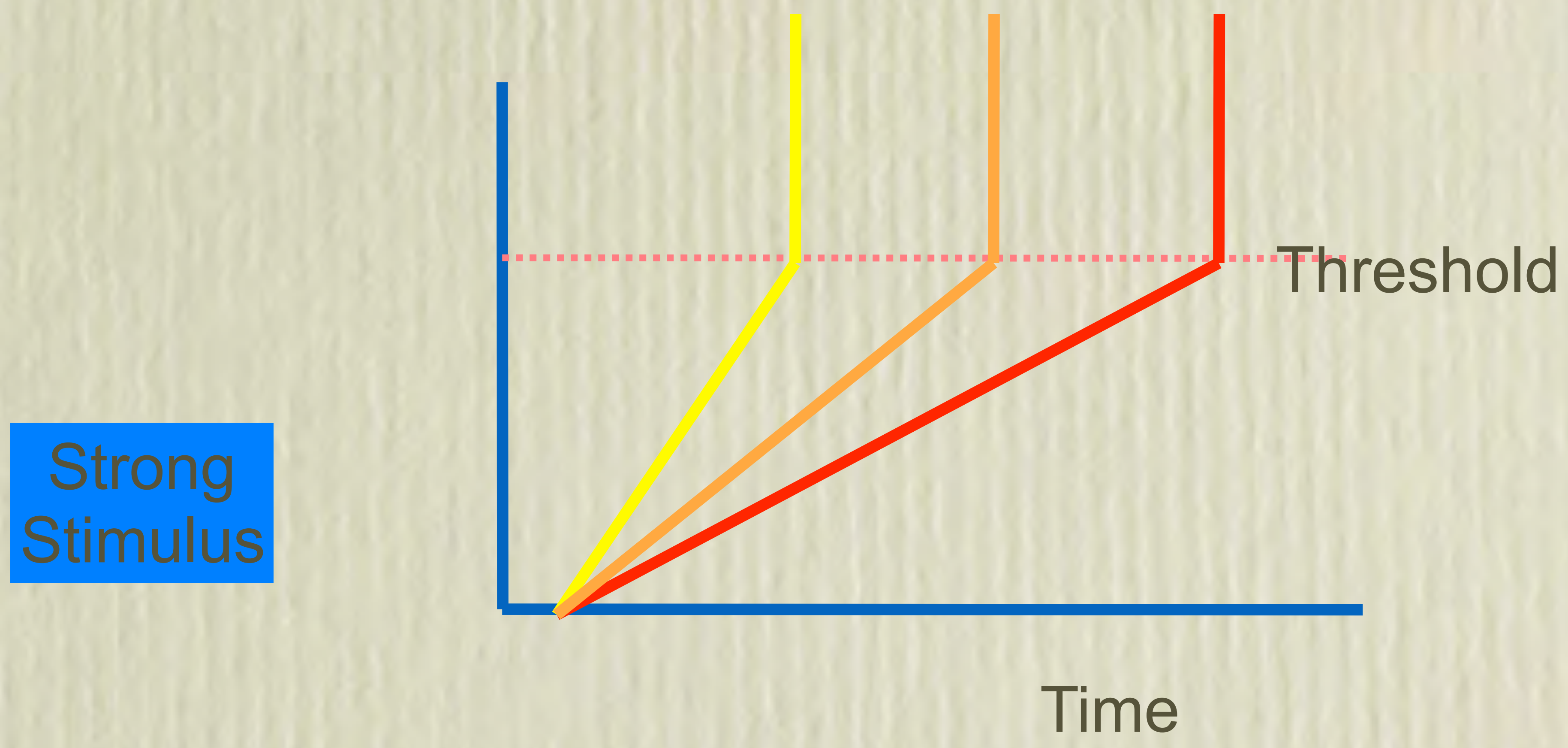
- The temporal patterning of spikes across neurons is critical for computation
  - Synchrony
  - Repeating patterns
  - etc

- The apparent noise in spiking is unexplained variation

# Order based coding

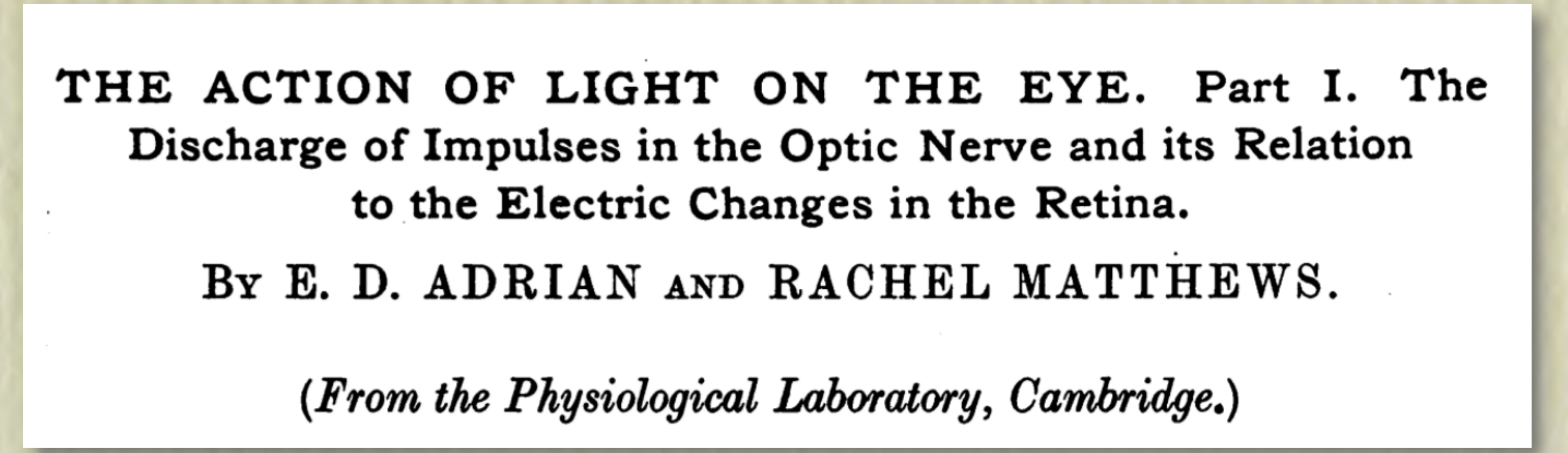



- Ordering of spikes is critical
- The most activated neurons fire first
- Temporal coding is used even for stimuli that are not temporally structured
- Computation theoretically possible even when each neuron emits one spike



# Sensory Coding with Spikes

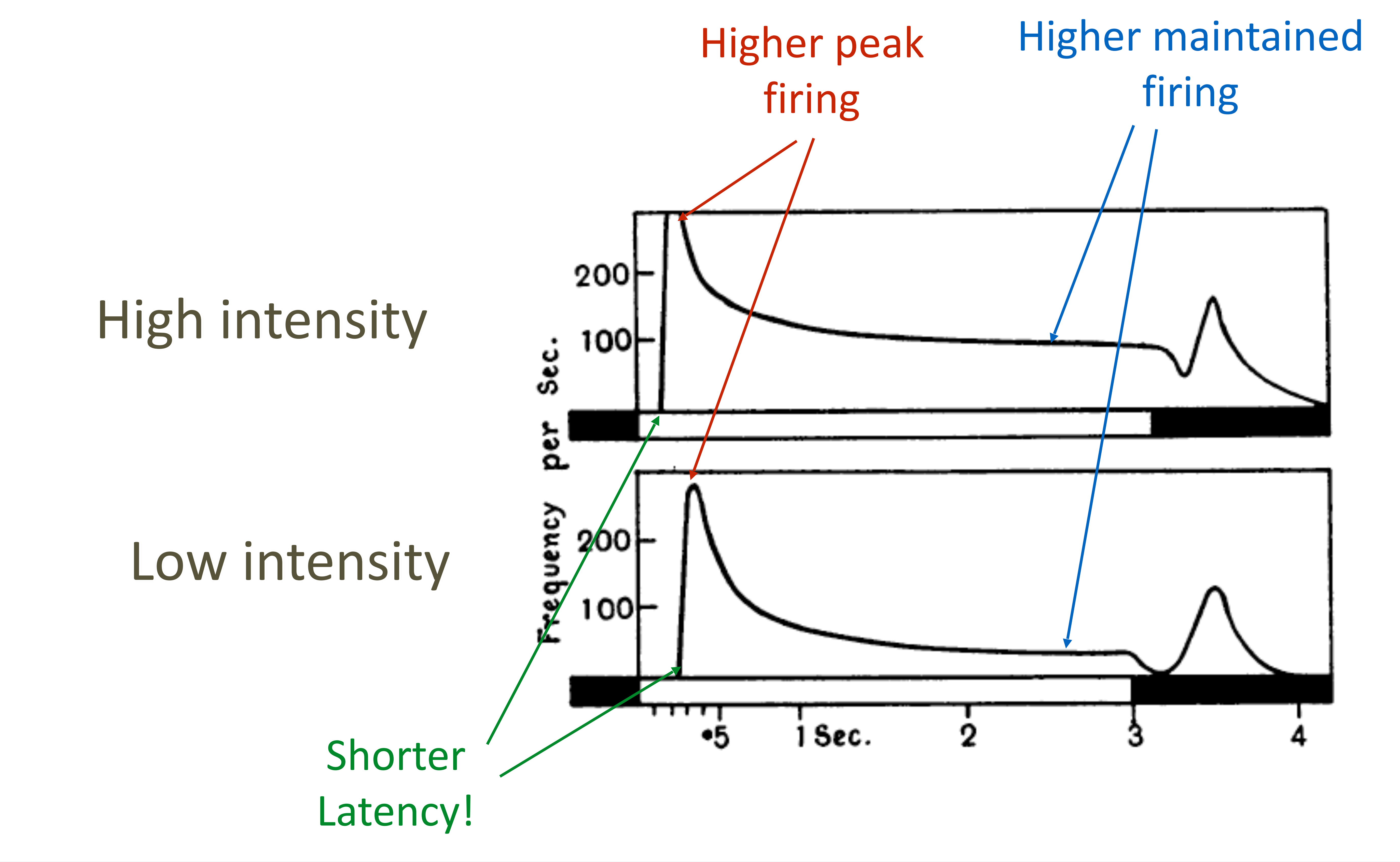
- Edgar Douglas Adrian (1920s)
  - First recordings from sensory fibres

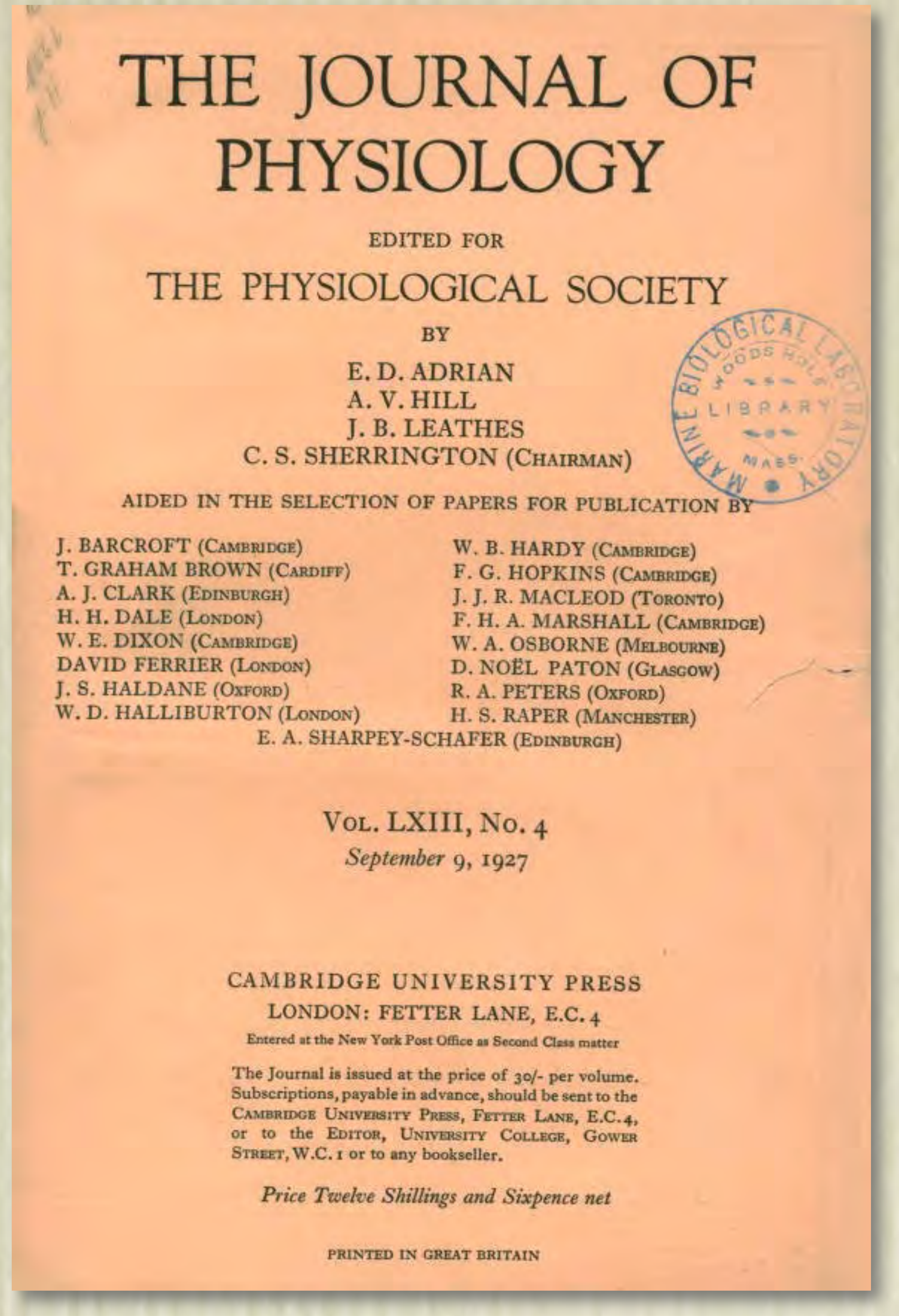


THE ACTION OF LIGHT ON THE EYE. Part I. The Discharge of Impulses in the Optic Nerve and its Relation to the Electric Changes in the Retina.

By E. D. ADRIAN and RACHEL MATTHEWS.

*(From the Physiological Laboratory, Cambridge.)*



THE JOURNAL OF PHYSIOLOGY

EDITED FOR

THE PHYSIOLOGICAL SOCIETY

BY

E. D. ADRIAN
A. V. HILL
J. B. LEATHES
C. S. SHERRINGTON (CHAIRMAN)

AIDED IN THE SELECTION OF PAPERS FOR PUBLICATION BY

J. BARCROFT (CAMBRIDGE)
T. GRAHAM BROWN (CARDIFF)
A. J. CLARK (EDINBURGH)
H. H. DALE (LONDON)
W. E. DIXON (CAMBRIDGE)
DAVID FERRIER (LONDON)
J. S. HALDANE (OXFORD)
W. D. HALLIBURTON (LONDON)
W. B. HARDY (CAMBRIDGE)
F. G. HOPKINS (CAMBRIDGE)
J. J. R. MACLEOD (TORONTO)
F. H. A. MARSHALL (CAMBRIDGE)
W. A. OSBORNE (MELBOURNE)
D. NOËL PATON (GLASGOW)
R. A. PETERS (OXFORD)
H. S. RAPER (MANCHESTER)
E. A. SHARPEY-SCHAFER (EDINBURGH)

VOL. LXIII, No. 4
*September* 9, 1927

CAMBRIDGE UNIVERSITY PRESS
LONDON: FETTER LANE, E.C. 4
Entered at the New York Post Office as Second Class matter

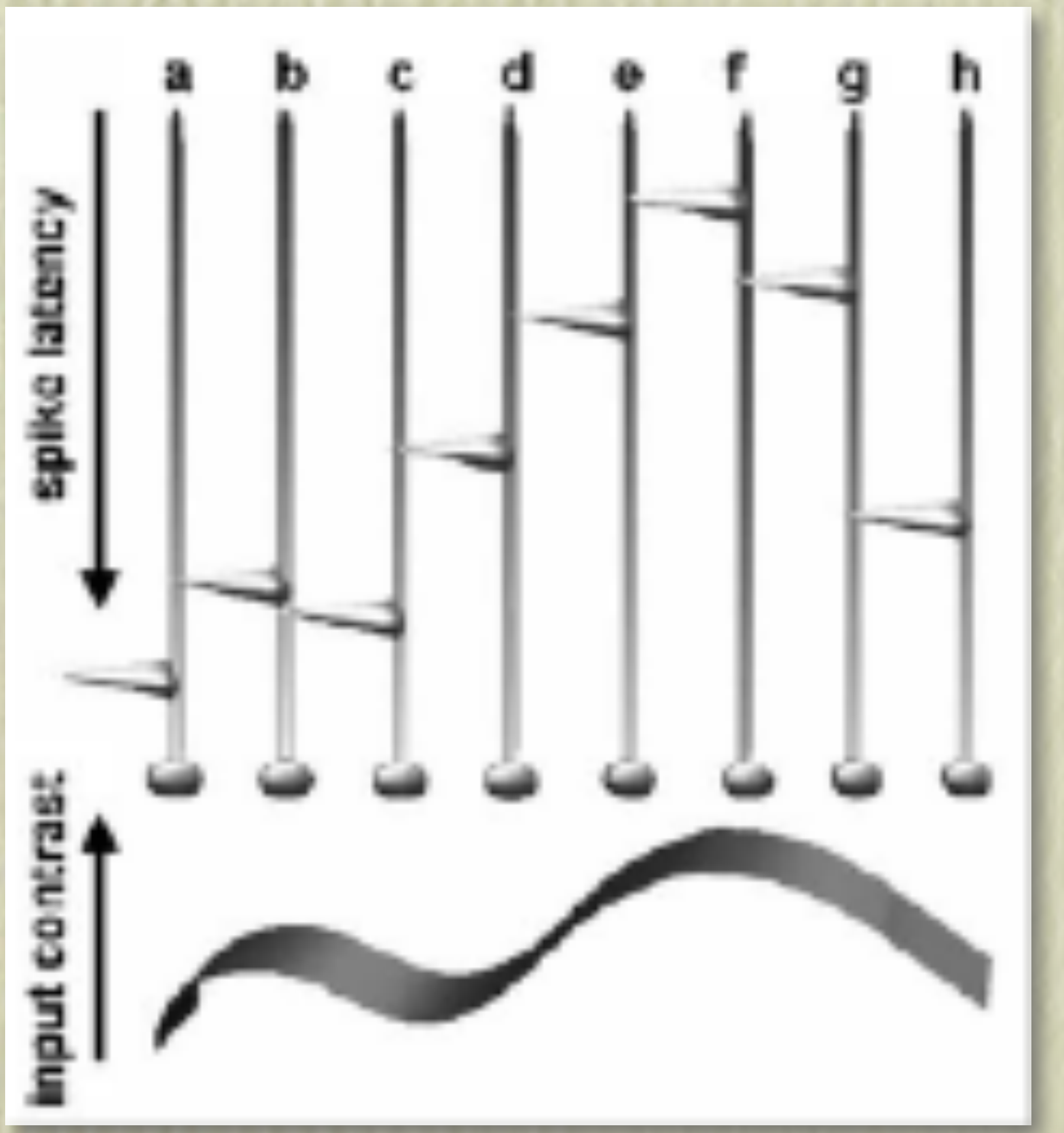
The Journal is issued at the price of 30/- per volume. Subscriptions, payable in advance, should be sent to the CAMBRIDGE UNIVERSITY PRESS, FETTER LANE, E.C. 4, or to the EDITOR, UNIVERSITY COLLEGE, GOWER STREET, W.C. 1 or to any bookseller.

*Price Twelve Shillings and Sixpence net*

PRINTED IN GREAT BRITAIN

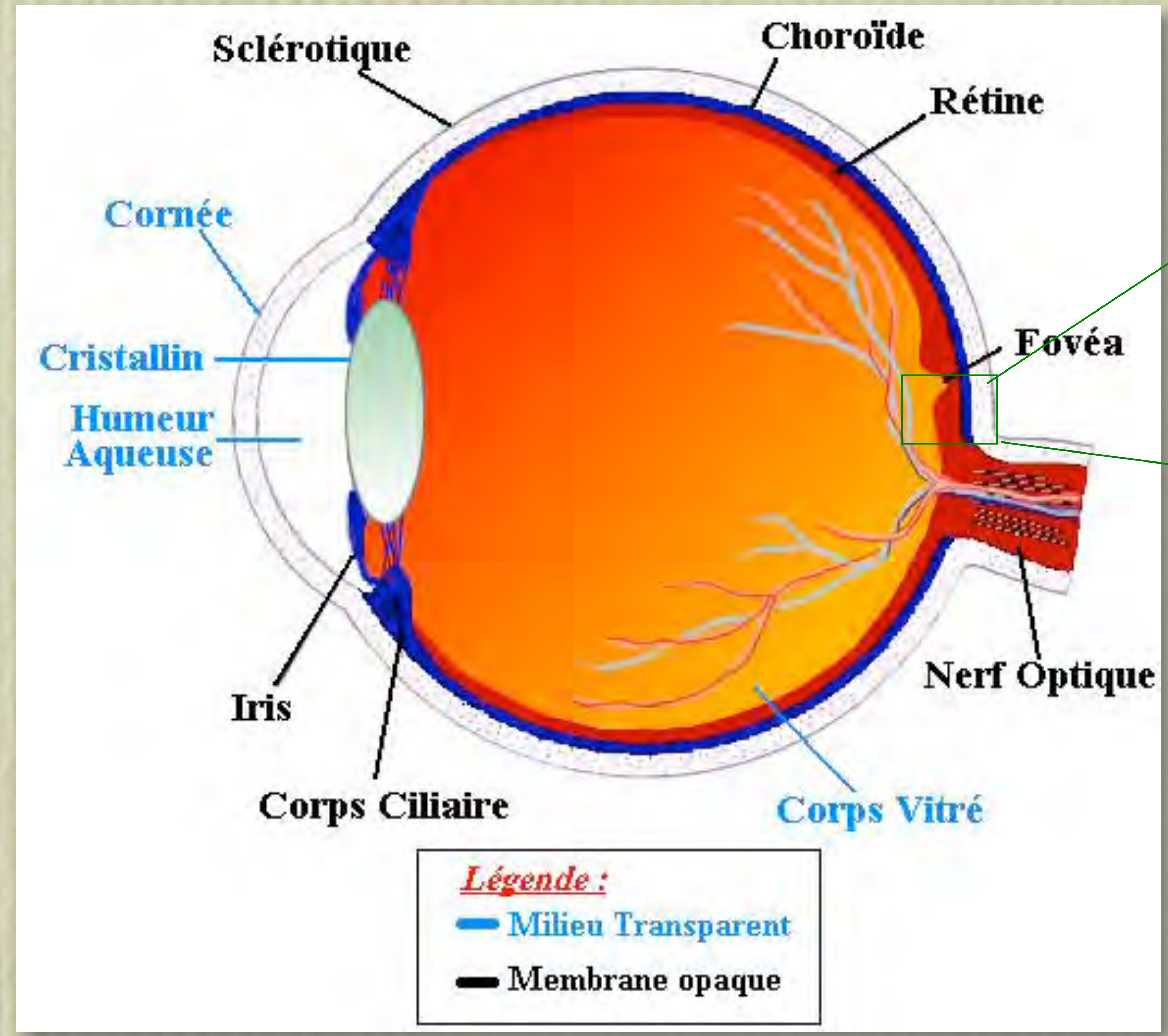
- The retina is an intensity to delay converter
- This basic physiological fact was ignored for over 60 years!

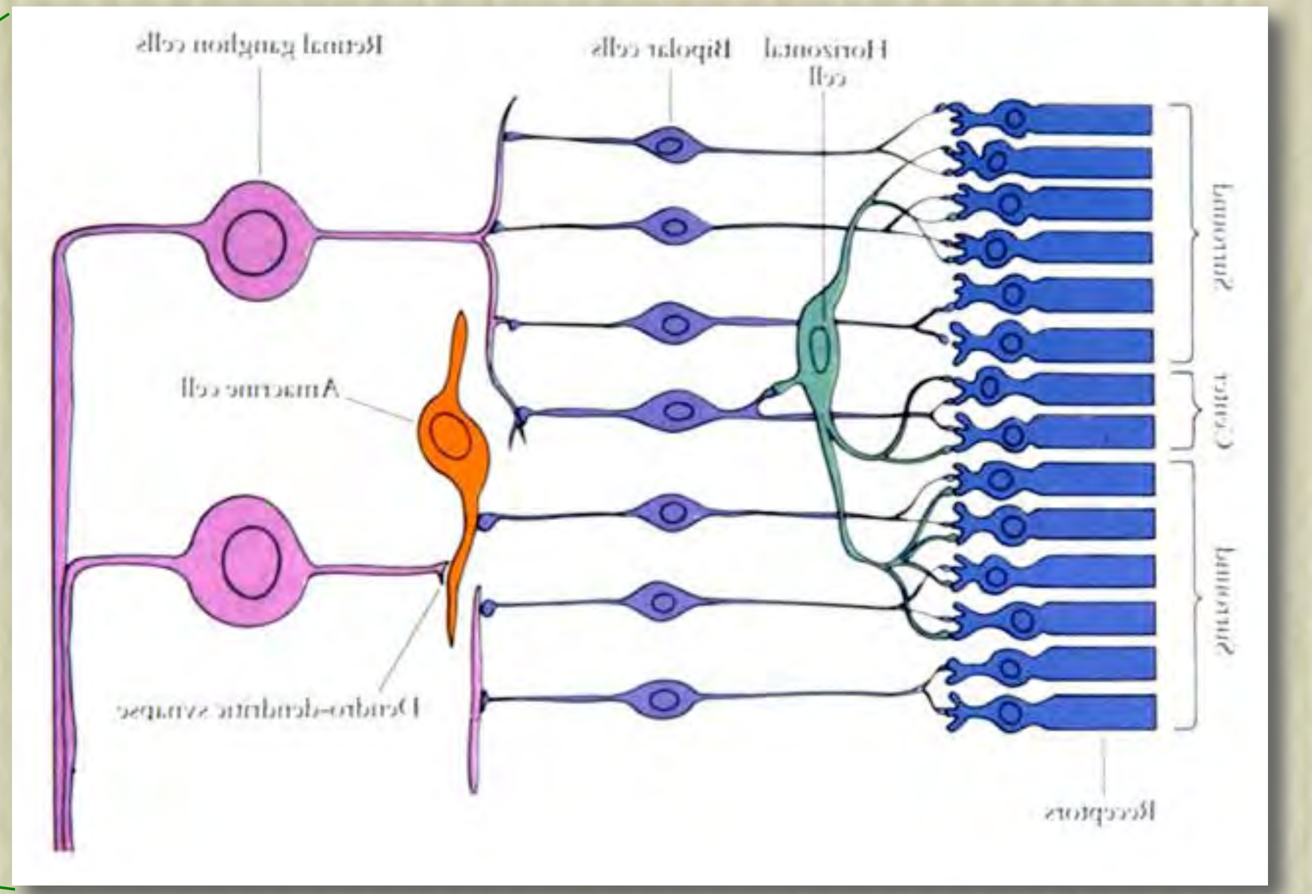
# Spike-based Processing



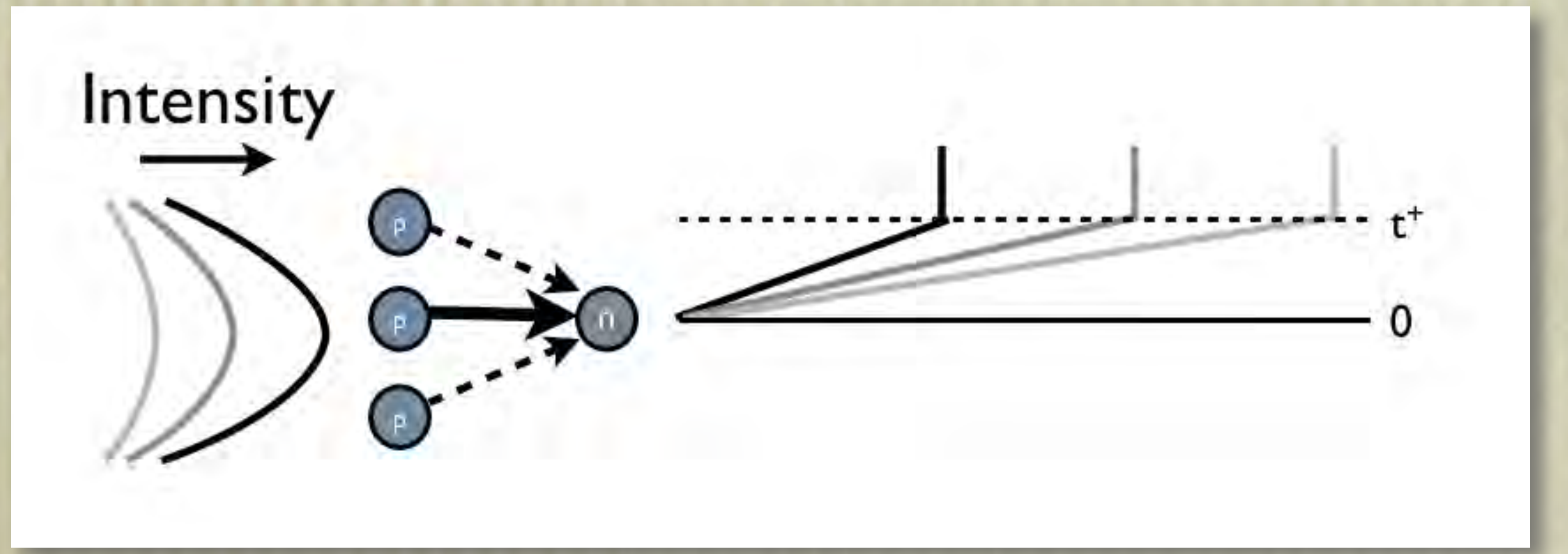

- Processing with a wave of spikes
  - The most strongly activated cells fire first
  - Information can be encoded in the order of firing

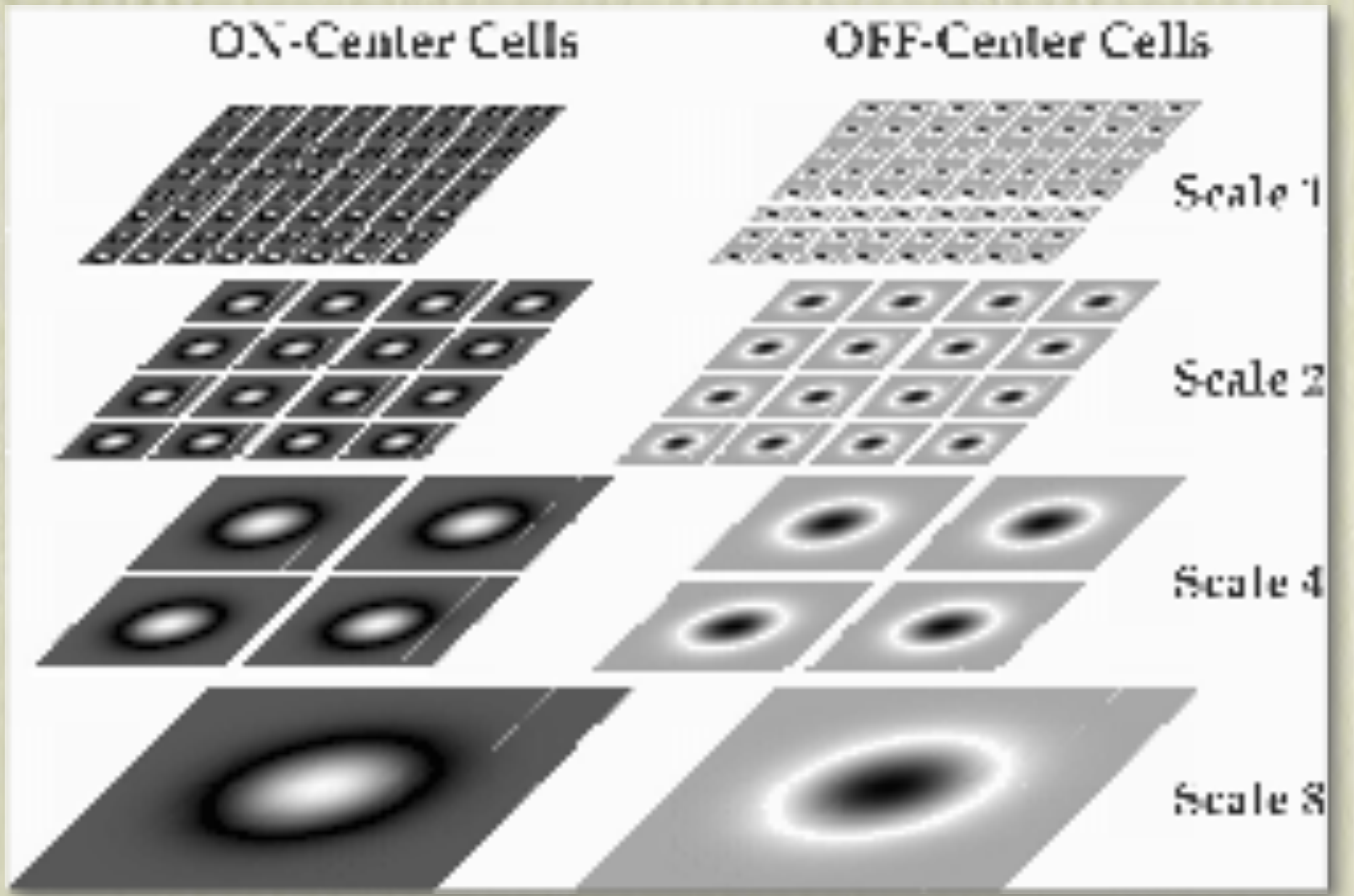
# Coding in the Optic Nerve



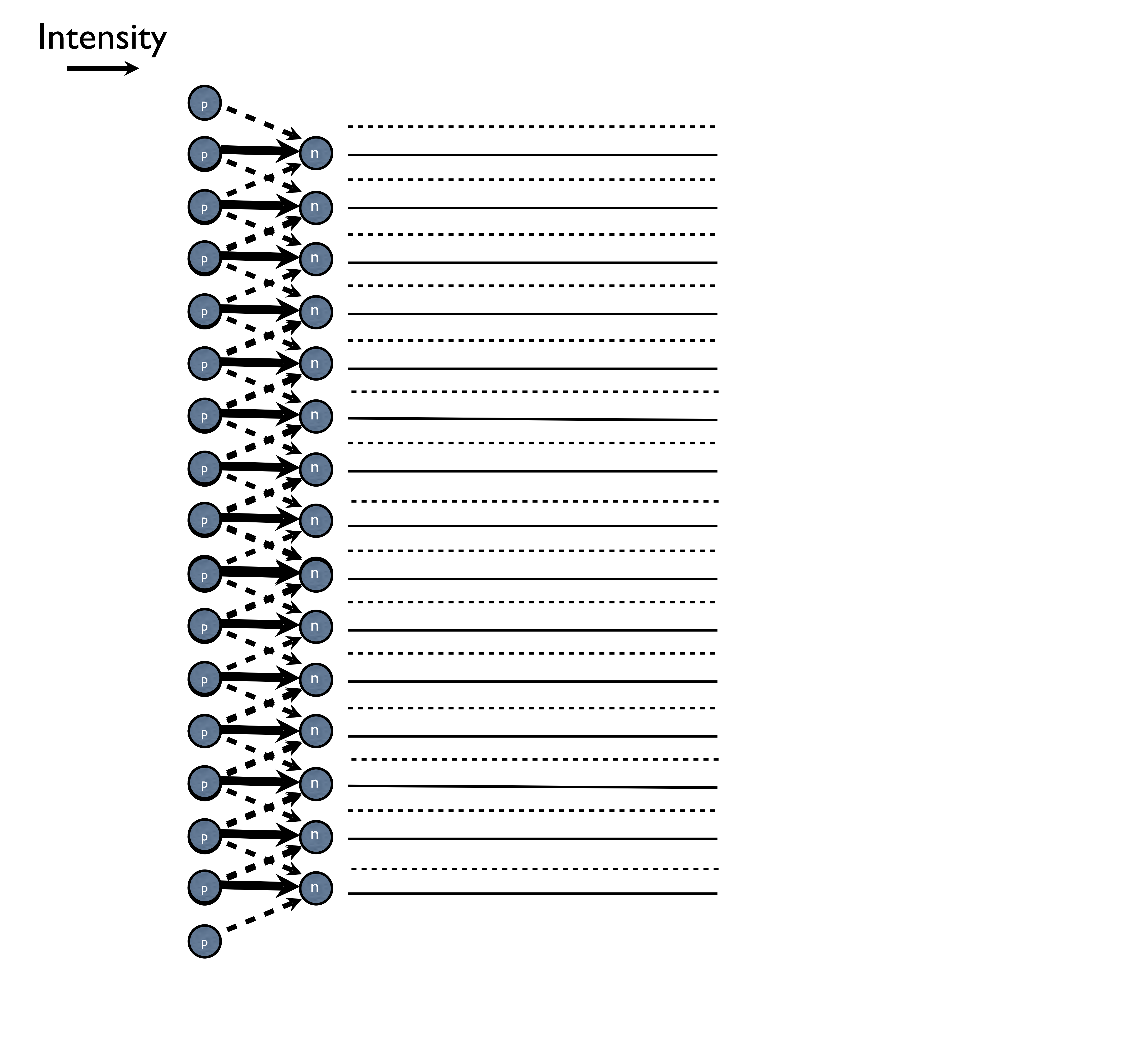



1,000,000
fibres

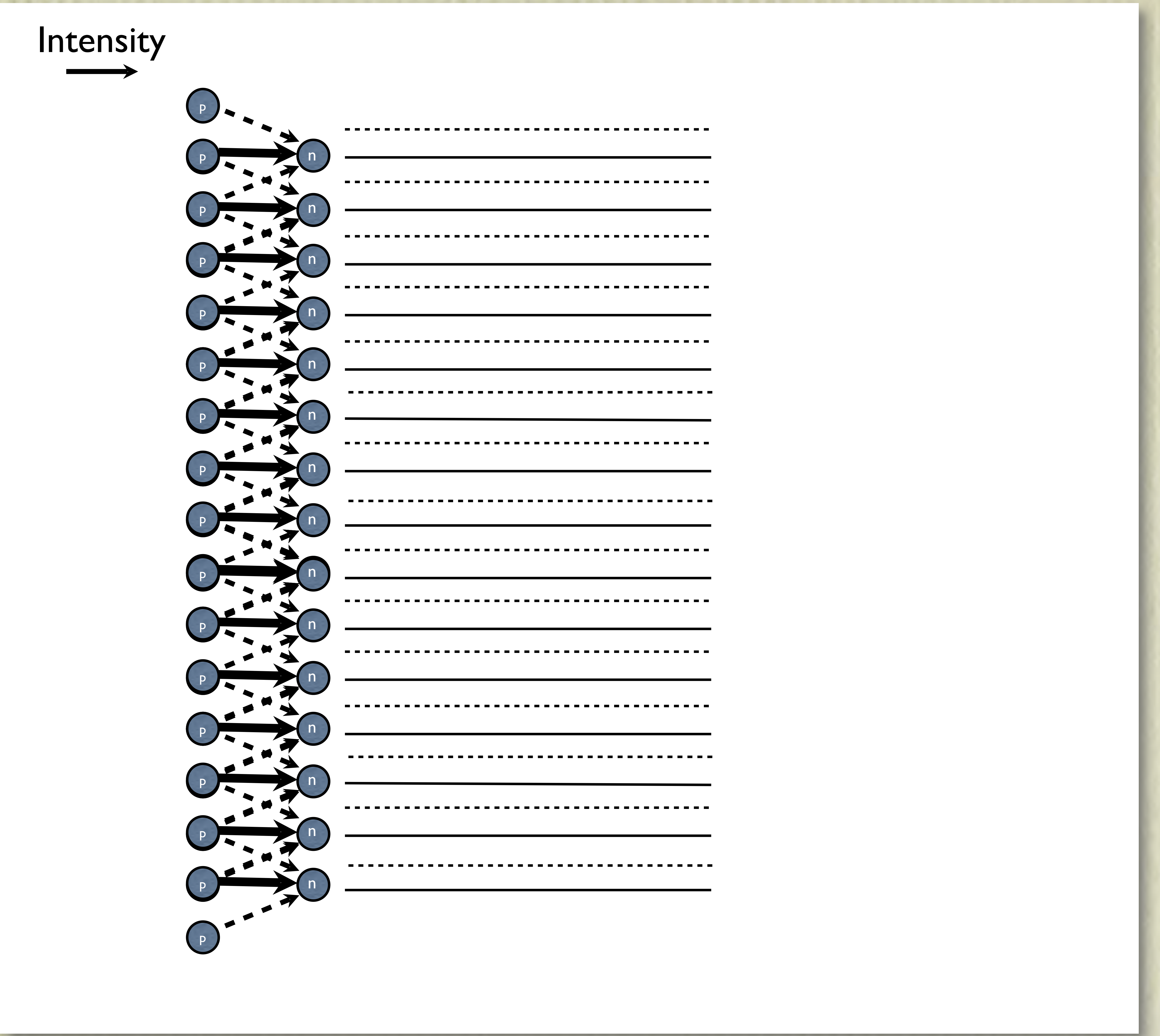


# Coding in the Optic Nerve



# Coding in the Optic Nerve




A mini retina
32 x 32 pixels

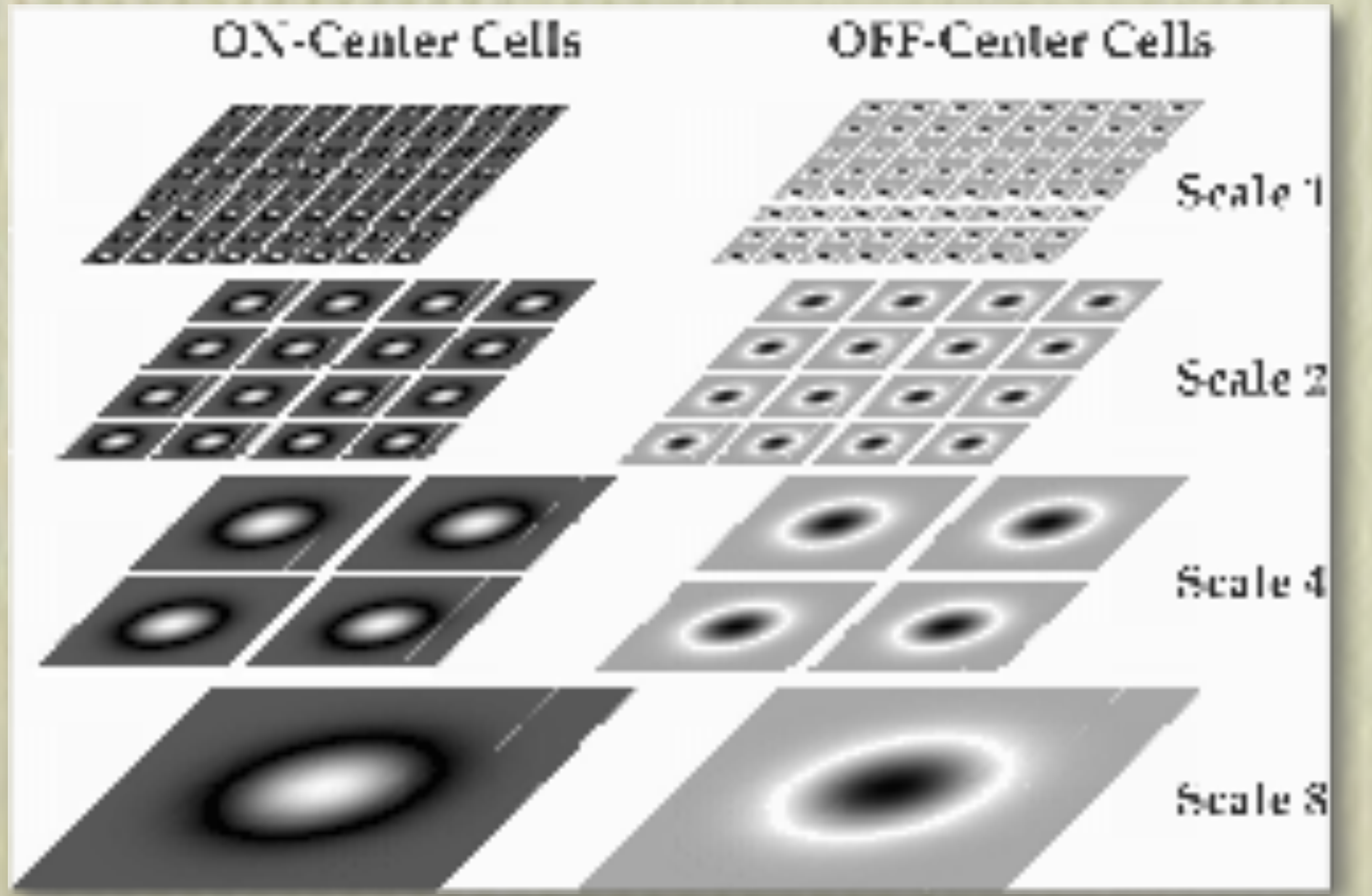
# Coding with Spike Ordering

Rate Coding Versus Temporal Order Coding: What the Retinal Ganglion Cells Tell the Visual Cortex

Rufin Van Rullen Simon J. Thorpe *Neural Computation* 13, 1255–1283 (2001)

Example

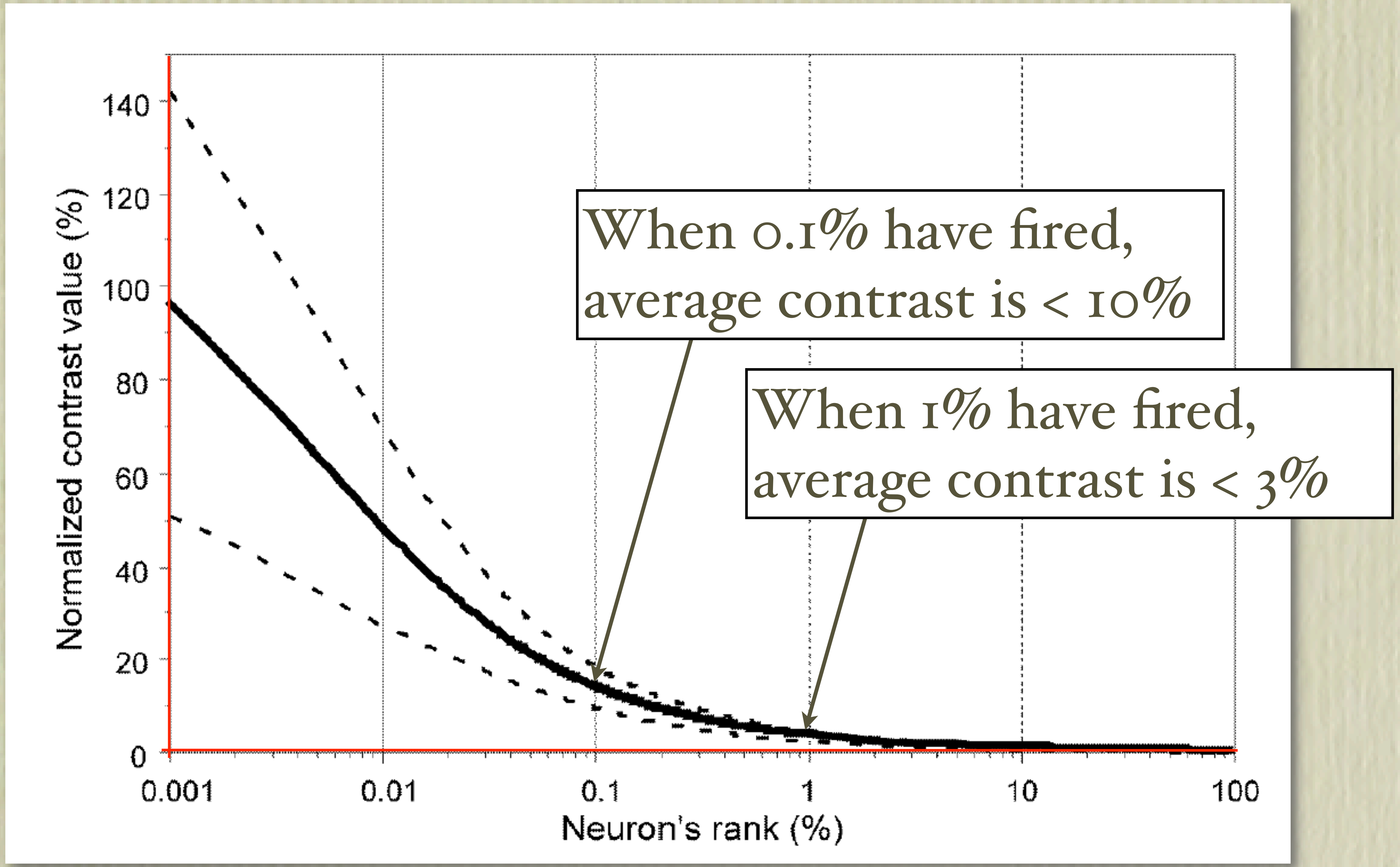
- A toy retina




Less than 1% of cells need to fire for recognition!

# Rank Order and Contrast

- How does contrast change with rank order?
  - Data from 3000 natural images




Forget the other 99%!

The first 1% does all the work!

# The Unreliability of Poisson processes

Rate coding versus temporal order coding: a theoretical approach

Jacques Gautrais *, Simon Thorpe BioSystems 48 (1998)

If a Poisson process generates 1 spike in 10 ms, you can be 90% confident that the underlying rate lies between 5 and 474 Hz

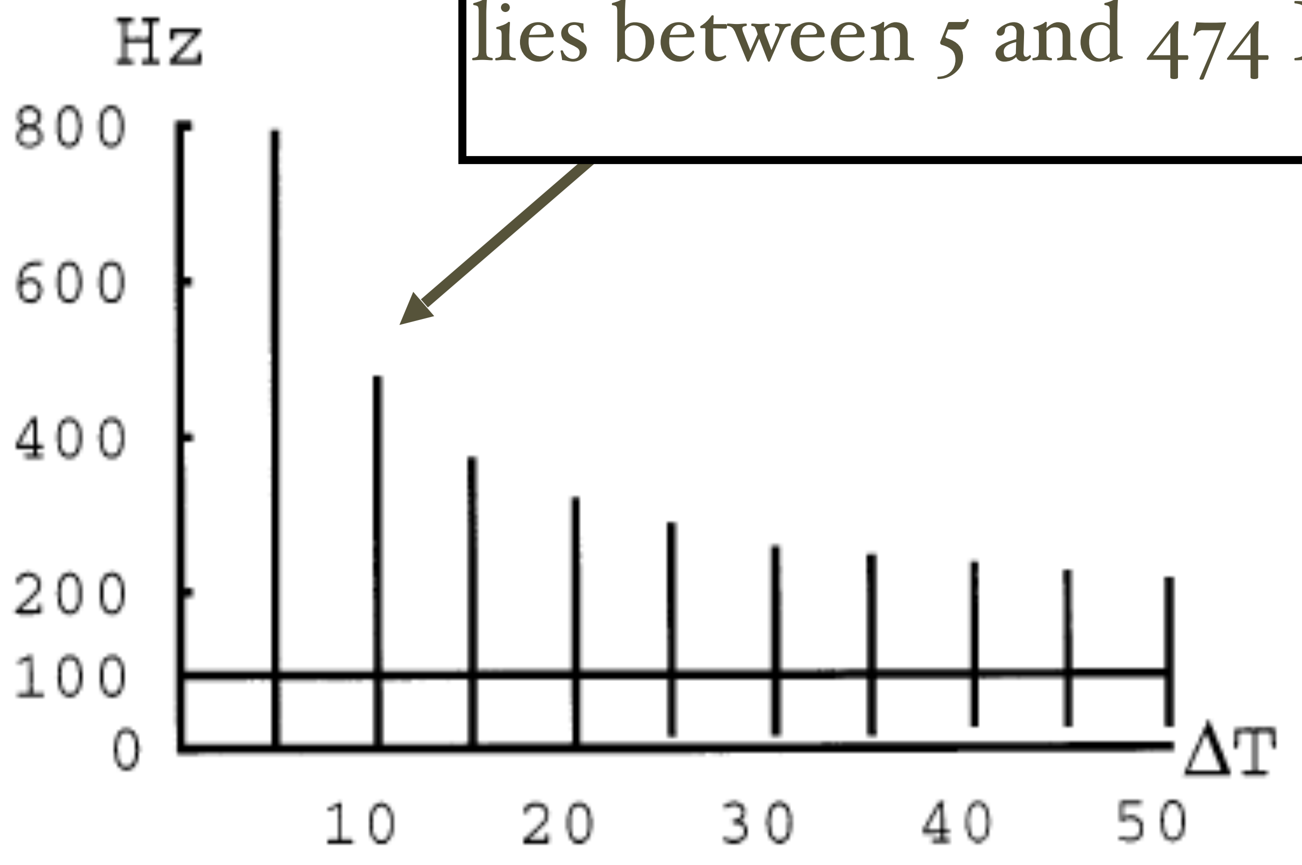



Fig. 1. Confidence interval (90%) on the true frequency of a Poisson process as a function of the time window of evaluation, and given an observed frequency at 100 Hz.

Even with 30 redundant neurons, and 30 spikes in 10 ms, the confidence interval would be 72-135 Hz

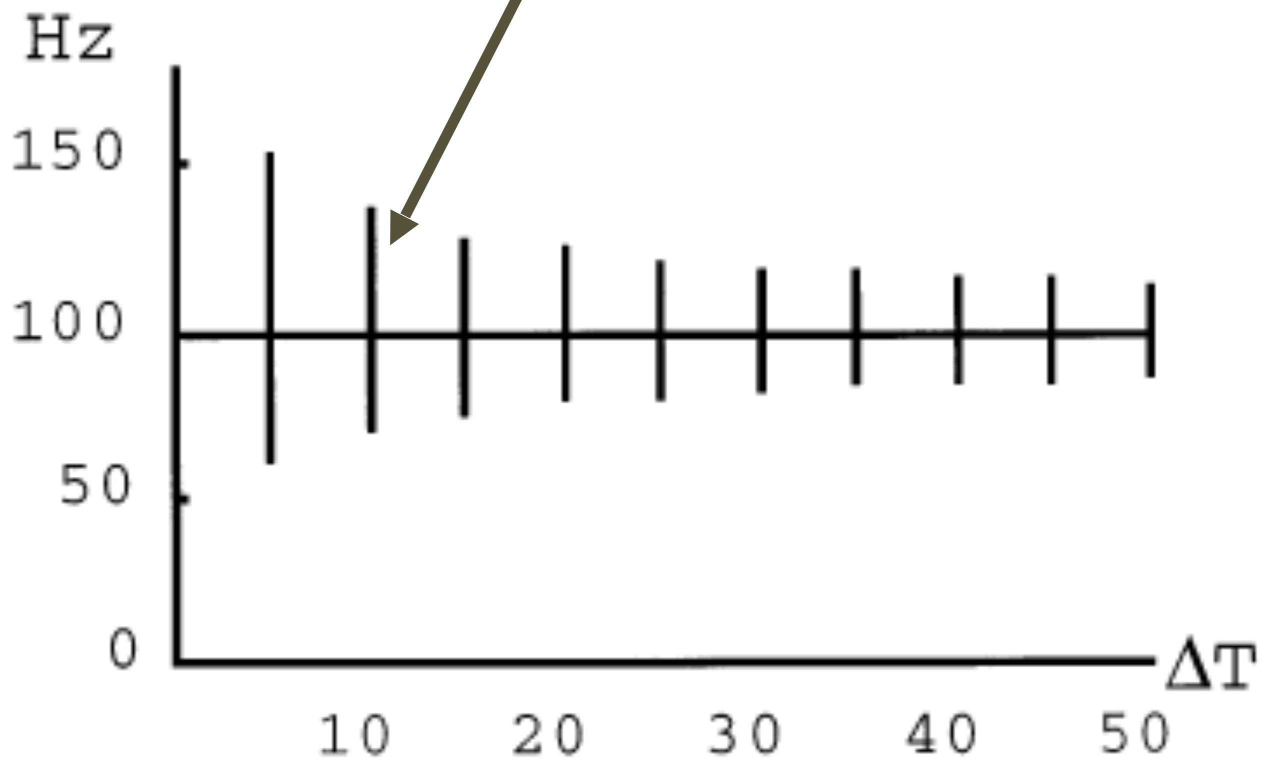
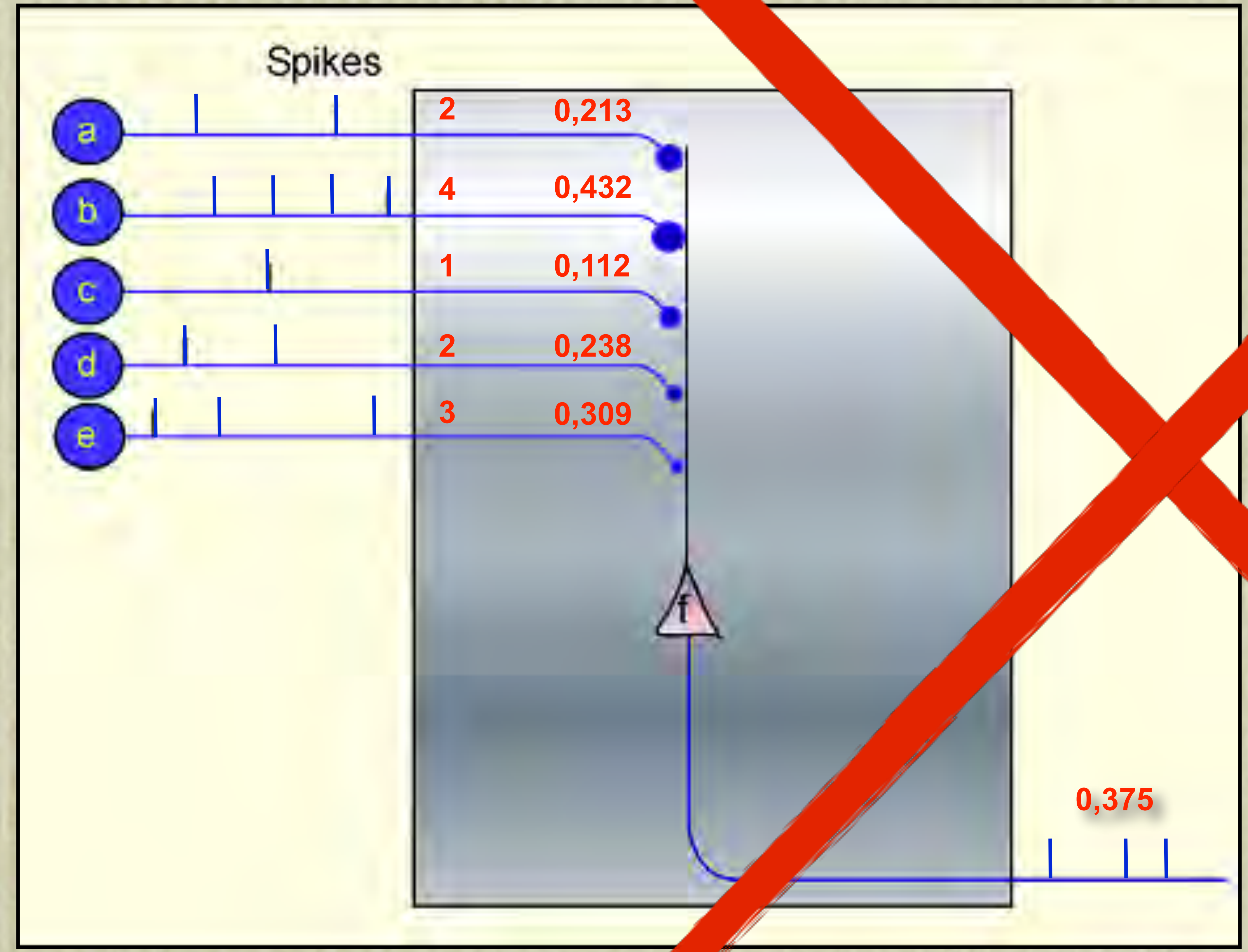
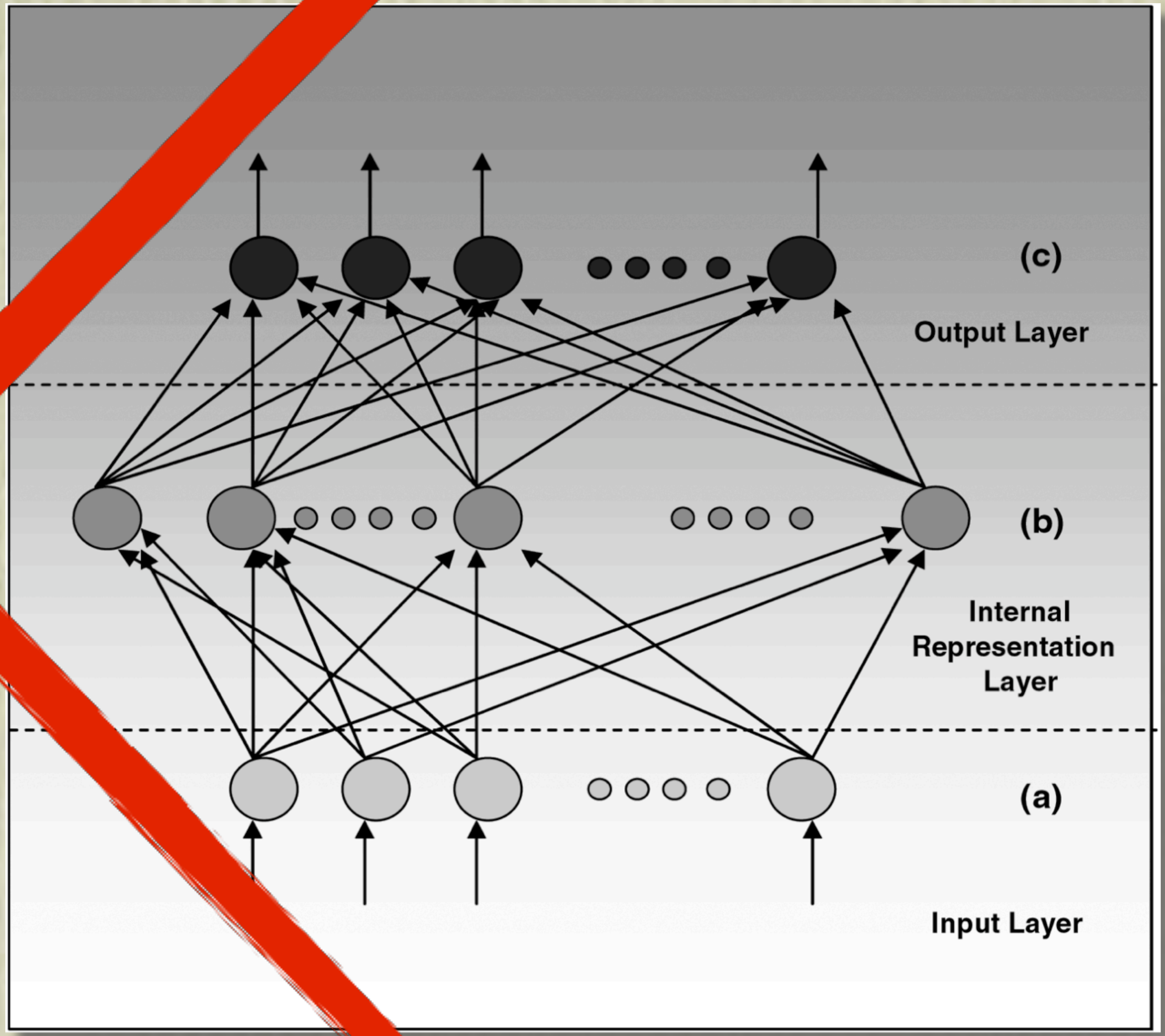



Fig. 2. Confidence interval (90%) on the true frequency of 30 redundant Poisson processes as a function of the time window of evaluation, and given an observed frequency at 100 Hz.

- To obtain 100 ± 10 Hz in 10 ms would need 281 redundant neurons!

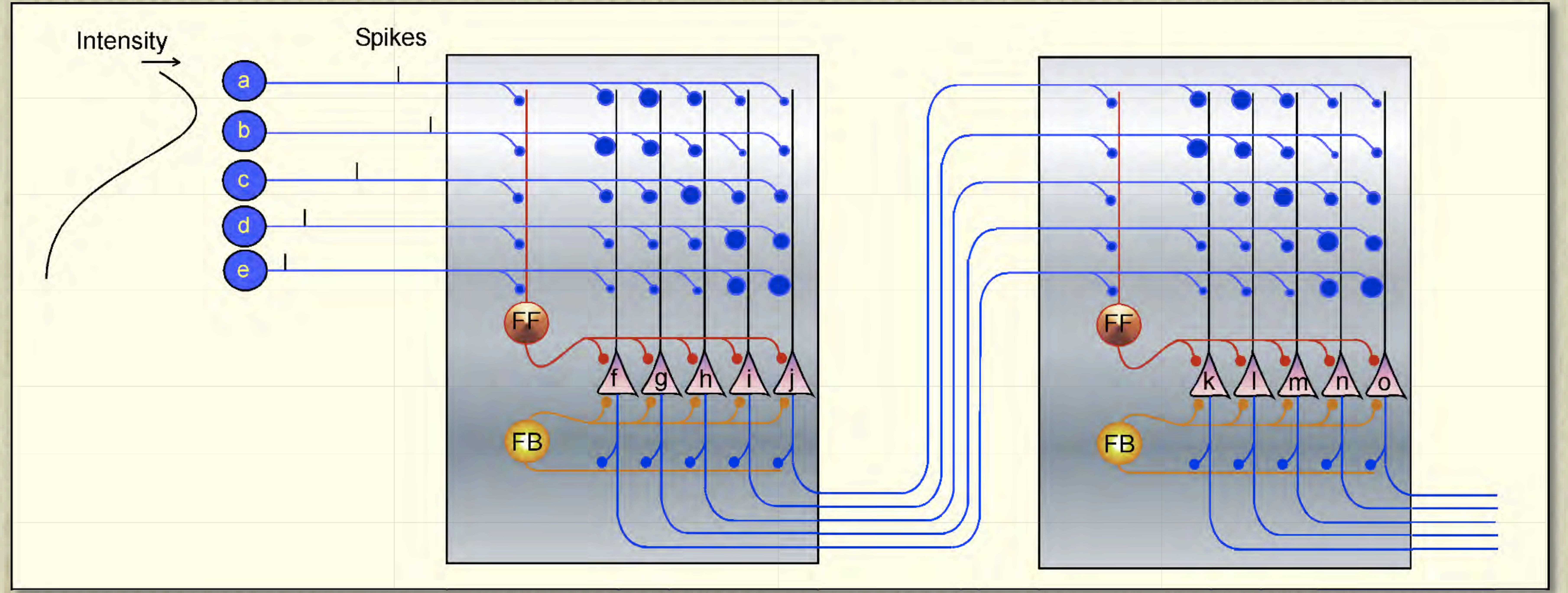
# The Classic View




Spikes really are important!

# Cortical Circuits




- Feed-forward inhibition
  - desensitisation
  - gives maximum importance to the first spikes

- Feedback inhibition
  - k- Winner take all
  - Controls the number of cells that are allowed to fire

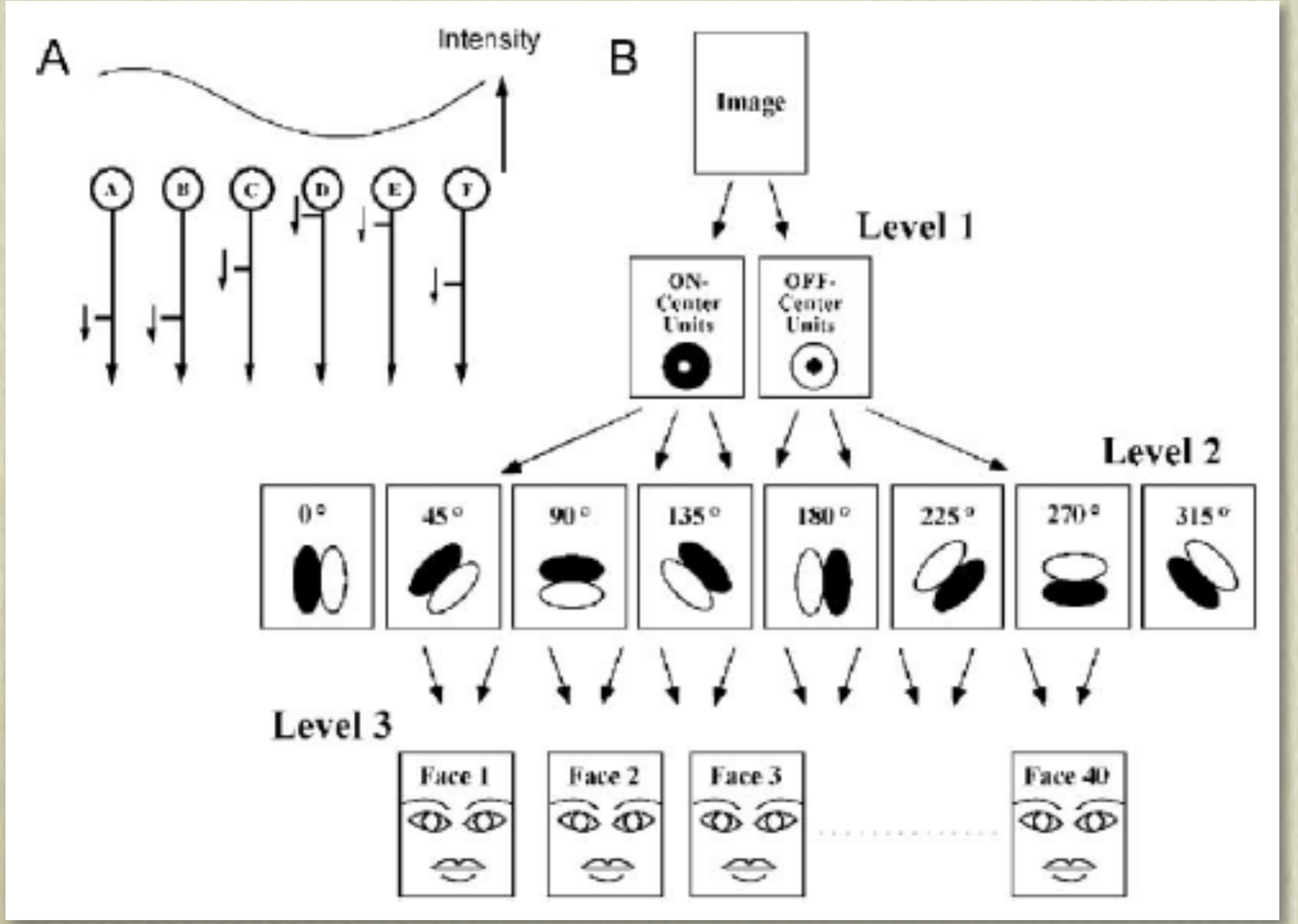
Can these ideas be used to solve real problems in vision?

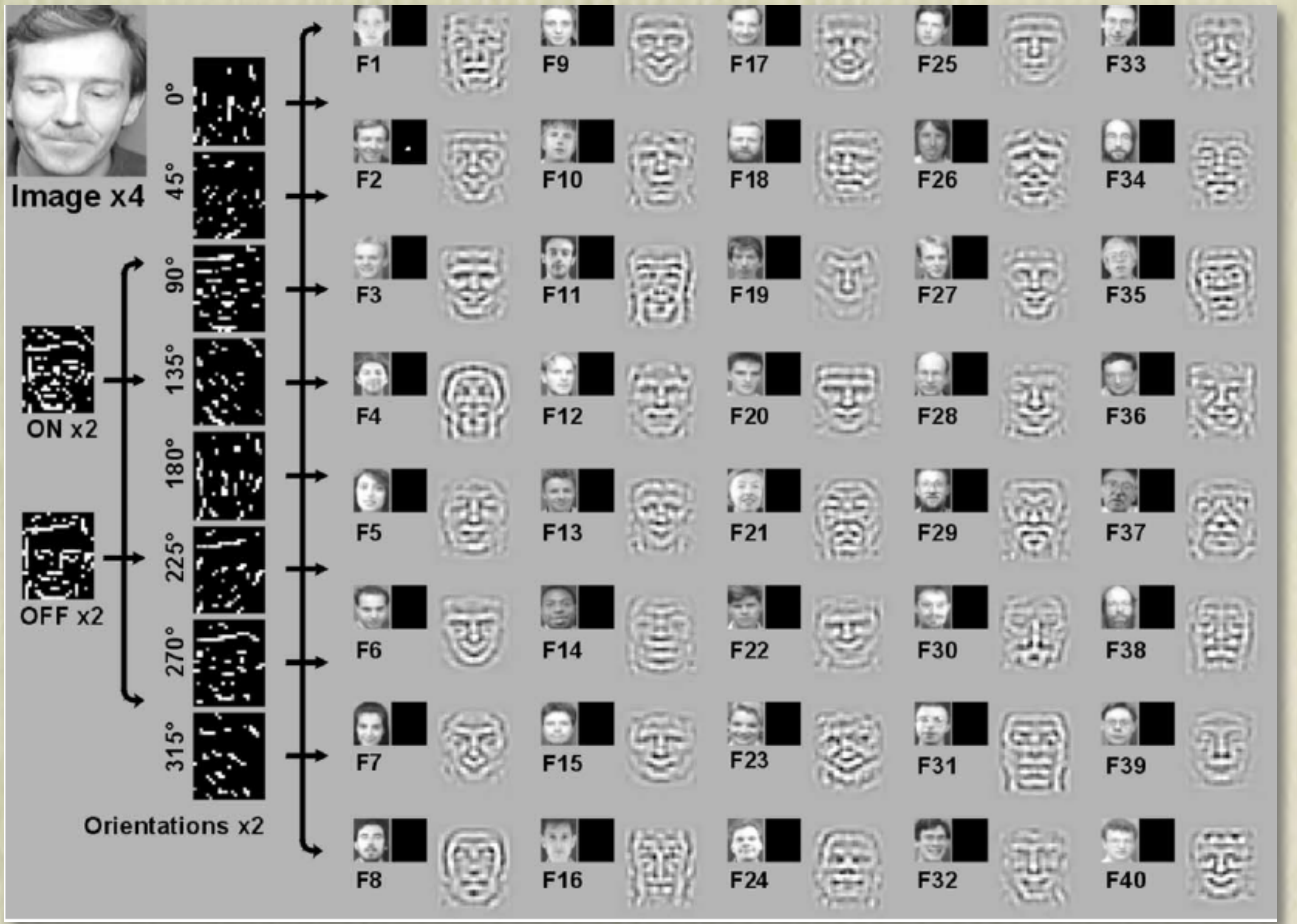
# One spike processing

Face identification using one spike per neuron: resistance to image degradations

A. Delorme*, S.J. Thorpe

Neural Networks 14 (2001) 795–803




- Face identification directly using oriented filters
- Relevance to Face Zapping task?

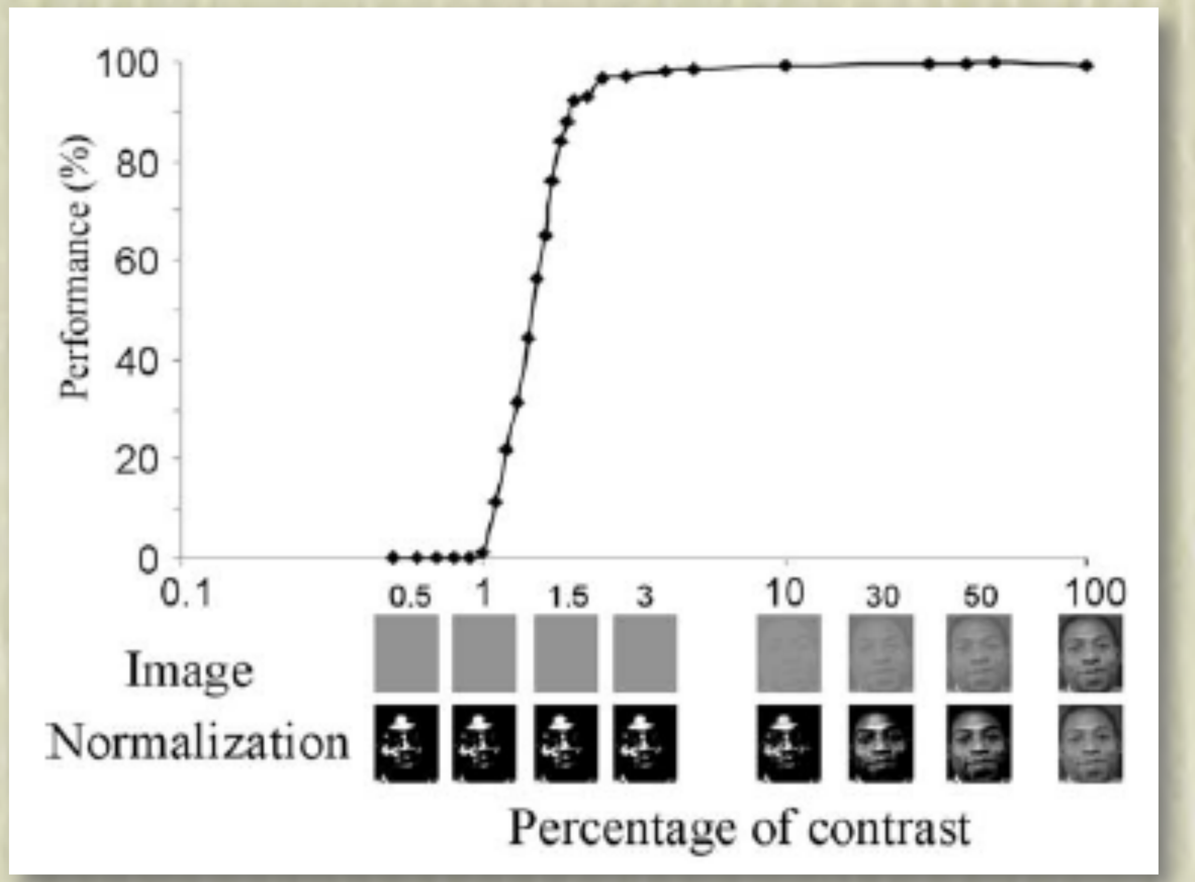
# One spike processing

Face identification using one spike per neuron: resistance to image degradations

A. Delorme*, S.J. Thorpe    Neural Networks 14 (2001) 795–803

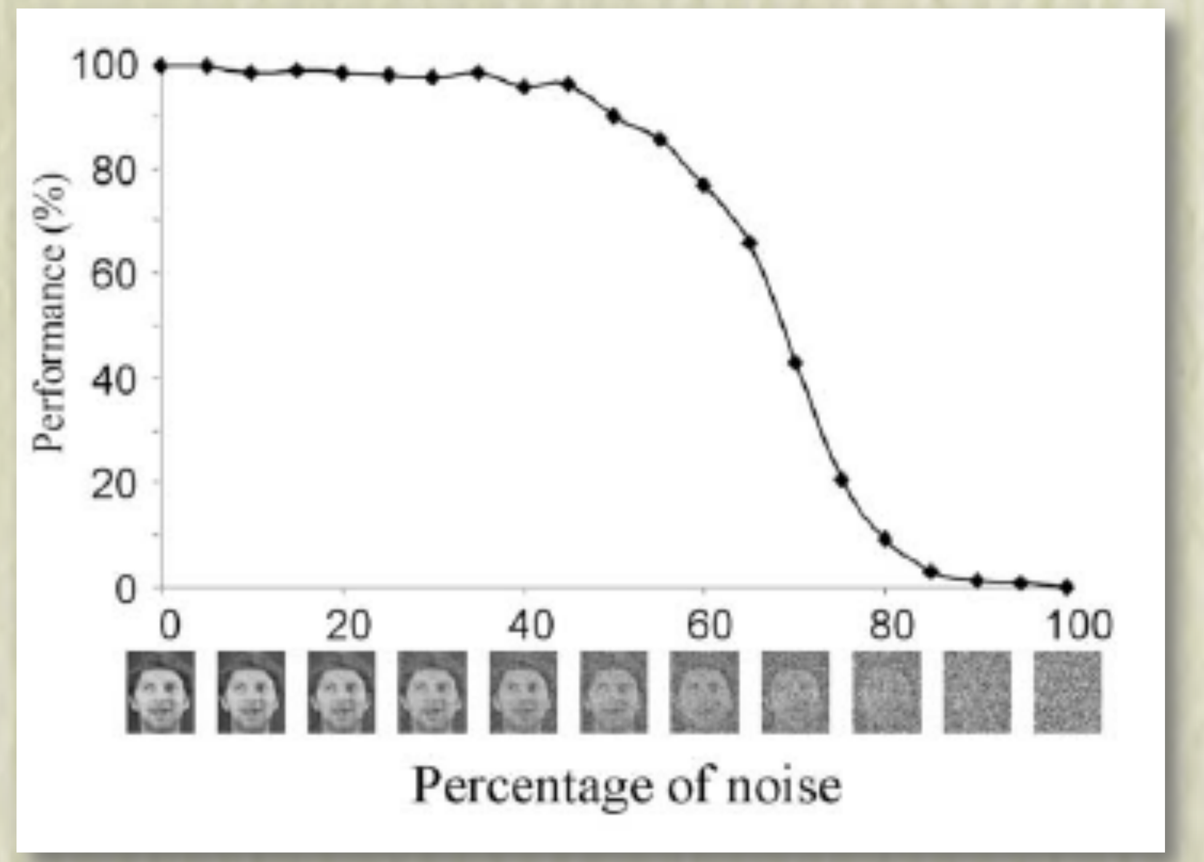
- Virtually all the faces correctly identified (393/400)

Very robust to low contrast




- 30 minutes of simulation time!

Very robust to noise



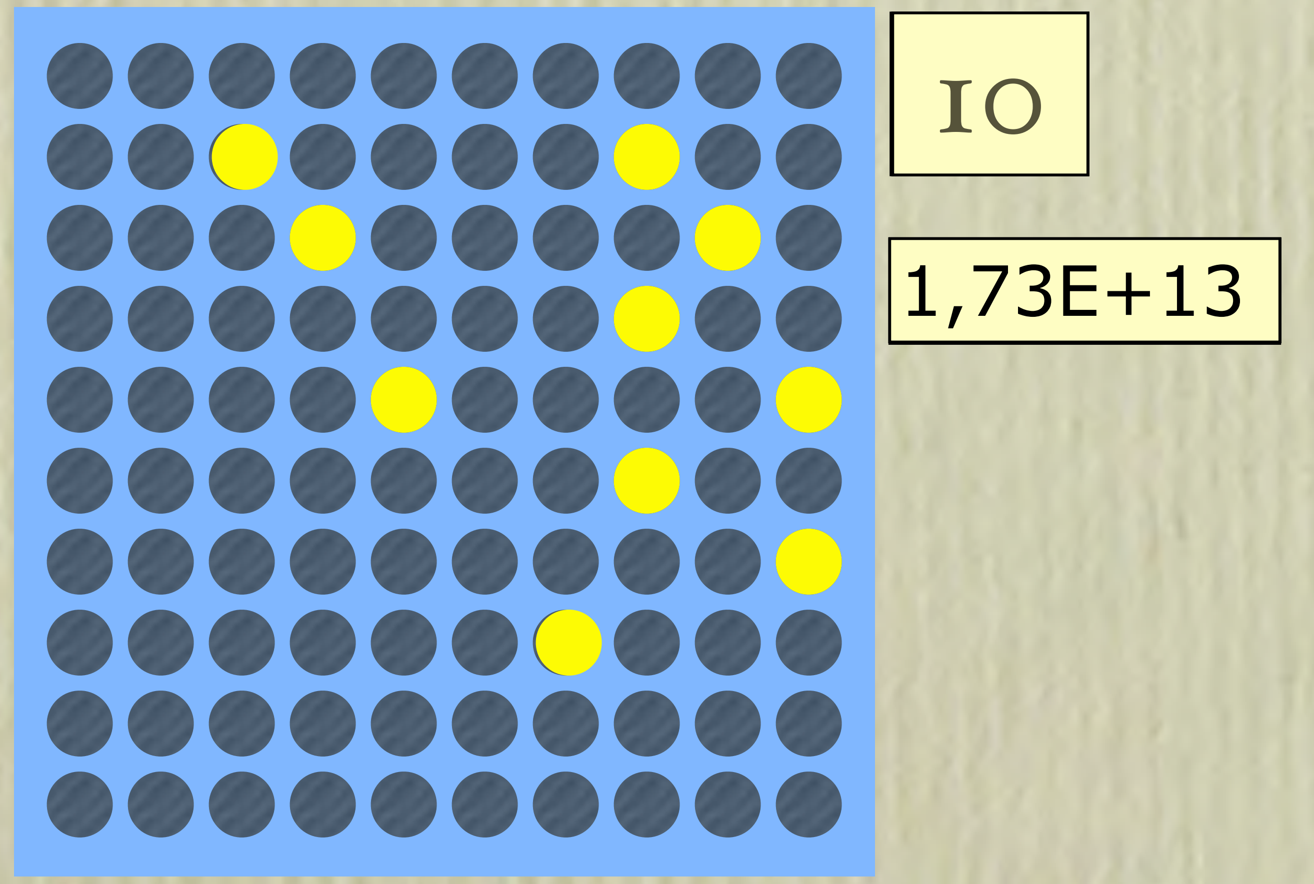

- 1999 Creation of SpikeNet Technology
- 2016 Acquition by BrainChip Inc

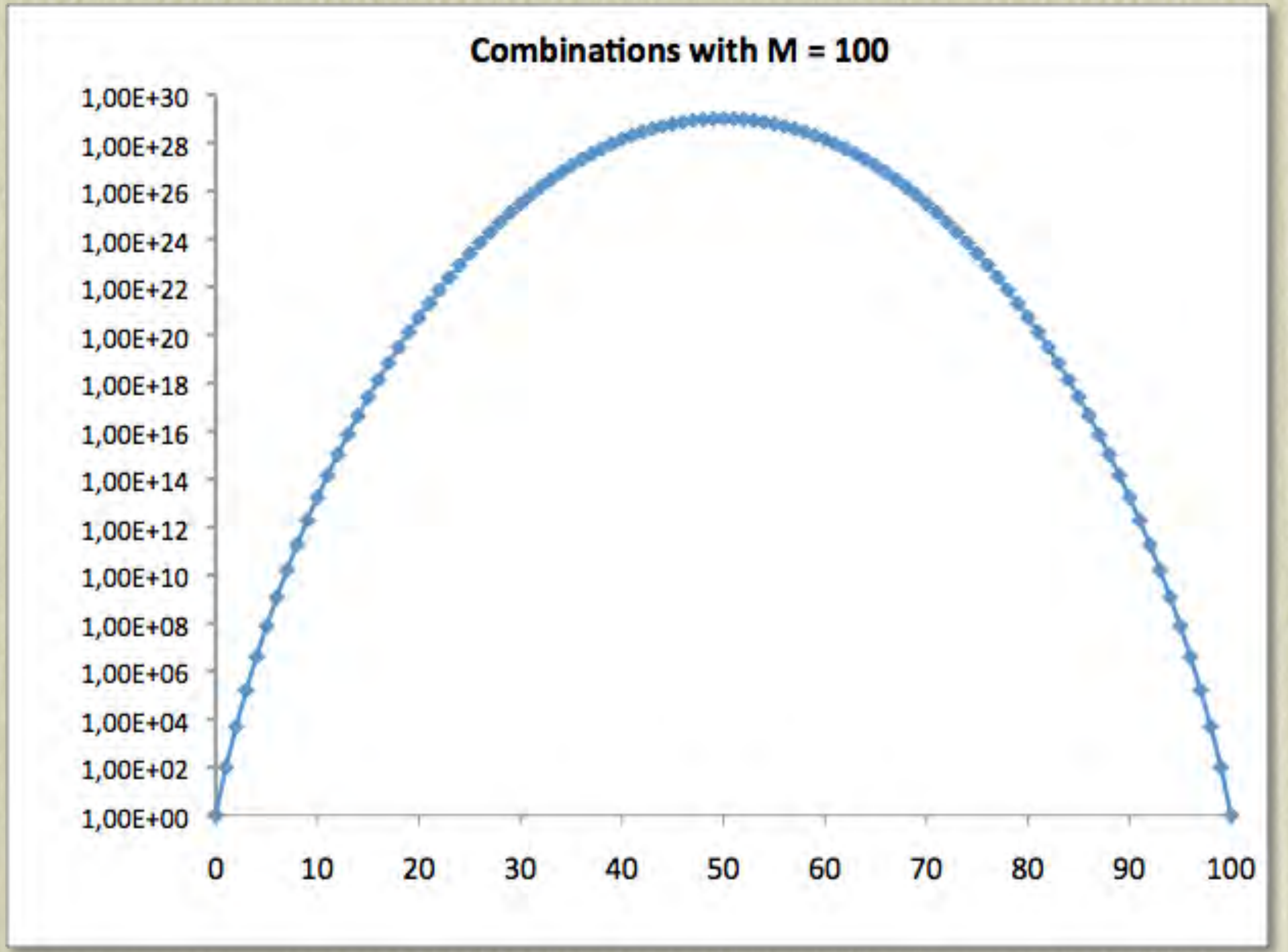
# Spikes and Sparsity

- With spikes
  - Easy to control the percentage of active neurons
- Solution to the k-Winner Take All problem
  - A counter circuit prevents more than k inputs from firing

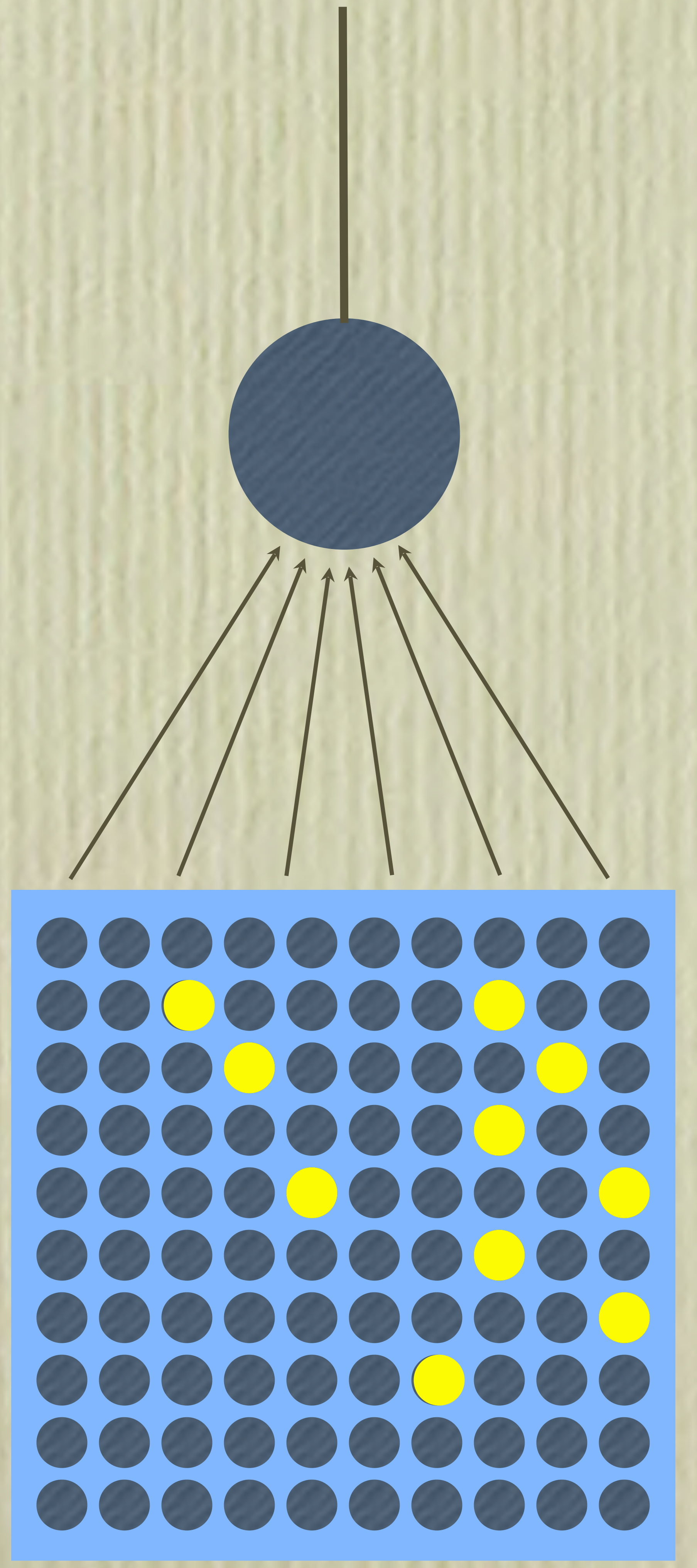
# N of M coding

$$Combinations = \frac{m!}{n!(m-n)!}$$



# Generating Selectivity



- Suppose that the neuron has 10 synapses
- Suppose that the percentage of active inputs is fixed at 10%
- What is the likelihood of having a given number of synapses active?

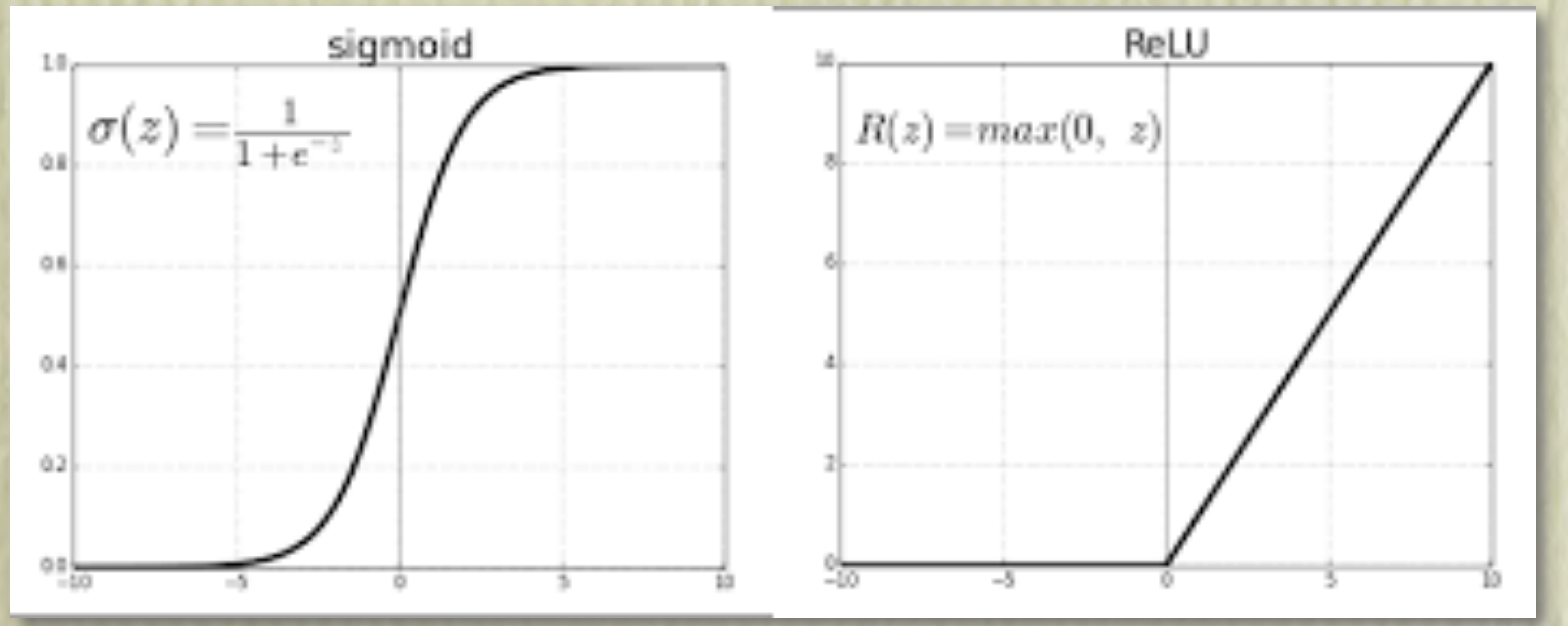
| Hits | Probability | Probability of exceeding threshold |
|---|---|---|
| 0 | 0,3486784401 | 1,0000000000 |
| 1 | 0,3874204890 | 0,6513215599 |
| 2 | 0,1937102445 | 0,2639010709 |
| 3 | 0,0573956280 | 0,0701908264 |
| 4 | 0,0111602610 | 0,0127951984 |
| 5 | 0,0014880348 | 0,0016349374 |
| 6 | 0,0001377810 | 0,0001469026 |
| 7 | 0,0000087480 | 0,0000091216 |
| 8 | 0,0000003645 | 0,0000003736 |
| 9 | 0,0000000090 | 0,0000000091 |
| 10 | 0,0000000001 | 0,0000000001 |

- With a threshold of 4 the neuron would only have a 1% chance of firing with a random input
- With a threshold of 5, the probability drops to 0.1%

- Spikes make it easy to make neurons that are arbitrarily selective

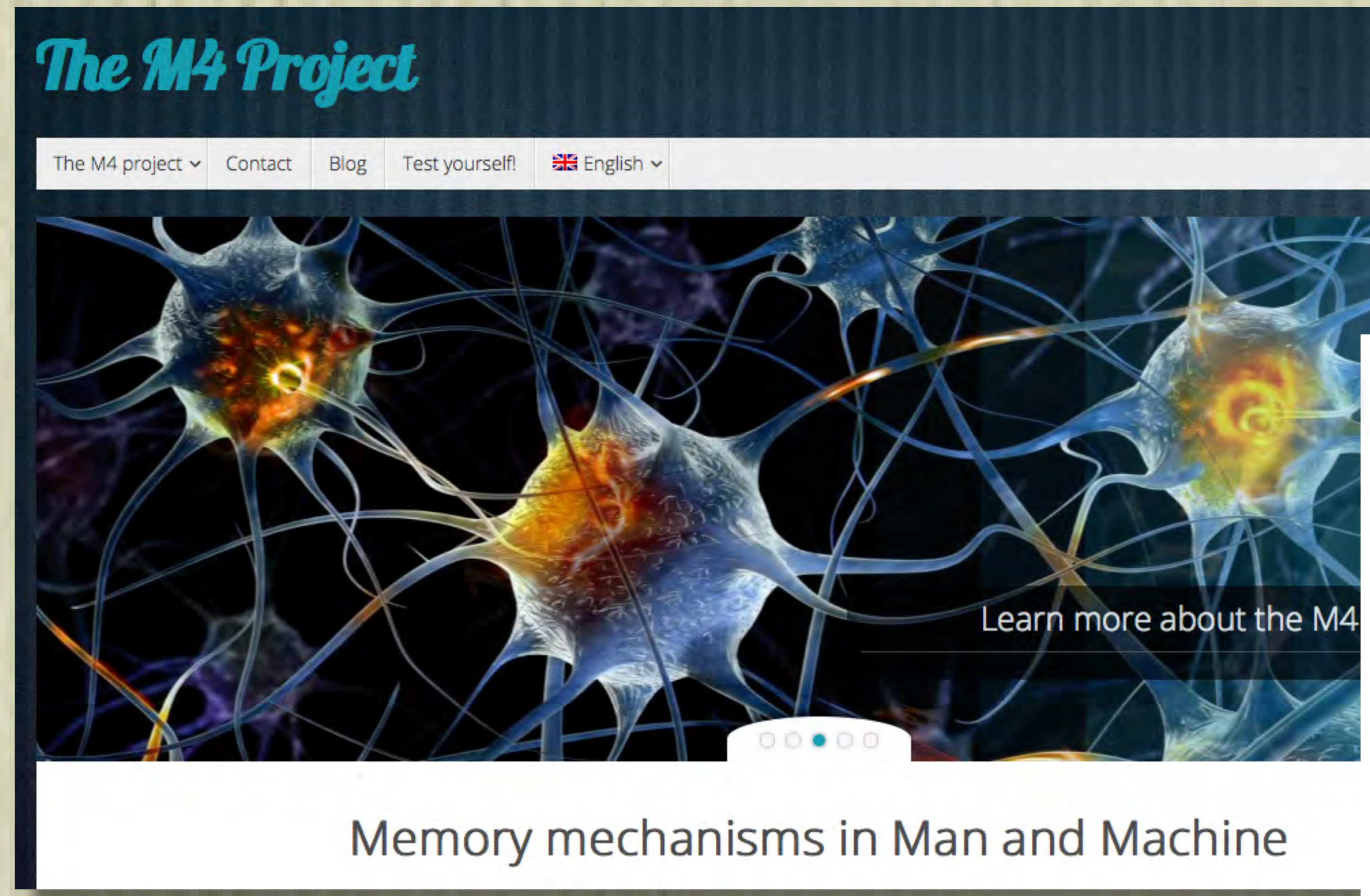
# The Other Trick

- What changed between the 1980s and 2012?
  - Availability of very fast GPU hardware
  - Availability of huge amounts of labelled data for training
  - A switch from Sigmoidal to ReLU functions




- This is equivalent to using Spiking Neurons!

# ERC Grant (2013-19)




**Proposal summary**

The project aims to validate a set of 10 provocative claims.

1) Humans can recognise visual and auditory stimuli that they have not experienced for decades.
2) Recognition after very long delays is possible without ever reactivating the memory trace in the intervening period.
3) These very long term memories require an initial memorisation phase, during which memory strength increases roughly linearly with the number of presentations
4) A few tens of presentations can be enough to form a memory that can last a lifetime.
5) Attention-related oscillatory brain activity can help store memories efficiently and rapidly
6) Storing such very long-term memories involves the creation of highly selective "Grandmother Cells" that only fire if the original training stimulus is experienced again.
7) The neocortex contains large numbers of totally silent cells ("Neocortical Dark Matter") that constitute the long-term memory store.
8) Grandmother Cells can be produced using simple spiking neural network models with Spike-Time Dependent Plasticity (STDP) and competitive inhibitory lateral connections.
9) This selectivity only requires binary synaptic weights that are either "on" or "off", greatly simplifying the problem of maintaining the memory over long periods.
10) Artificial systems using memristor-like devices can implement the same principles, allowing the development of powerful new processing architectures that could replace conventional computing hardware.

# Grandmother Cells?

On the Biological Plausibility of Grandmother Cells: Implications for Neural Network Theories in Psychology and Neuroscience

Jeffrey S. Bowers
University of Bristol

Psychological Review
2009, Vol. 116, No. 1, 220–251

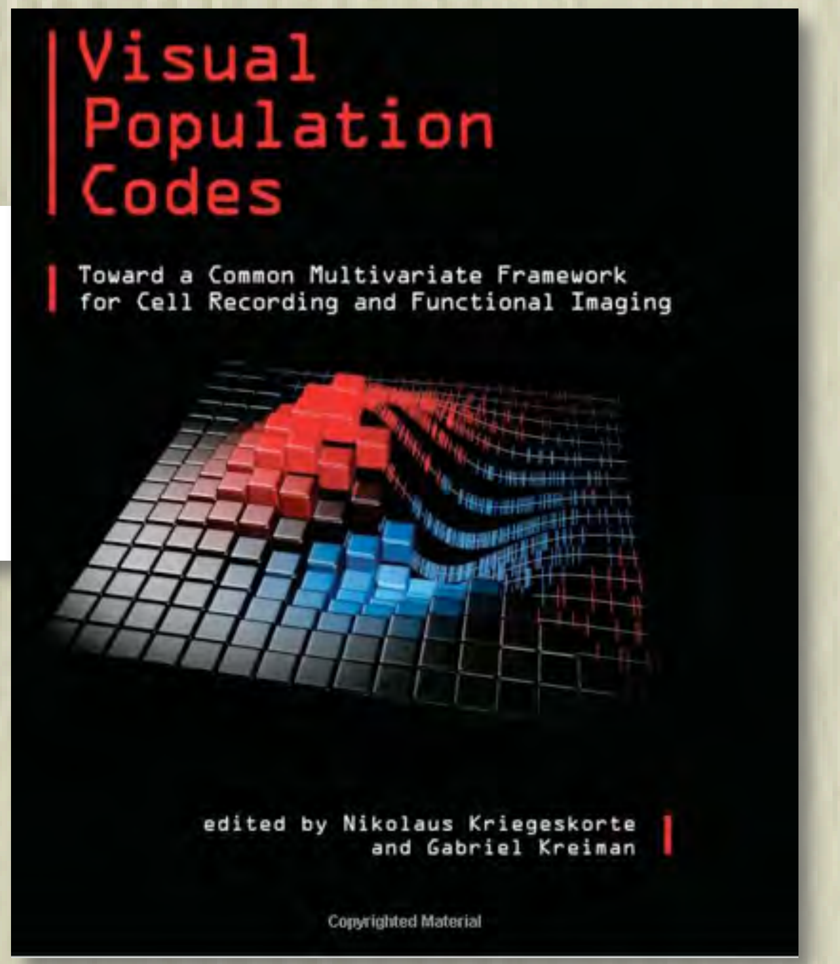
**What is a grandmother cell? And how would you know if you found one?**

Jeffrey S. Bowers*

*Connection Science*
Vol. 23, No. 2, June 2011, 91–95

**1 Grandmother Cells and Distributed Representations**

**Simon J. Thorpe**




**Neural Coding: Non-Local but Explicit and Conceptual**
**Peter Földiák**
**Current Biology Vol 19 No 19**

Cells

| Stimuli | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $c_6$ | $c_7$ | $c_8$ | $c_9$ | $c_{10}$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $S_1$ | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | .2 |
| $S_2$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | .2 |
| $S_3$ | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | .2 |
| $S_4$ | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | .1 local |
| $S_5$ | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | .4 |
| $S_6$ | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | .2 |
| $S_7$ | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | .2 sparse |
| $S_8$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | .2 |
| $S_9$ | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | .5 dense |
| $S_{10}$ | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | .2 |
| | .3 | .1 | .3 | .1 | .4 | .2 | .1 | .2 | .5 | .2 | .24 |
| | | grandm. | | | | narrow | | | broad | | |

Current Biology

- Do Grandmother Cells Exist?

# Local vs Distributed Coding


Simon J. THORPE Intellectica, 1989/2, 8, 3-40

LOCAL vs DISTRIBUTED CODING

- Very few proponents of Local coding
  - Jerzy Konorski (1967)
    - “gnostic neurons”
  - Horace Barlow (1972)
    - “cardinal cells”
  - Alberta Gilinsky (1984)
    - “cognons”

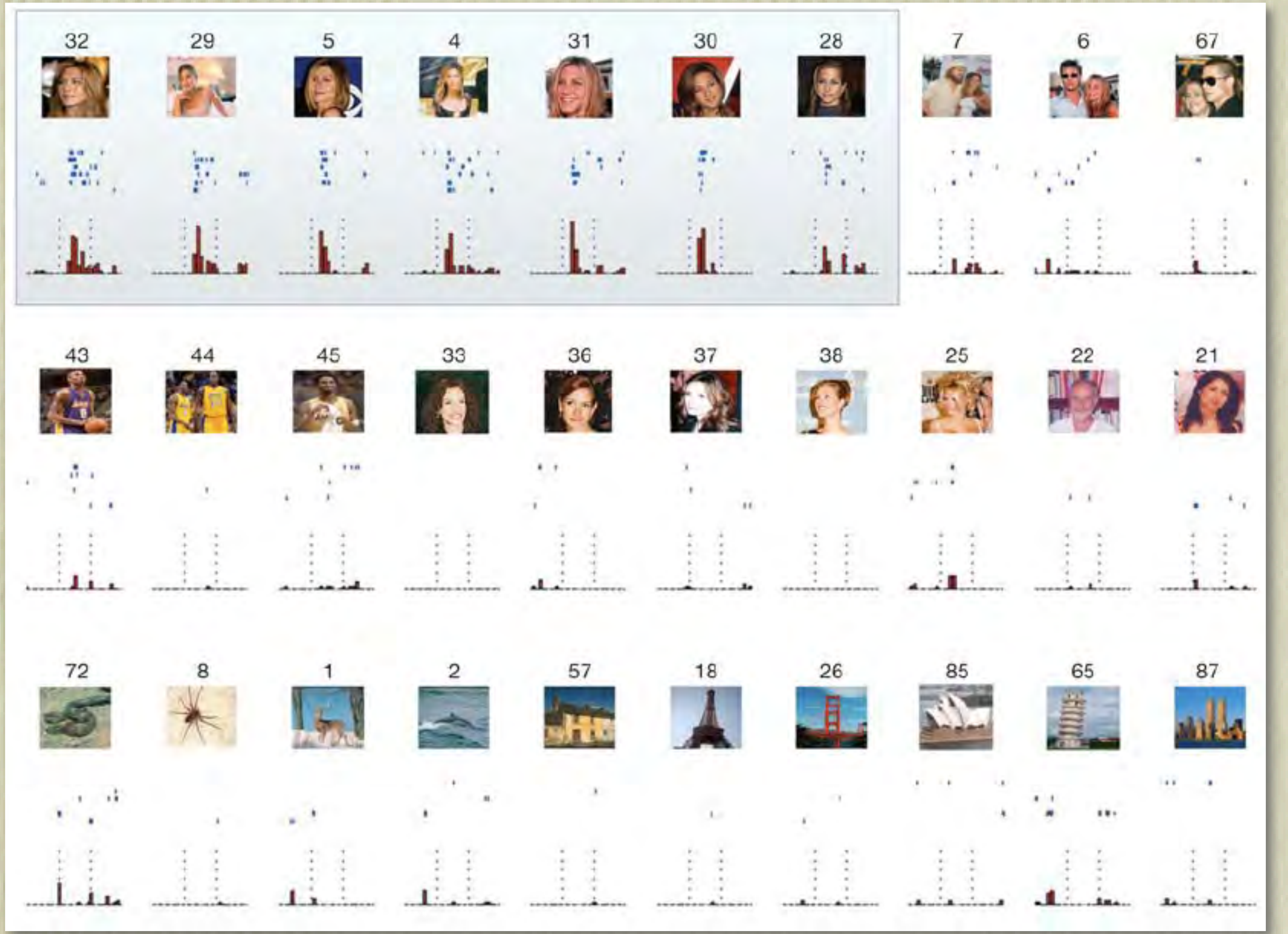
- Criticisms of localist coding
  - There are not enough neurons in the brain
  - Coding by single cells is too risky - not enough redundancy
  - No-one has ever found a “Grandmother cell”
  - Individual neurons are too unreliable
  - You would lose the advantages of distributed coding
  - No-one knows how to make a “Grandmother cell”

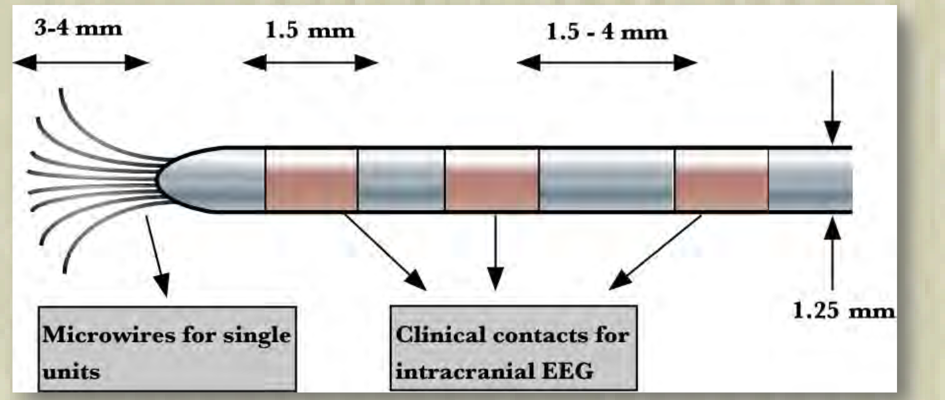
# Grandmother Cells in Man?

**Invariant visual representation by single neurons in the human brain**

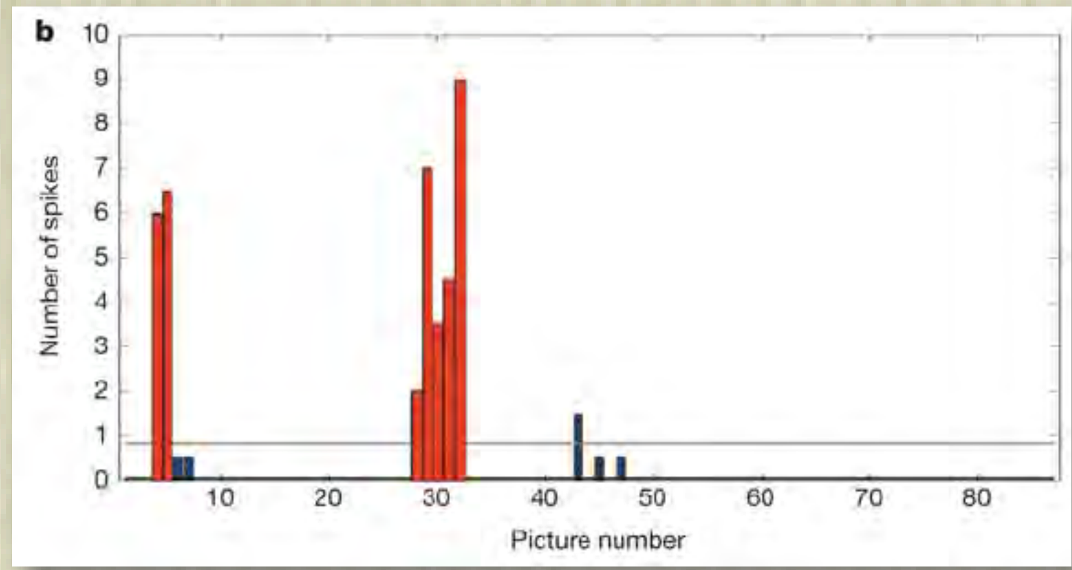
R. Quian Quiroga[1,2]†, L. Reddy[1], G. Kreiman[3], C. Koch[1] & I. Fried[2,4]

NATURE|Vol 435|23 June 2005




A Jennifer Anniston Cell !




- Halle Berry
- The Taj Mahal
- Bill Clinton
- Saddam Hussein
- The Simpsons
- The patient's brother
- Members of the research team
- ...

# Neocortical Dark Matter?

- Question
  - What is the true distribution of firing rates in the cortex?
  - Are there neurons that never fire for very long periods of time?
- Problem
  - Nearly all single unit neurophysiological studies are biased towards neurons that have spontaneous activity
  - The number of neurons recorded seems far lower than would be expected
    - Only 5-15 neurons recorded per descent (out of hundreds)

PROCEEDINGS OF THE IEEE, VOL. 56, NO. 6, JUNE 1968

The Electrical Properties of Metal Microelectrodes

DAVID A. ROBINSON, MEMBER, IEEE

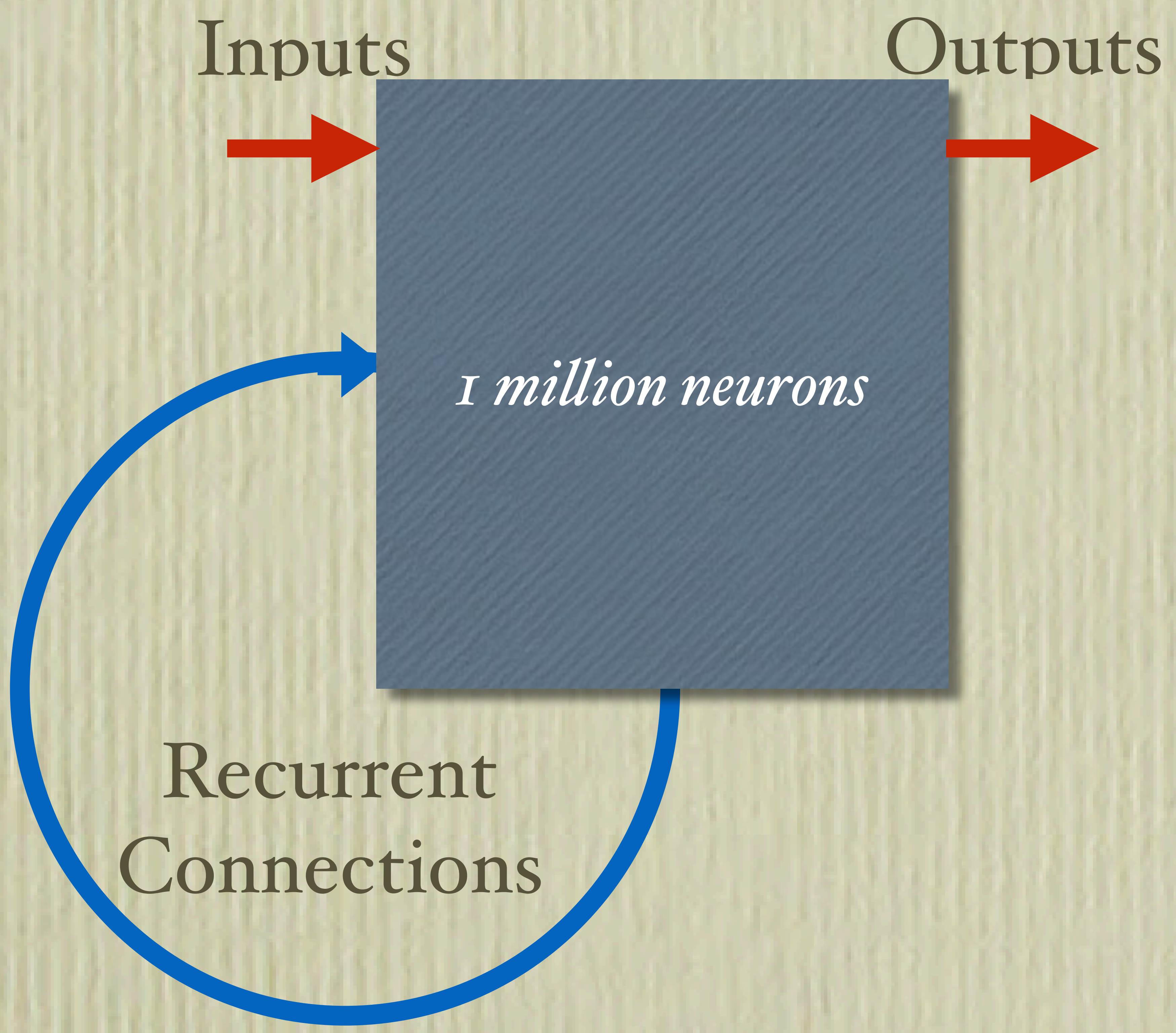
in a 2-mm electrode track the tip should record from 70 to 234 cells, depending on cell density. In actual practice, in gray matter, one sees only a tiny fraction of these cells, and why this is so is a very disturbing question to users of microelectrodes.

- The Brain could be extremely sparse!

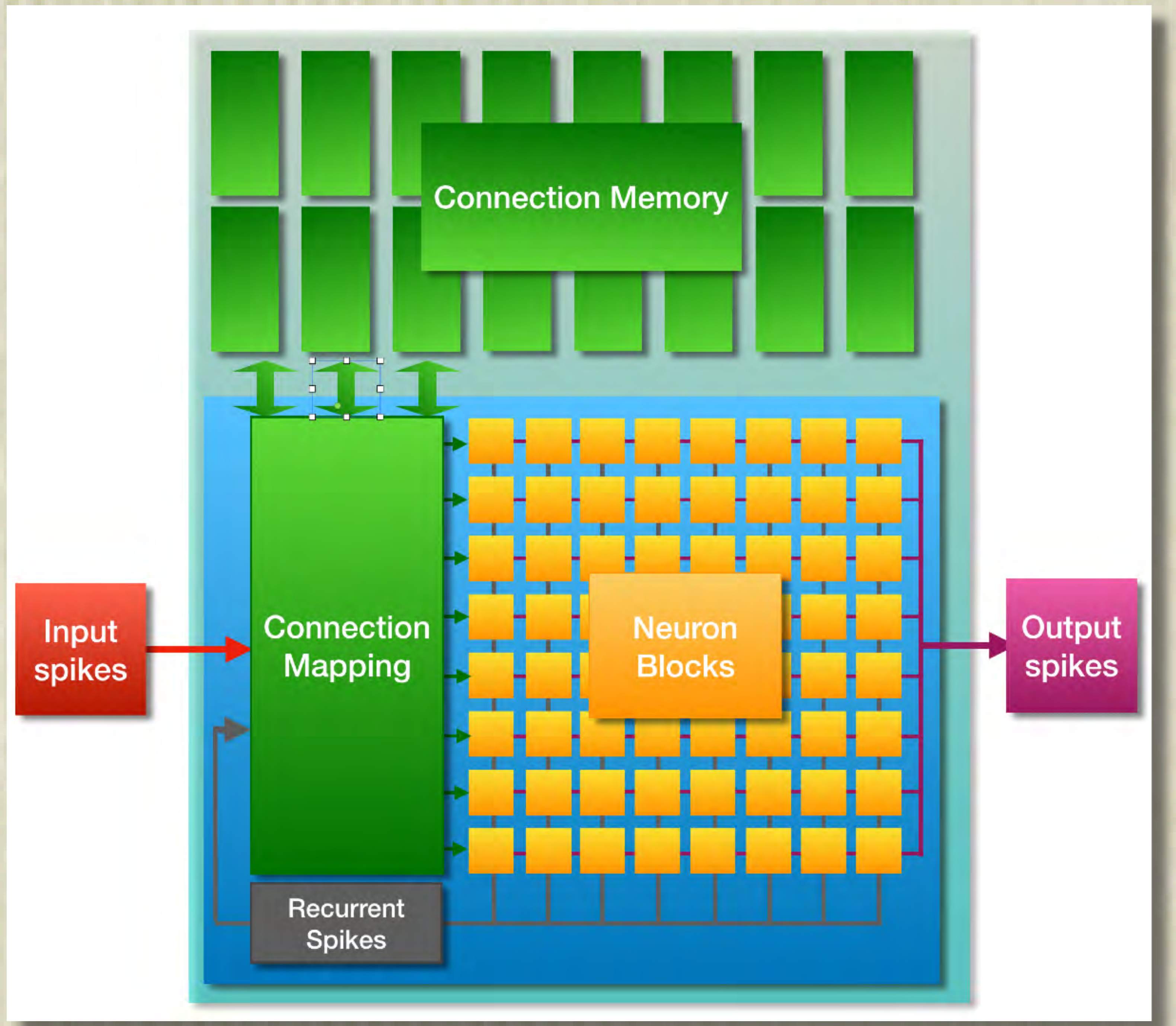
# The Importance of Representation

- Grandmother Cells are generally dismissed
  - But there is good evidence for them
    - Single unit recording
    - Analysis of Deep Learning networks
- Neocortical Dark Matter
  - Only very small percentage of neurons actually fires
- The secret of maintaining memories over the entire lifetime
  - We still have the neurons we had when we were infants
  - Grandmother cells and Dark Matter mean that memories are not overwritten
- Importance for Neural Network Hardware!

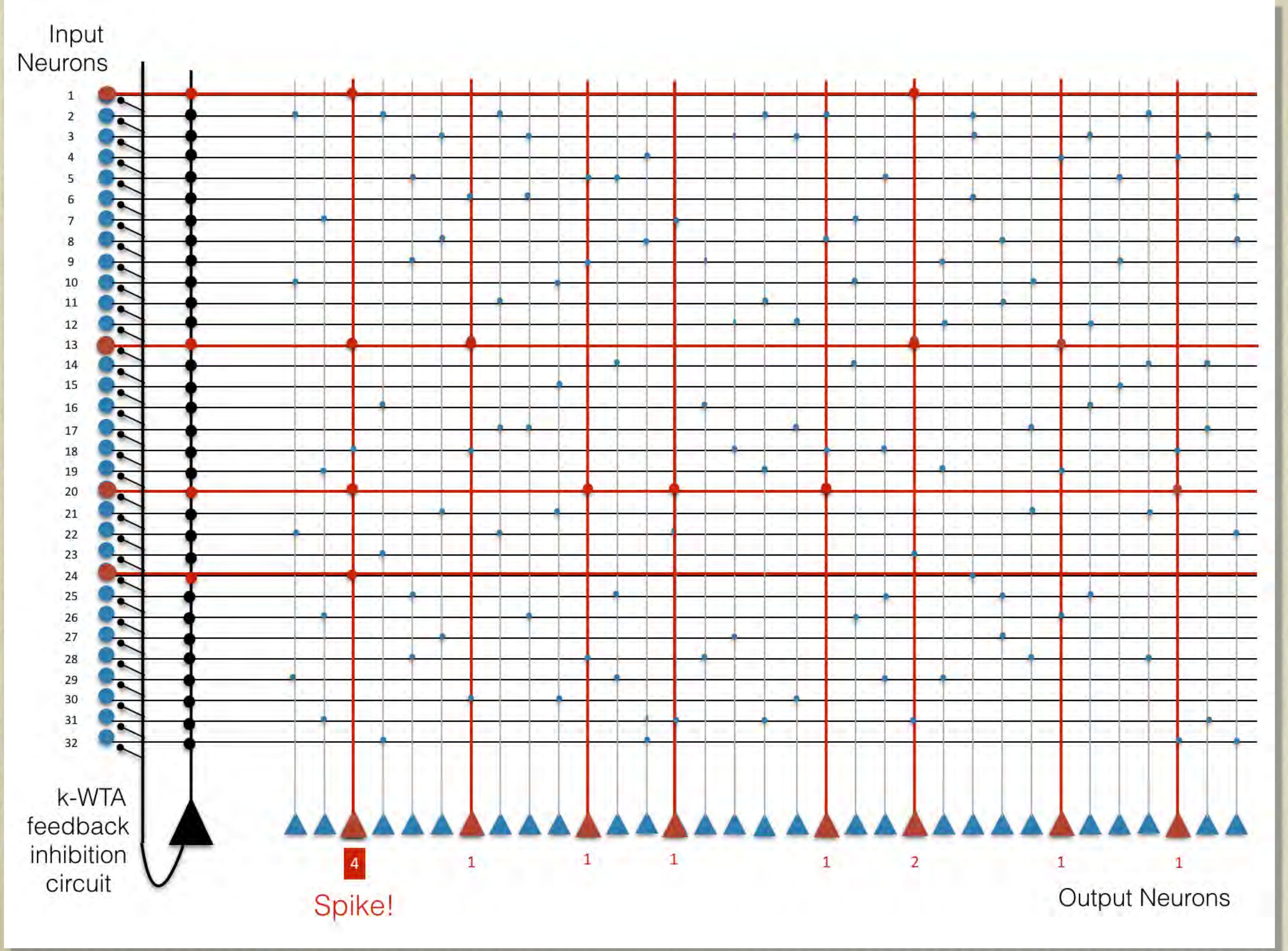
# Computational Costs



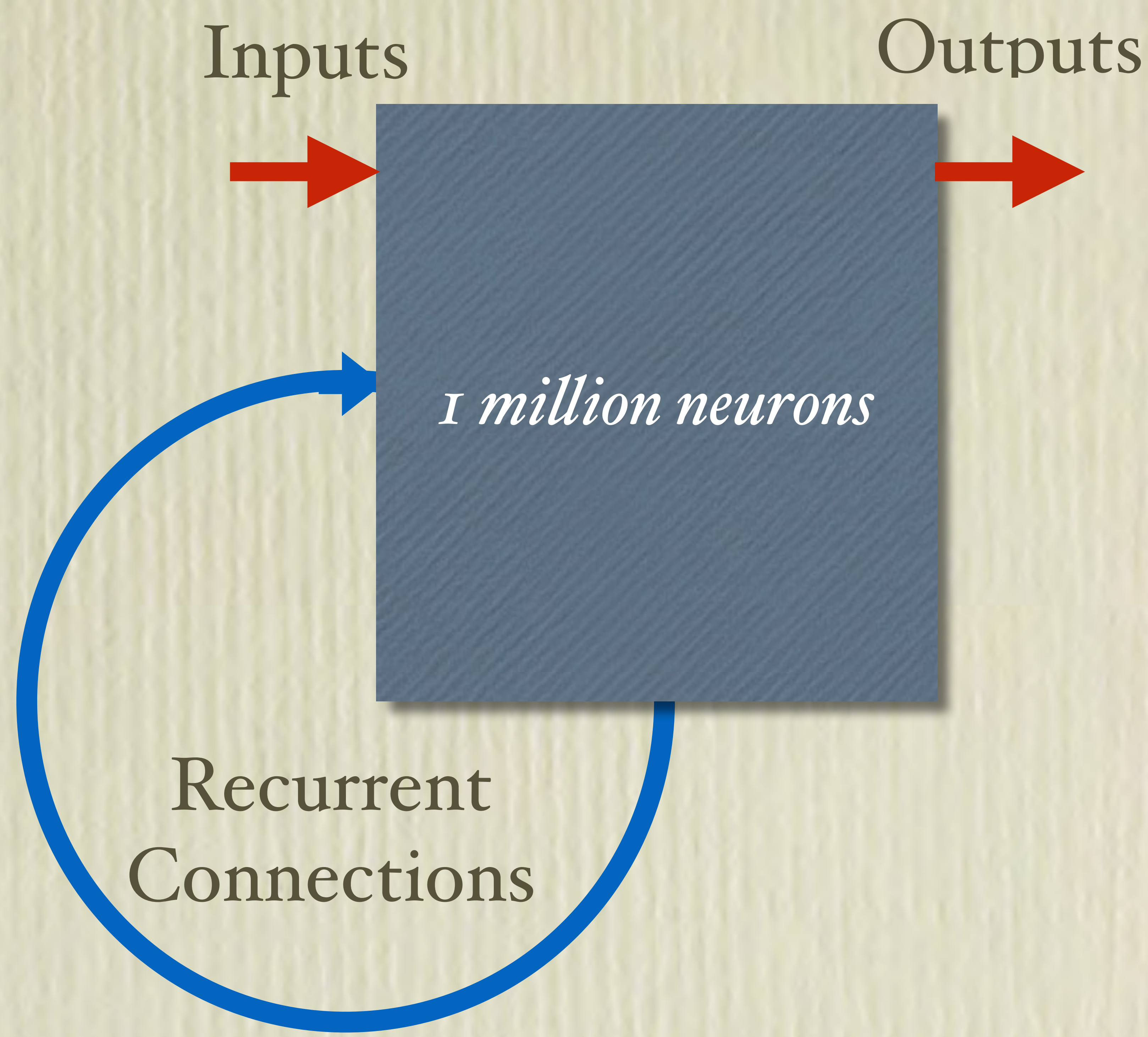

- Using Standard Neurons with floating point numbers
    - $10^{12}$ Floating point operations per clock for the recurrent connections
    - Additional Computations for each Input channel
- With Grandmother Cells and Dark Matter
    - Only the recurrent connections for the active neurons need to be calculated
    - Activity could be 0.1% or less!



# Packet Based Crossbar Processing

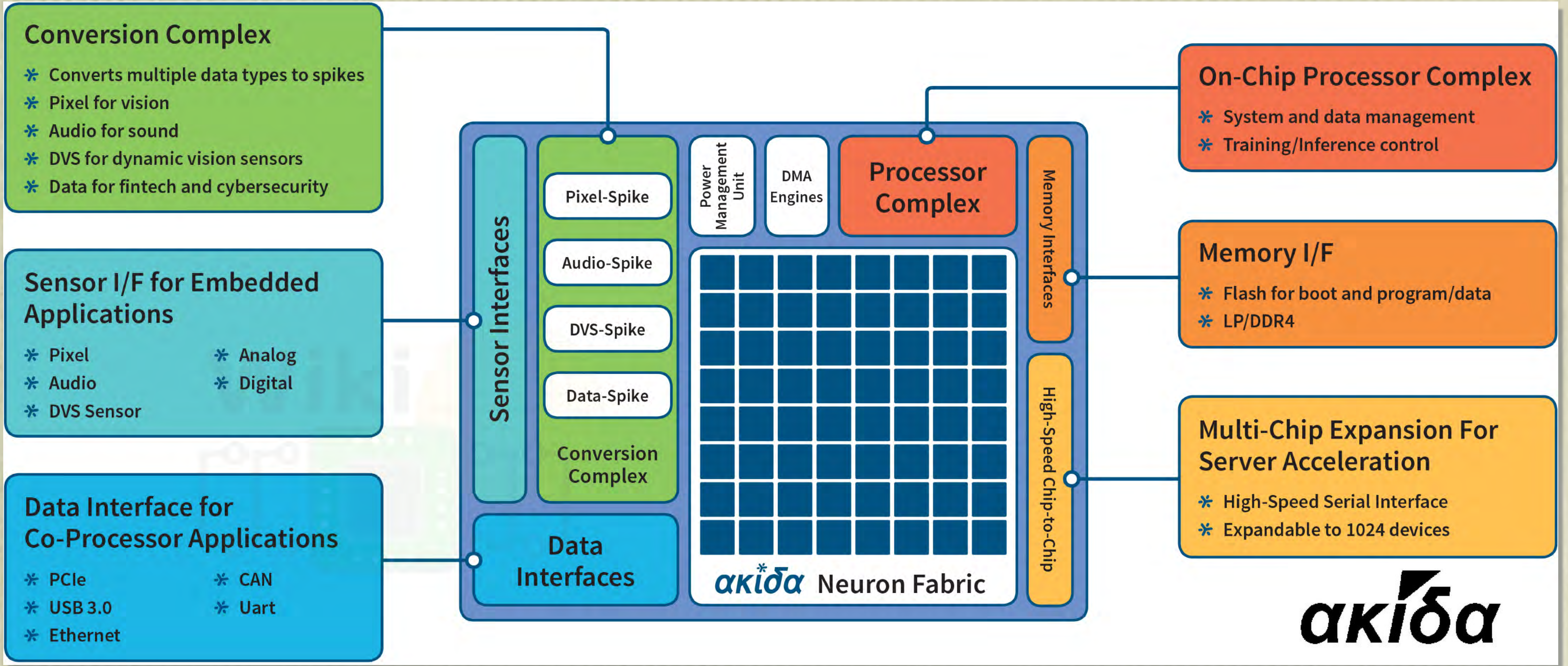



- k-WTA circuit sets the packet size
- Allows Spike by Spike processing
- Can be used with very large networks




- Ultra Sparse Representation!

# BrainChip's AKIDA chip




- 1.2 million neurons
- 10 billion synapses
- On chip JAST learning
- 50 mm2
- Production cost : $10-15
- Available before end 2019

# Final Thoughts

- Deep Learning Networks are the state of the art
  - Problem
    - Very computationally demanding
  - Inspiration from biology
    - Use spikes!
    - Use the order of firing to encode information
    - Use circuits to control the percentage of neurons that fire
    - Grandmother Cells and Neocortical Dark Matter
    - Spikes also important for learning
      - another story!
  - Development of Spiking Hardware
    - Ultra Low Power devices

Don't believe the Neuroscientists!

- Neurons don't always use rate coding
- They send pulses not floating point numbers
- Neurons can be extremely selective
- Neocortical Dark Matter may be the reality

- Ultra Sparse Coding!!