# From natural scene statistics to models of neural coding and representation (part 1)

**Bruno A. Olshausen**
Helen Wills Neuroscience Institute, School of Optometry
and Redwood Center for Theoretical Neuroscience
UC Berkeley

REDWOOD CENTER
for Theoretical Neuroscience

**Review article:**

*What Natural Scene Statistics Can Tell Us about Cortical Representation*

Bruno A. Olshausen & Michael S. Lewicki

To appear in:
*The New Visual Neurosciences*
Chalupa and Werner, Eds.
MIT Press

# Today's talk

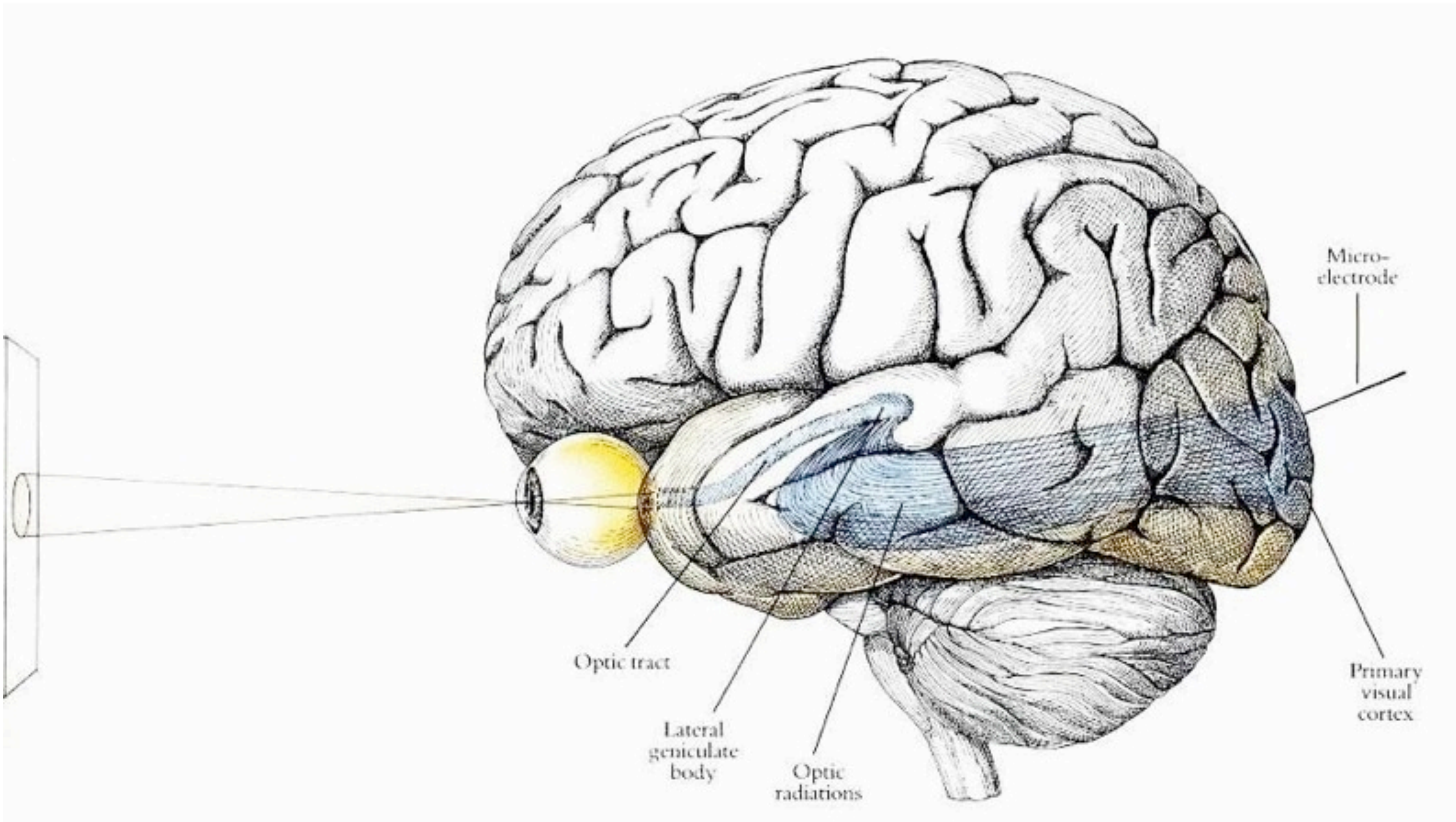
Why natural scene statistics?

- (a bit about biology)

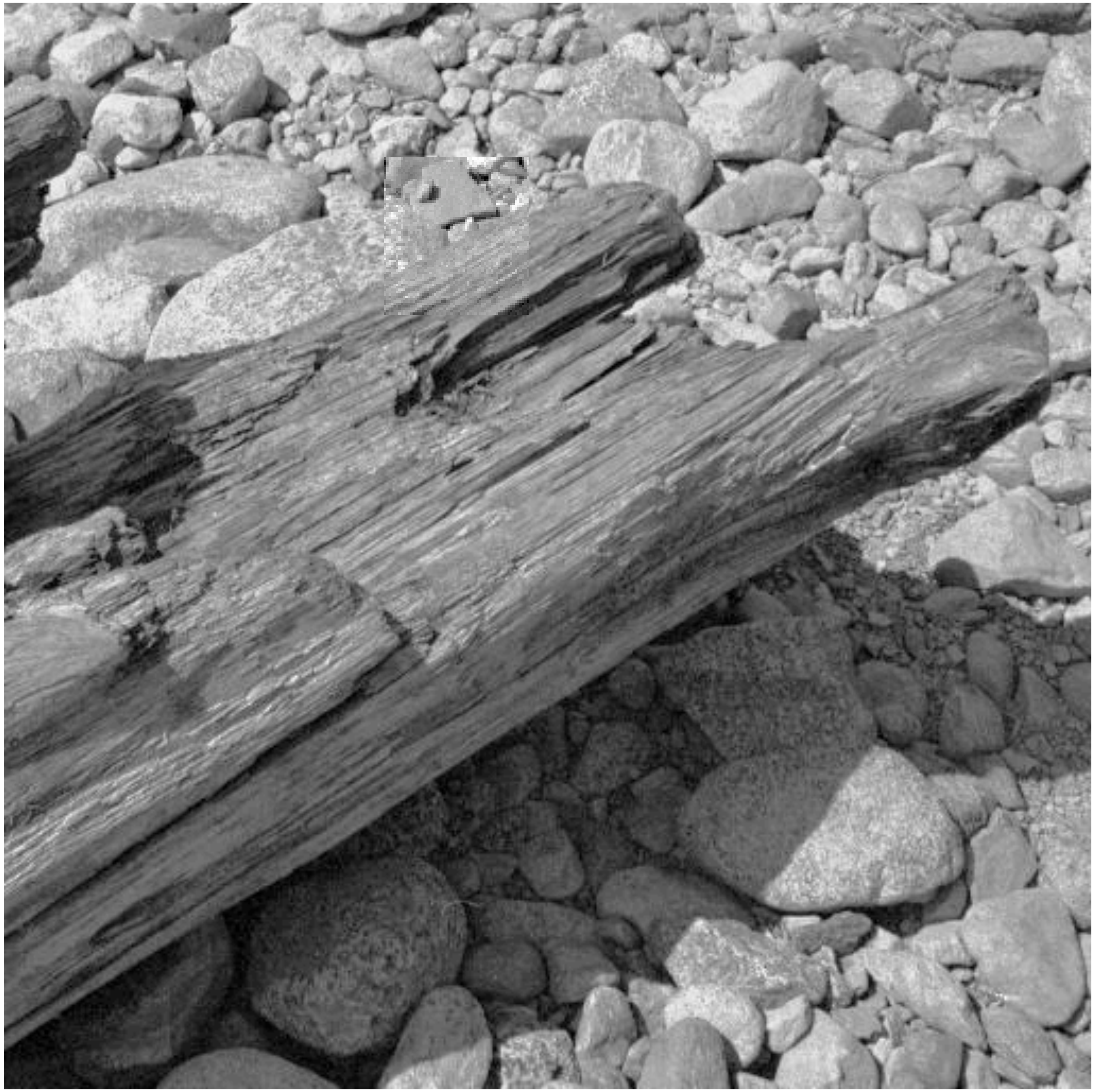
Theory of Redundancy Reduction

Sparse Coding

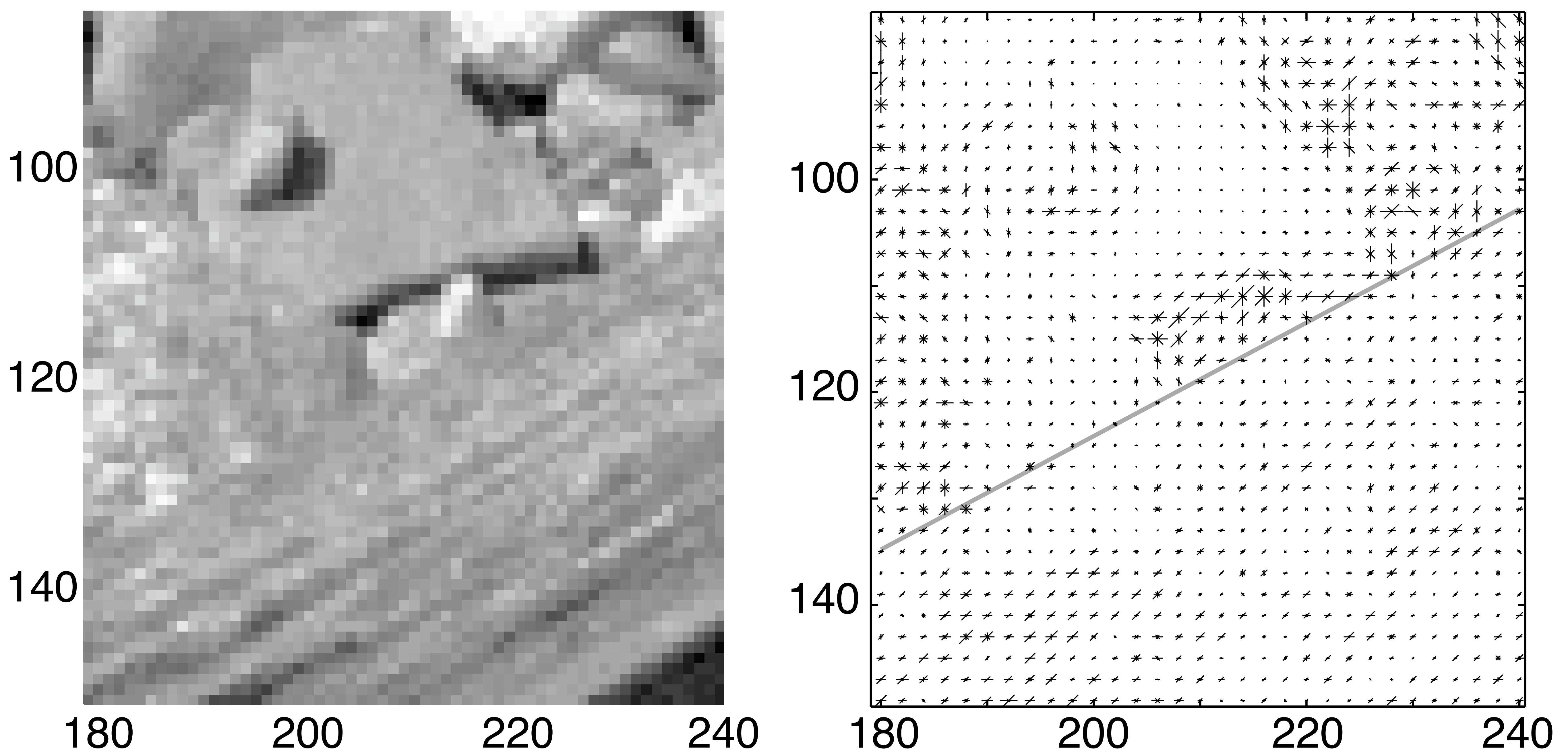
# What are the principles of computation and representation governing this system?



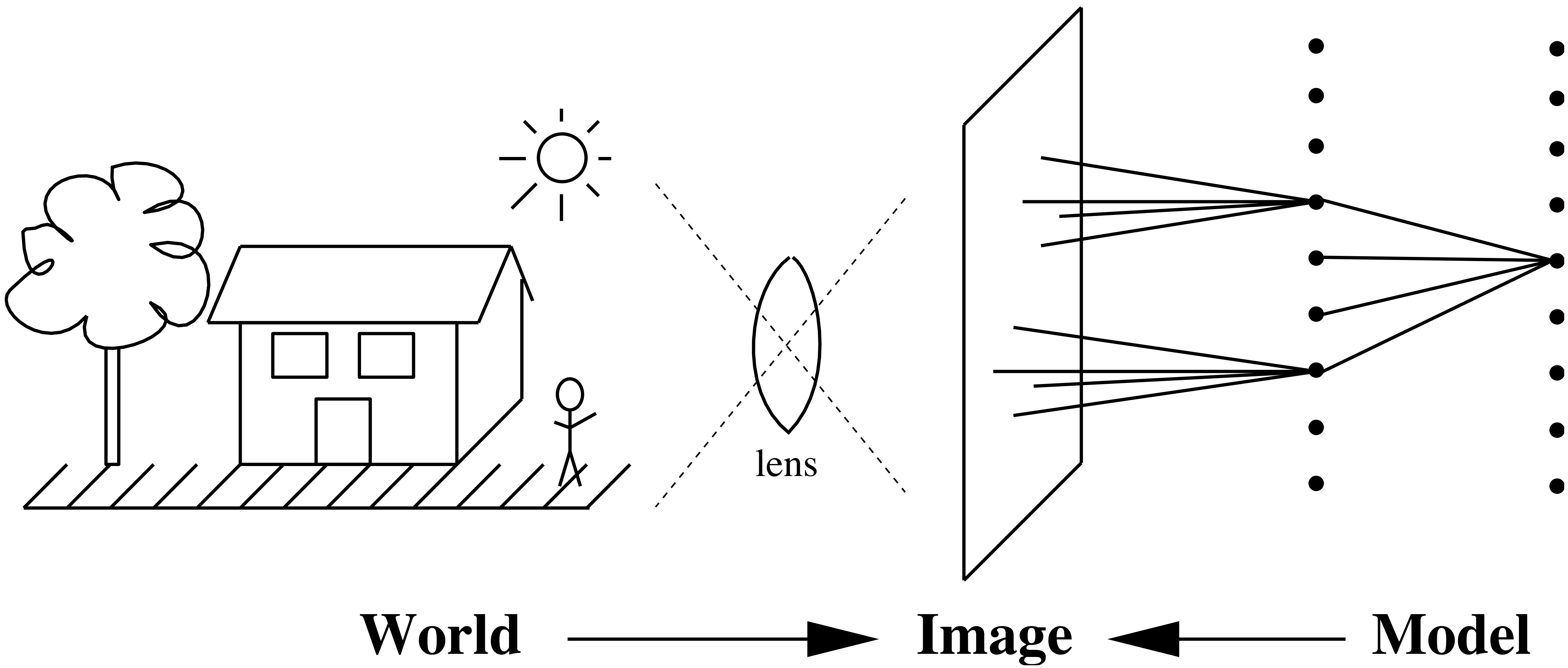
# Natural images are full of ambiguity



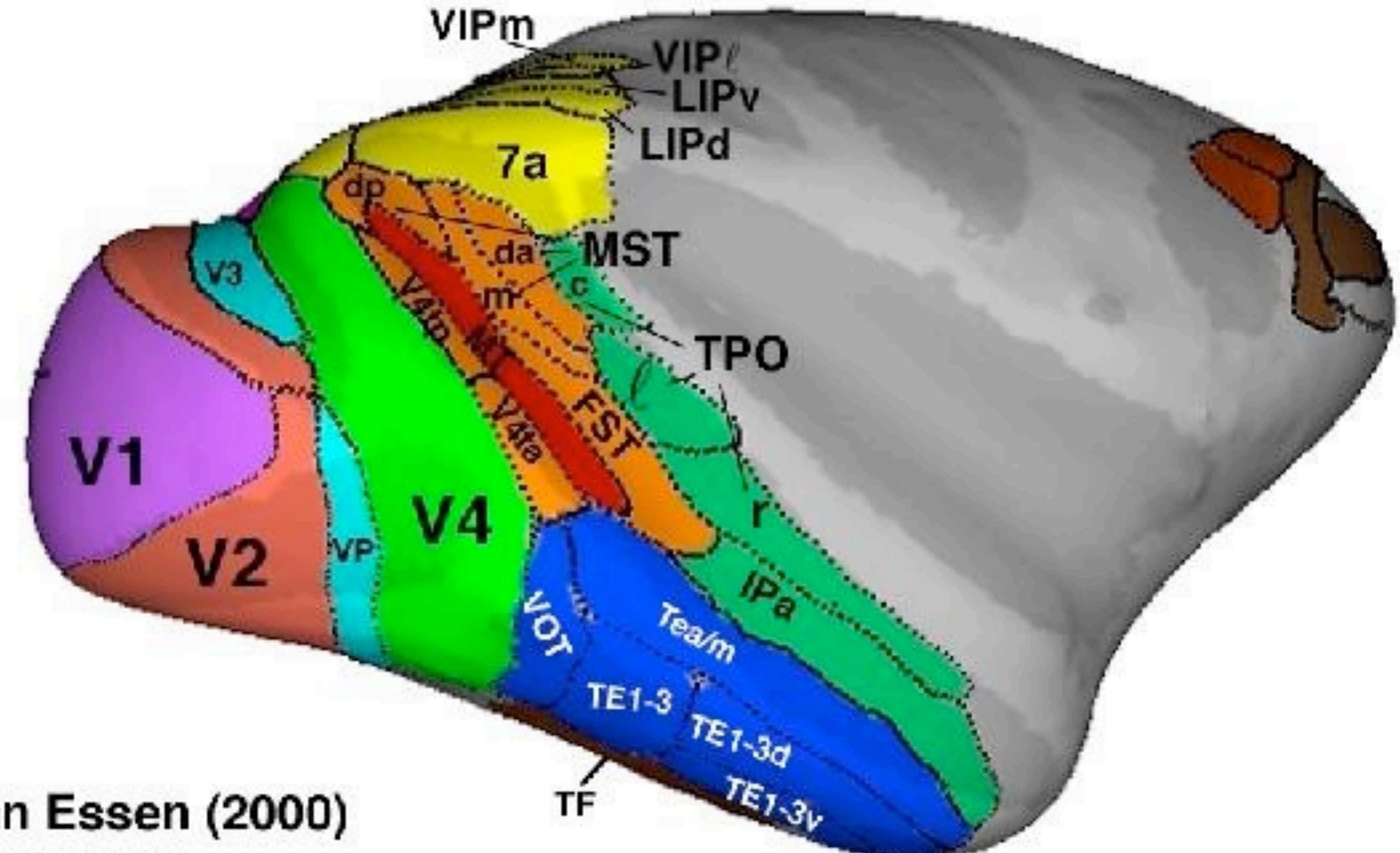
# Natural images are full of ambiguity



# Vision as inference

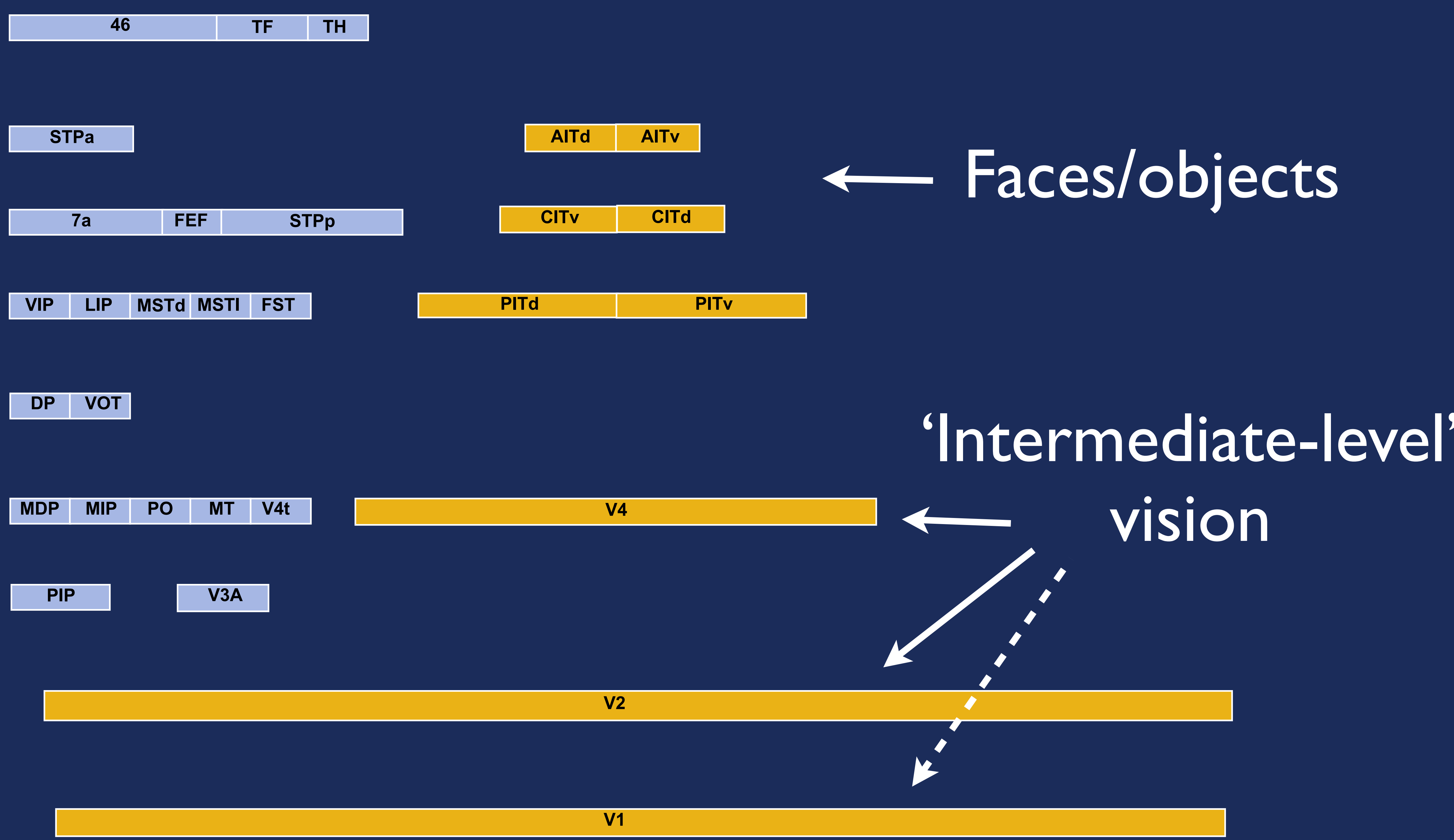


# Visual cortical areas - macaque monkey

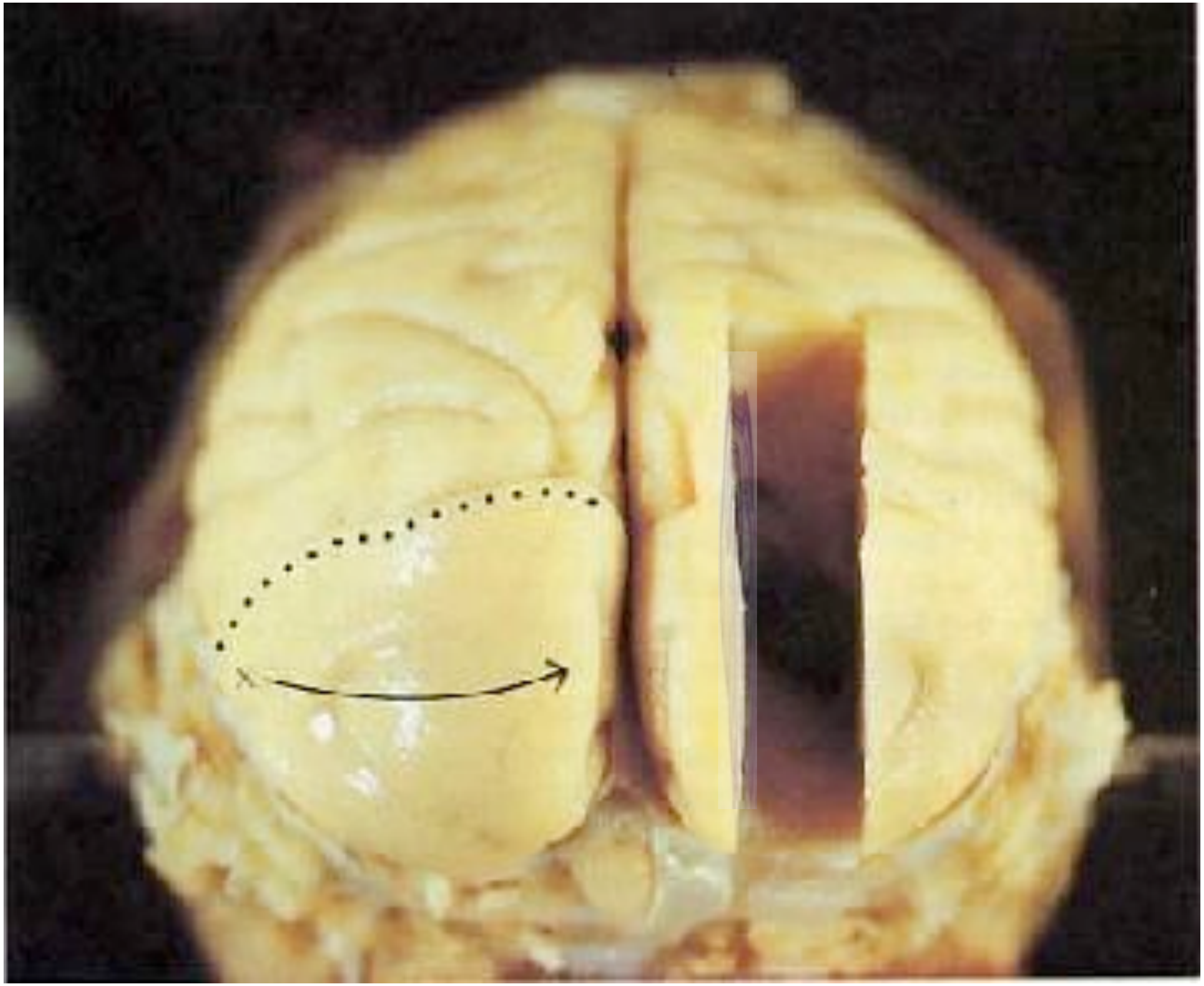


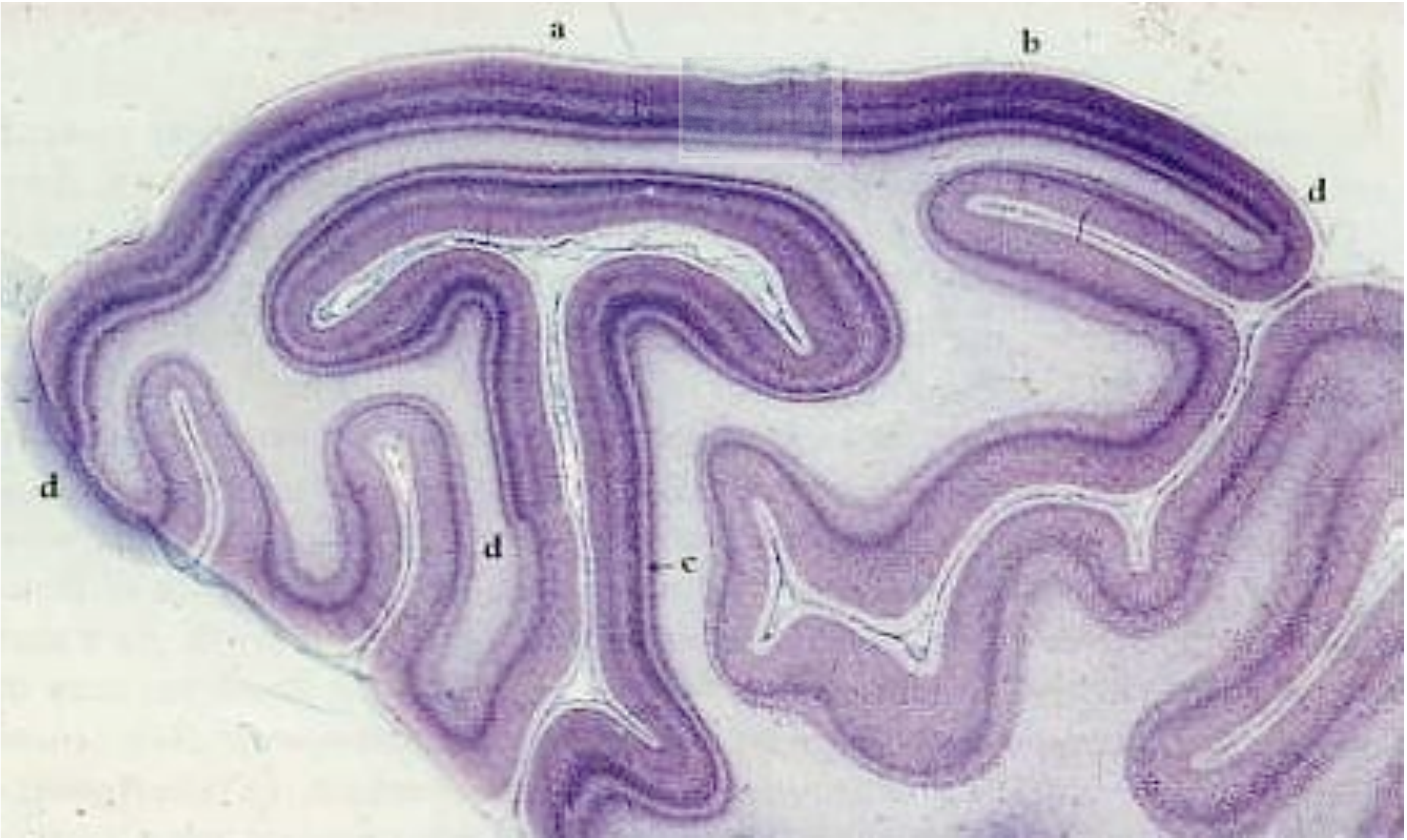

Lewis & Van Essen (2000)
Visual areas

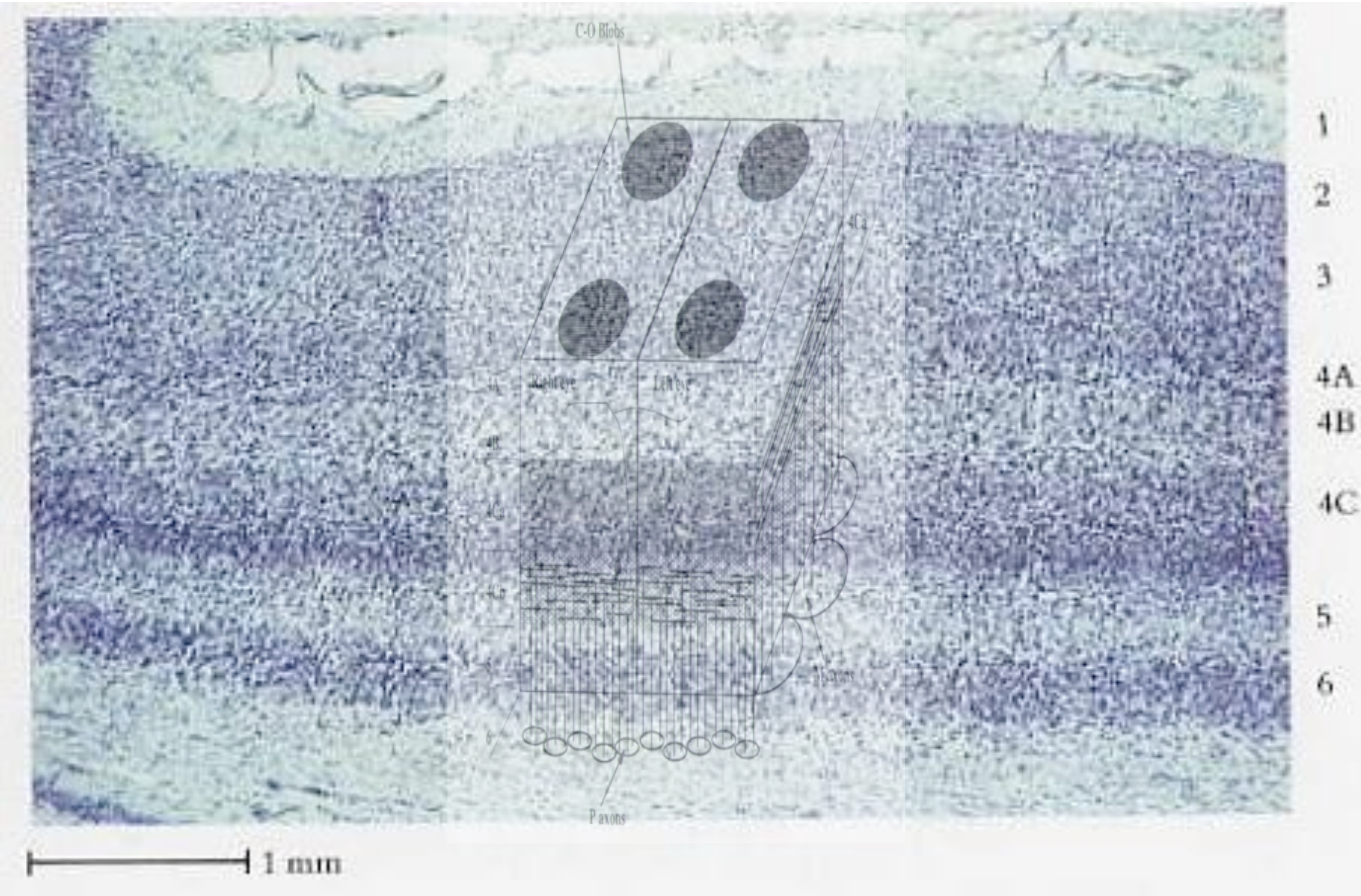
# Visual cortical areas



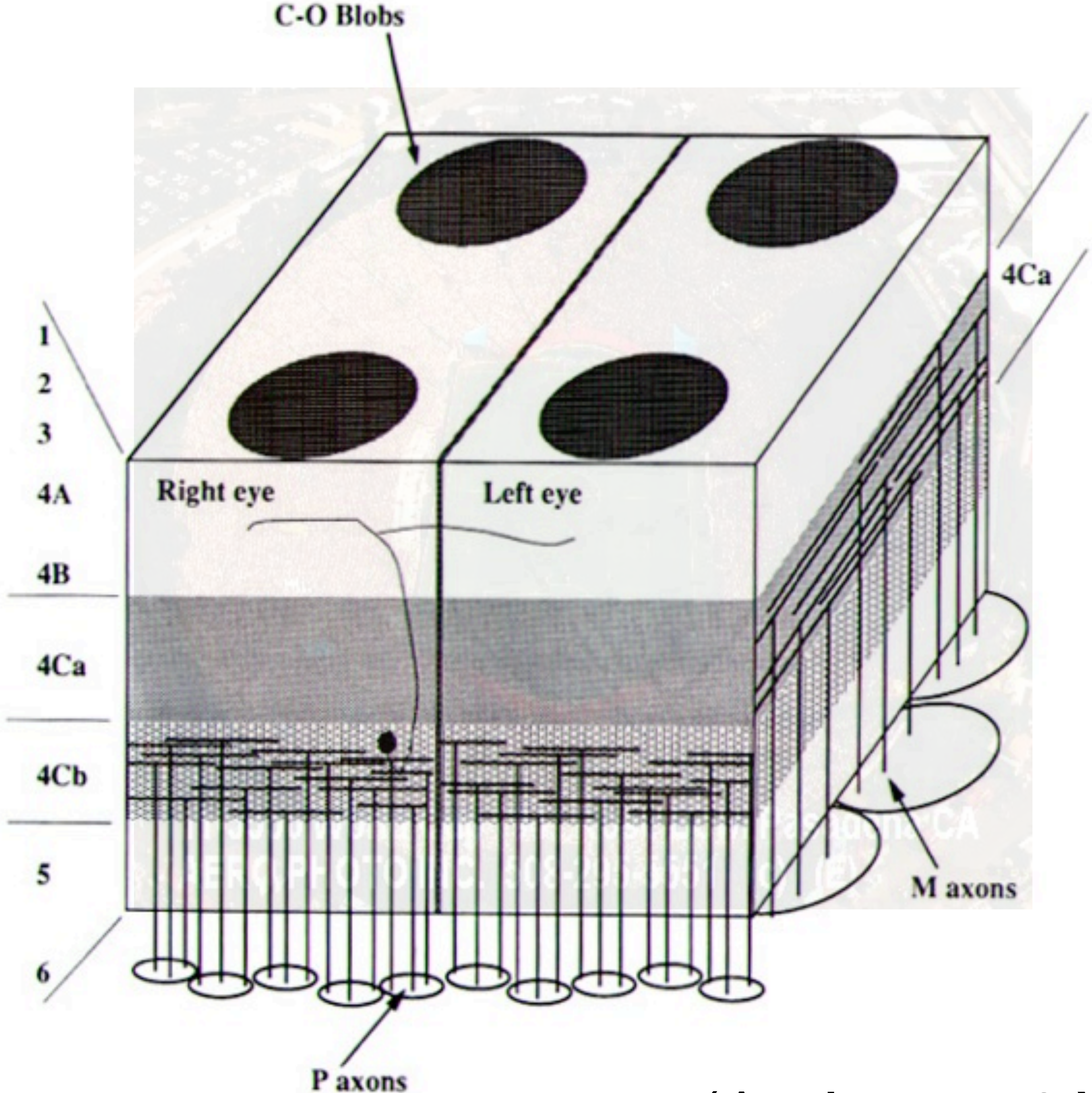

(courtesy of Jeff Hawkins)



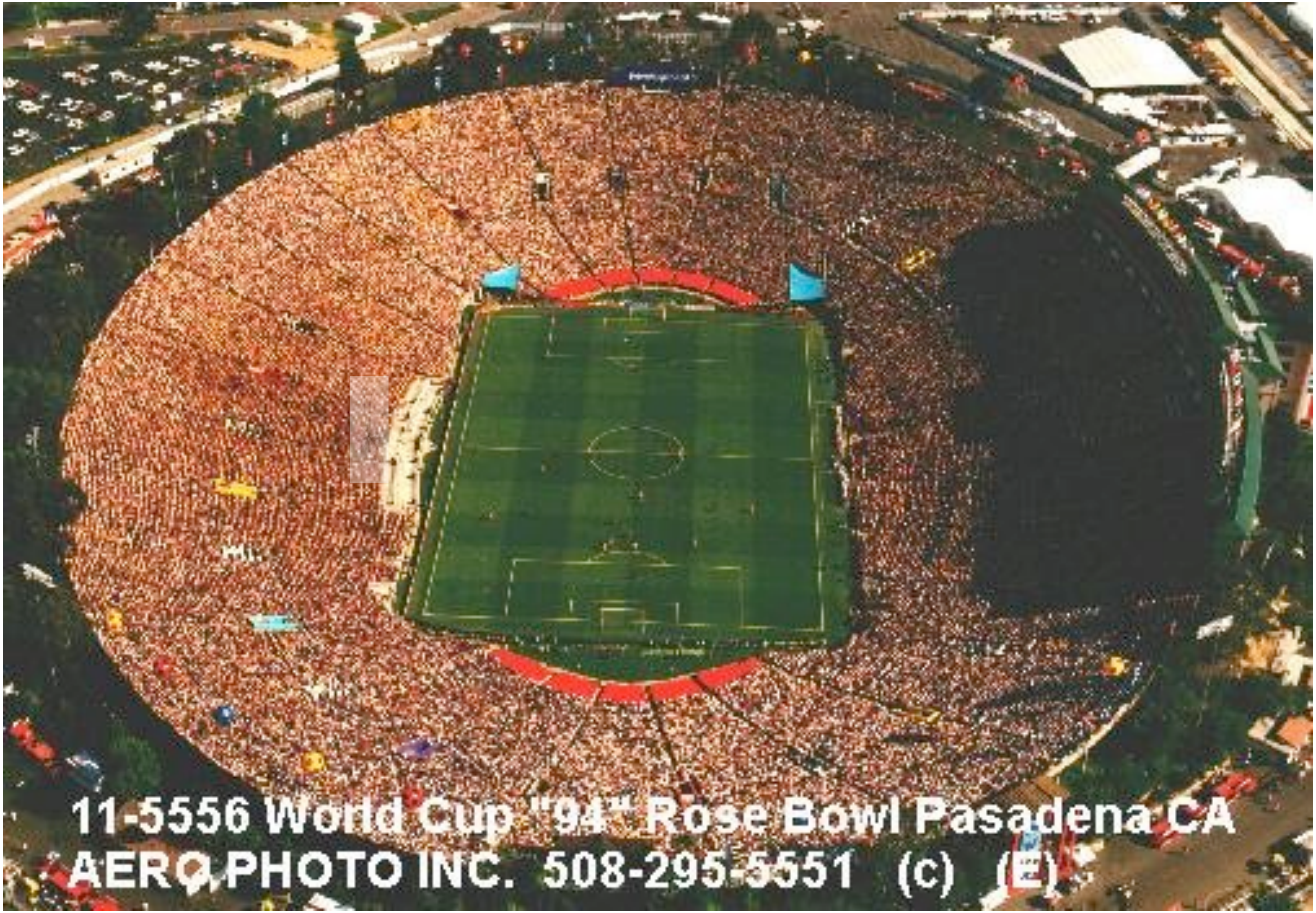

a
b
d
d
d
c

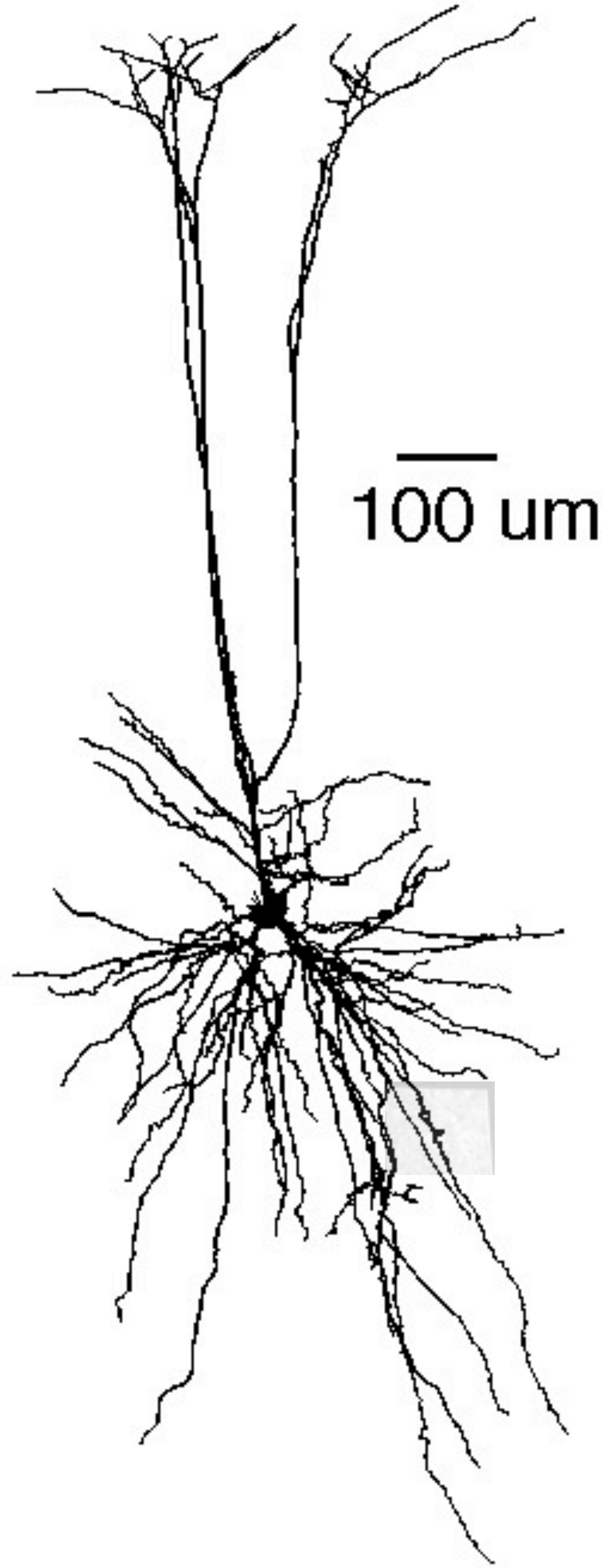

C-O Blobs
1
2
3
4A
4B
4C
5
6
P axons
1 mm

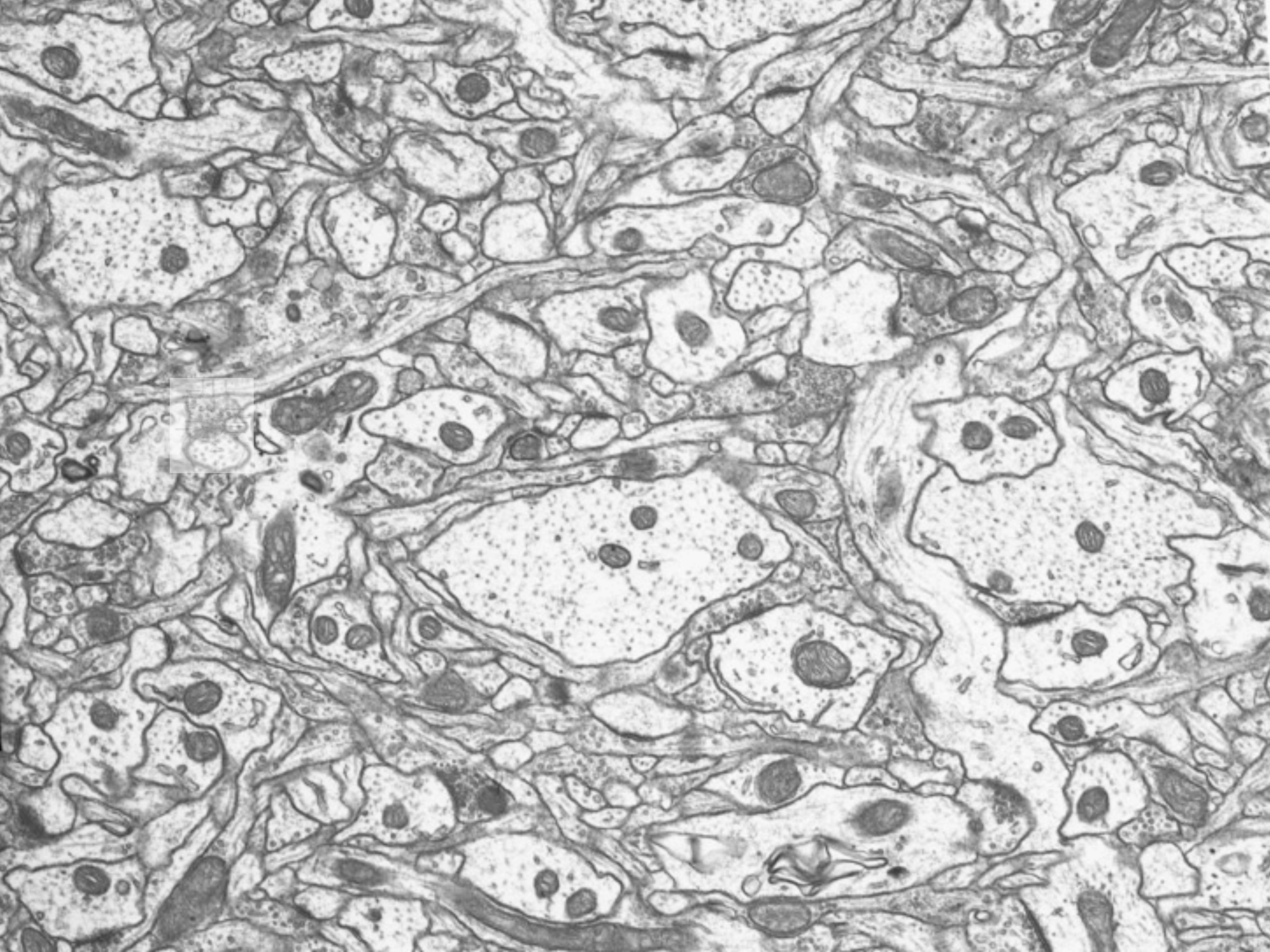
1 mm$^2$ of cortex analyzes ca. 14 x 14 array of retinal sample nodes and contains 100,000 neurons



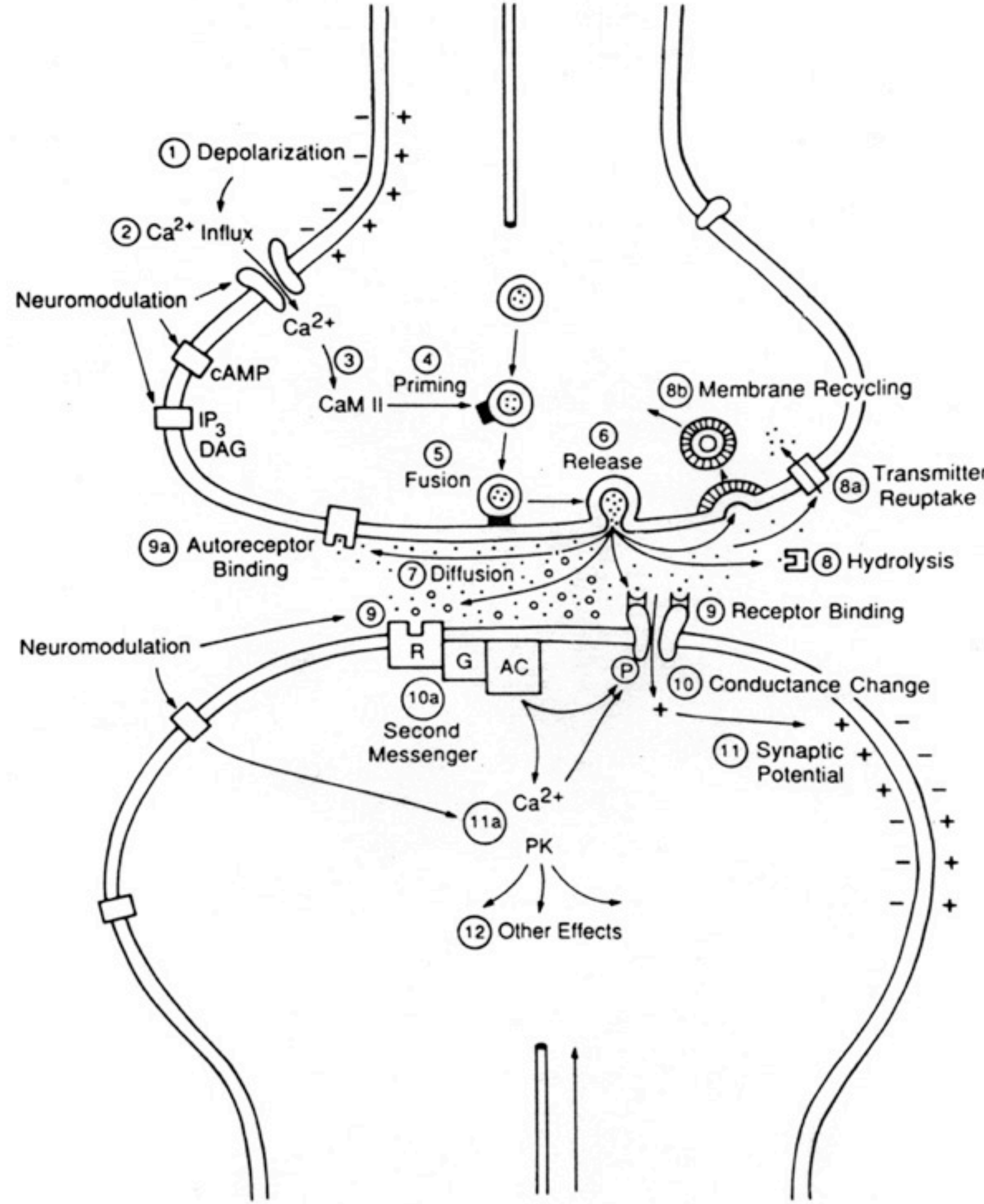

(Anderson & Van Essen, 1995)


11-5556 World Cup "94" Rose Bowl Pasadena CA
AERO PHOTO INC. 508-295-5551 (c) (E)


100 um

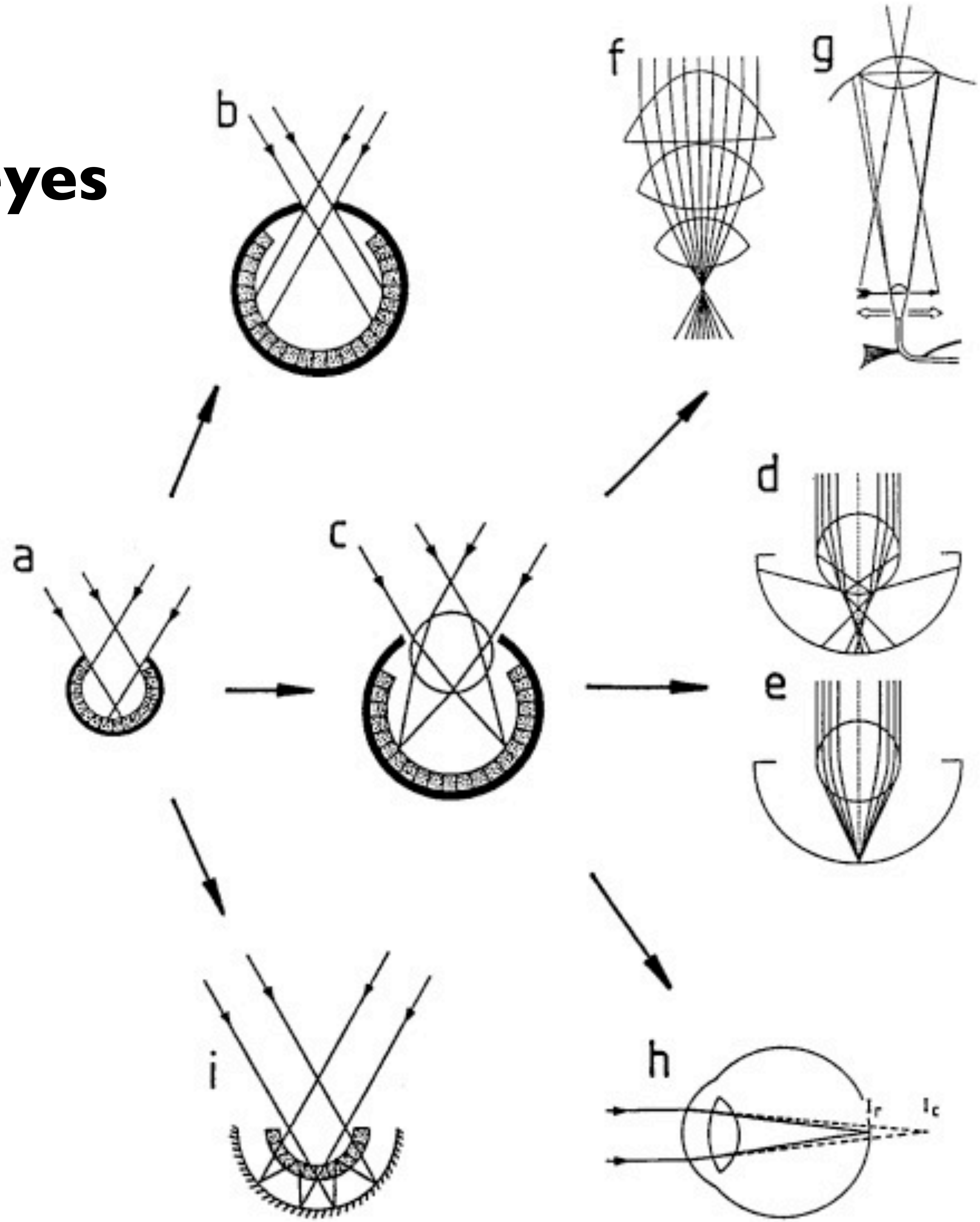


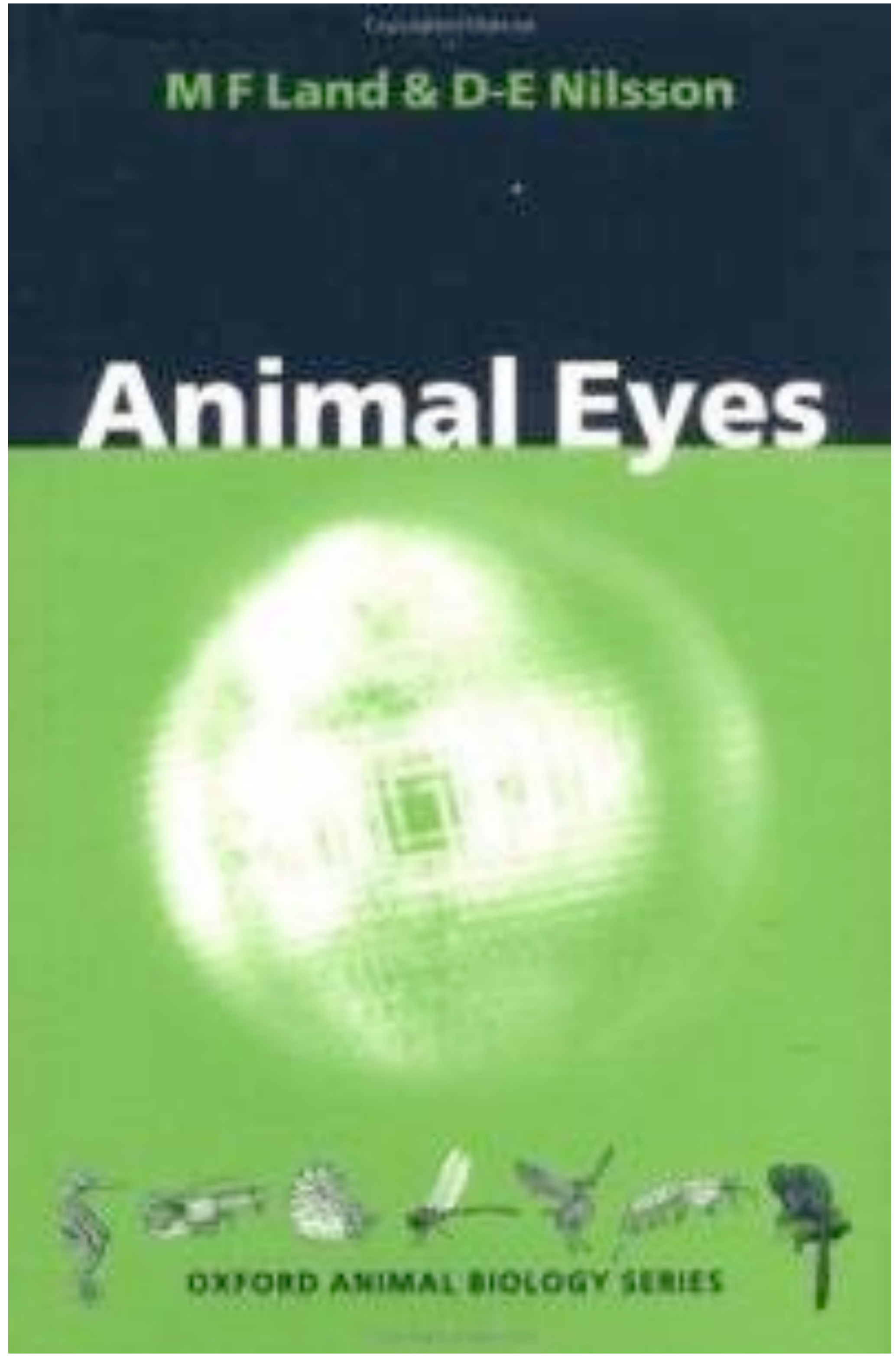
# Anatomy of a synapse



# The evolution of eyes

Land & Fernald (1992)






http://redwood.berkeley.edu/wiki/VS298:_Animal_Eyes

# Efficient Coding

*represent the most relevant visual information with the fewest physical and metabolic resources*

# Theory of ‘Redundancy Reduction’

Attneave (1954) *Some Informational Aspects of Visual Perception*

Barlow (1961) *Possible Principles Underlying the Transformations of Sensory Messages*
- nervous system should reduce redundancy
- makes more efficient use of neural resources
- enables storing information about prior probabilities since $P(\mathbf{x}) = \Pi_i P(x_i)$

➔ “suspicious coincidences”

# From theory to models

Laughlin (1981) - histogram equalization

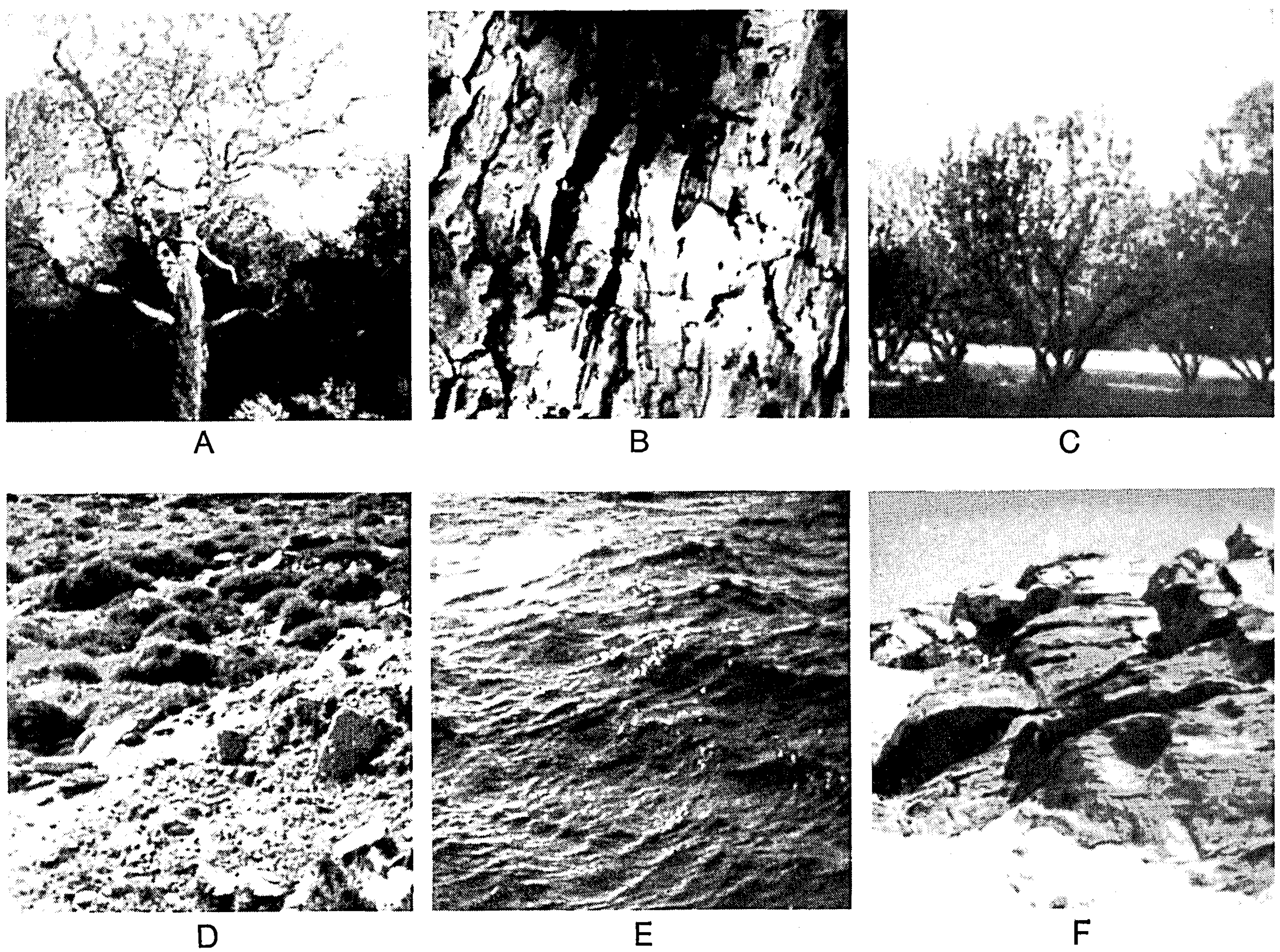
Laughlin, Srinivasan and Dubs (1982) - *Predictive Coding: A Fresh View of Inhibition in the Retina*

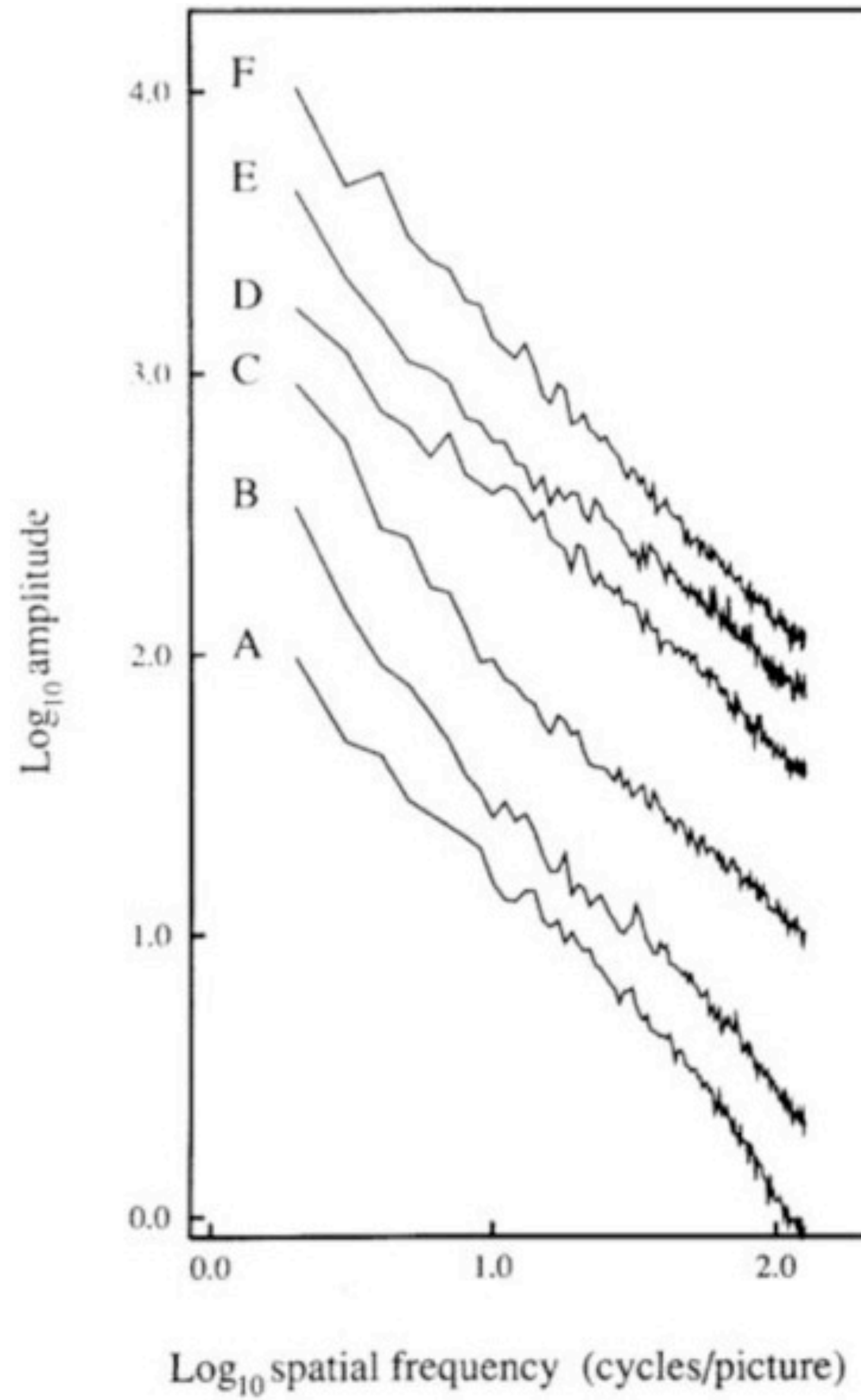
Field (1987) - natural images have $1/f^2$ power spectra

Atick & Redlich (1992); van Hateren (1992; 1993) - whitening

Dan, Atick, and Reid (1996) - LGN whitens natural movies

# Natural scene statistics and visual coding

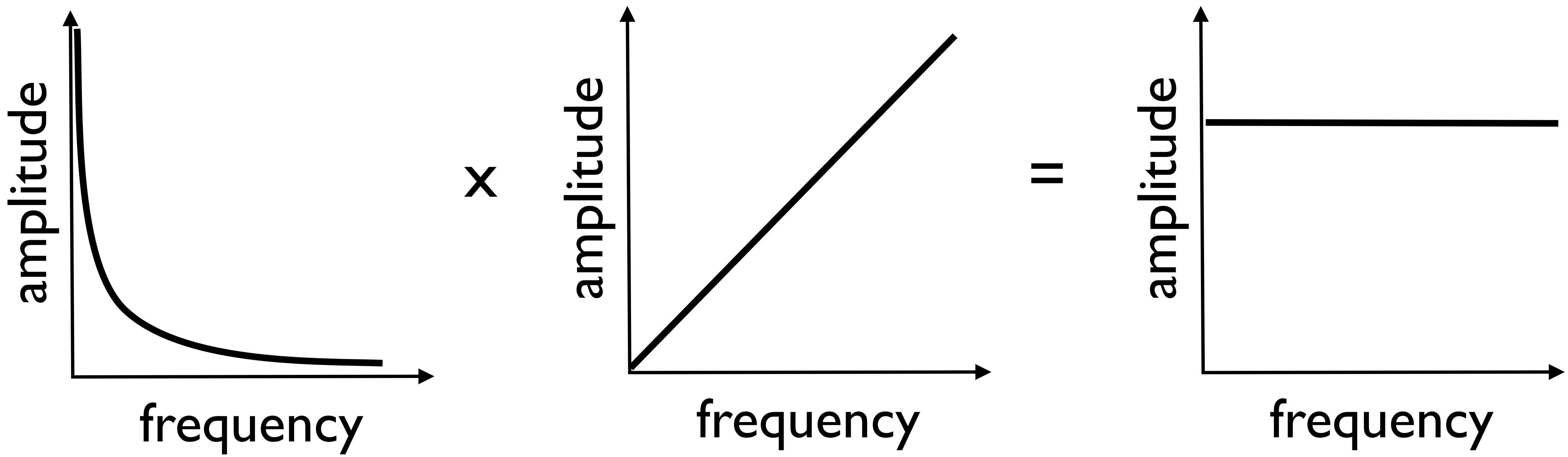



(Field 1987)

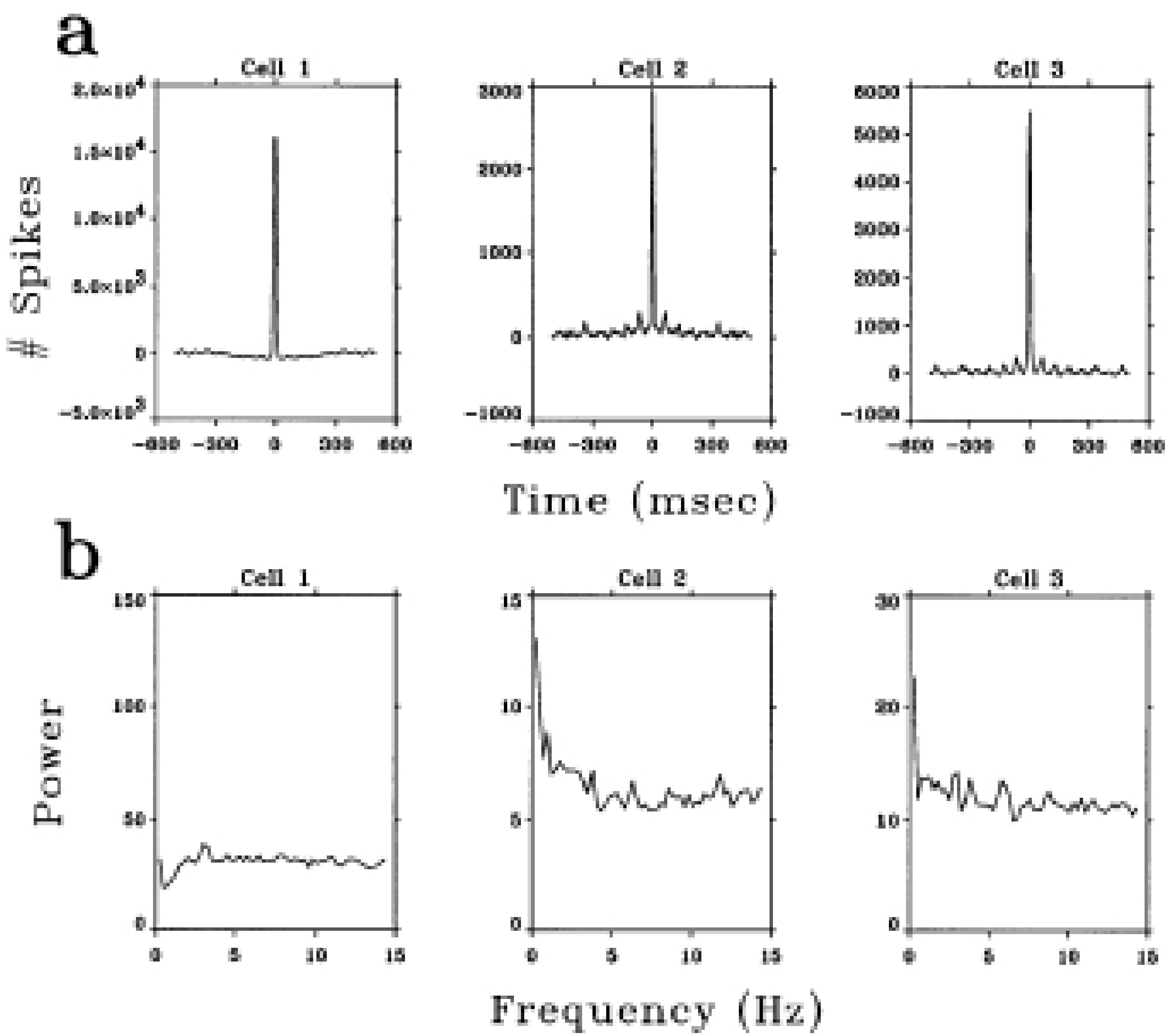
# Whitening (or decorrelation) theory
## (Atick & Redlich, 1992)

1/f image | whitening filter | decorrelated image

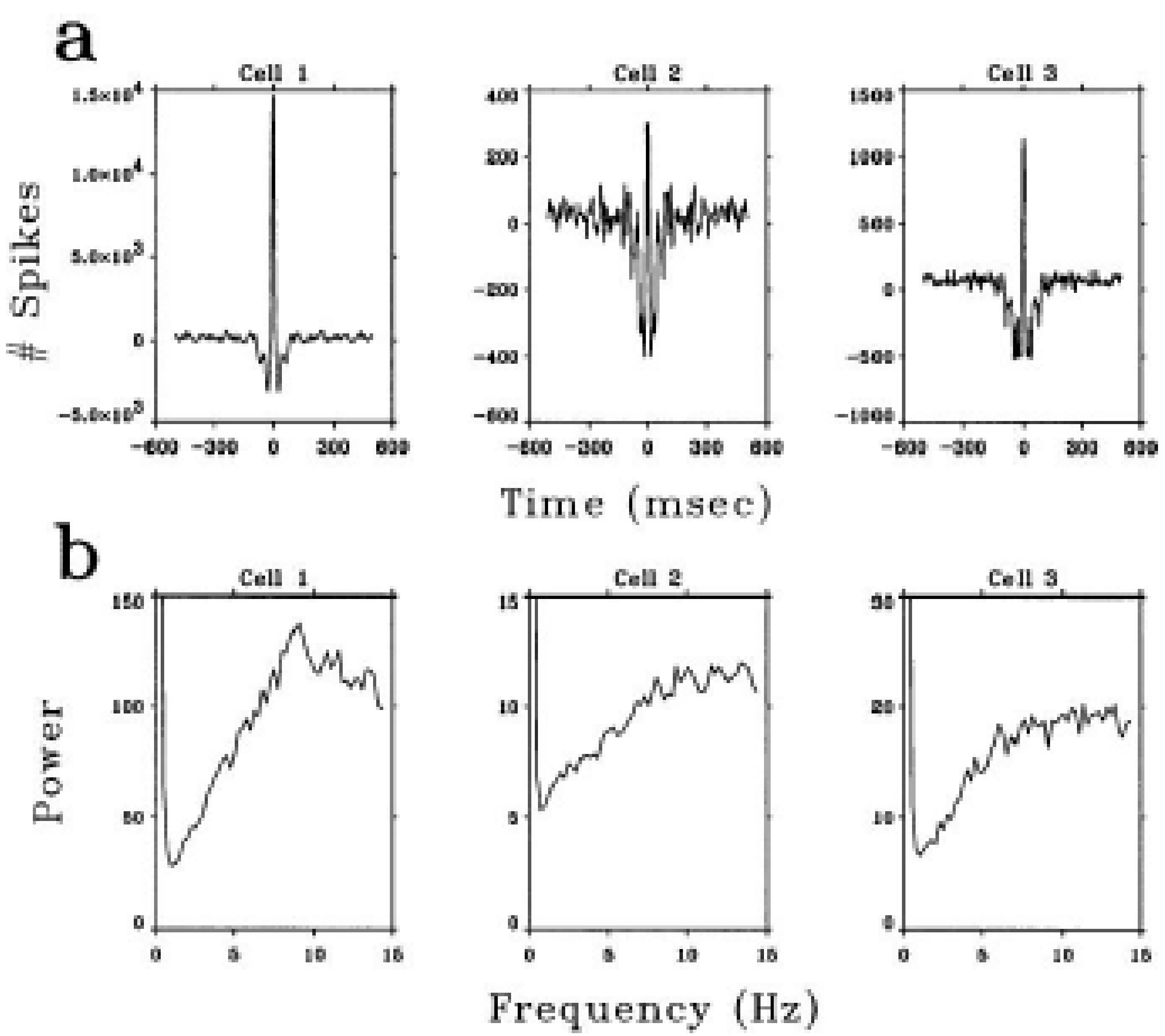


# LGN neurons whiten time-varying natural images

Dan et al, 1996

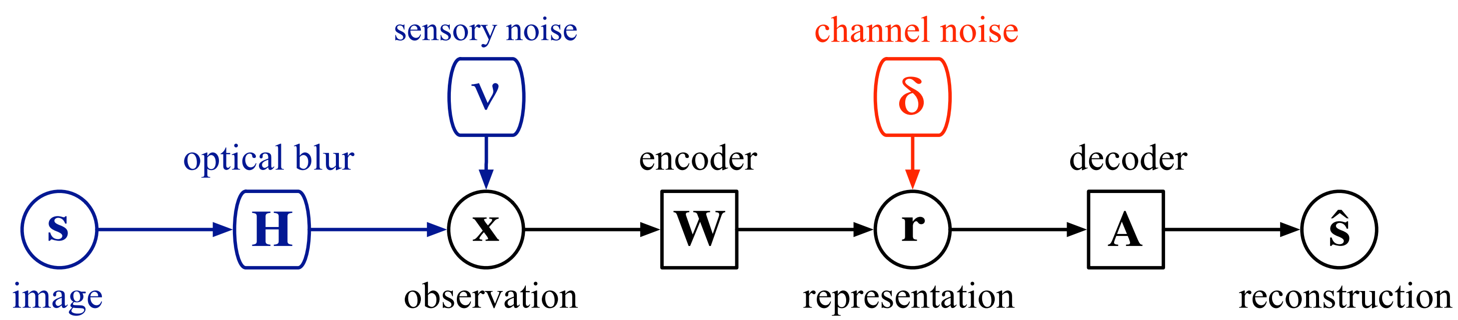


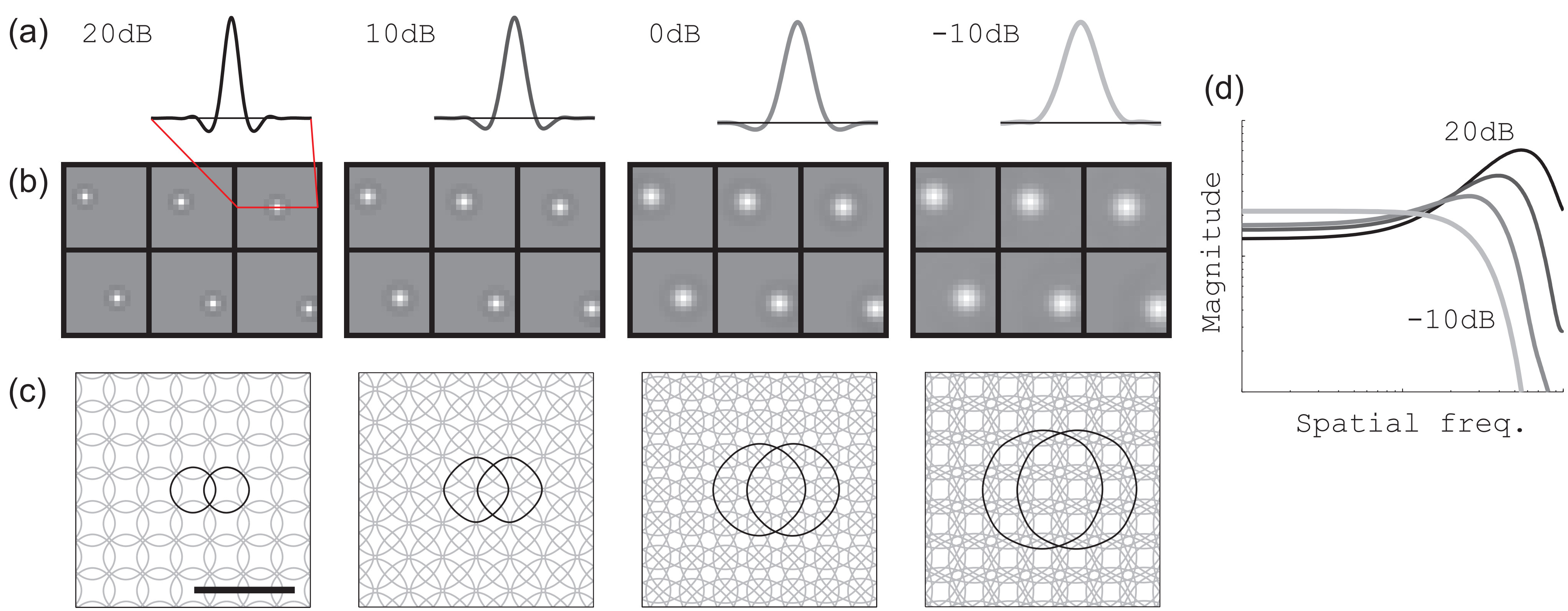
# … but not white noise



# ‘Robust Coding’

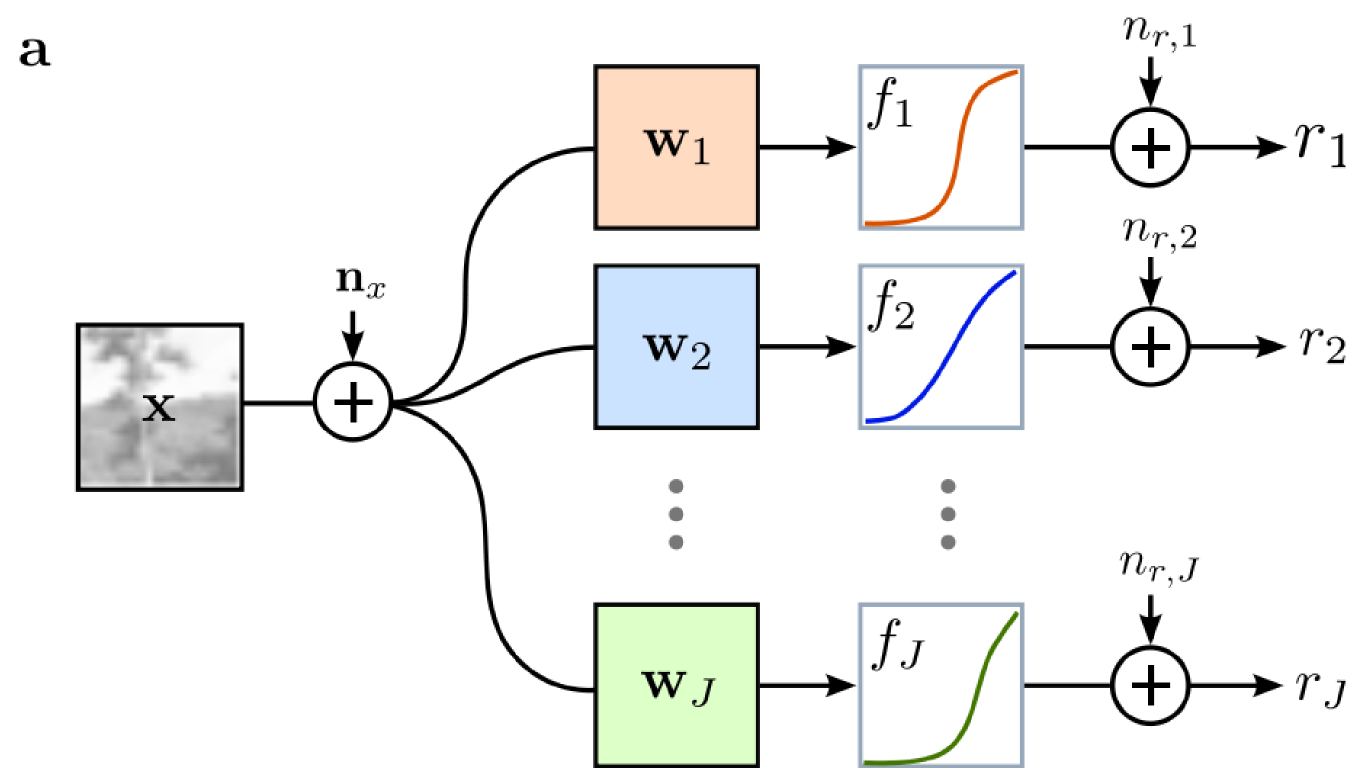
Doi & Lewicki (2006) - *A Theory of Retinal Population Coding*

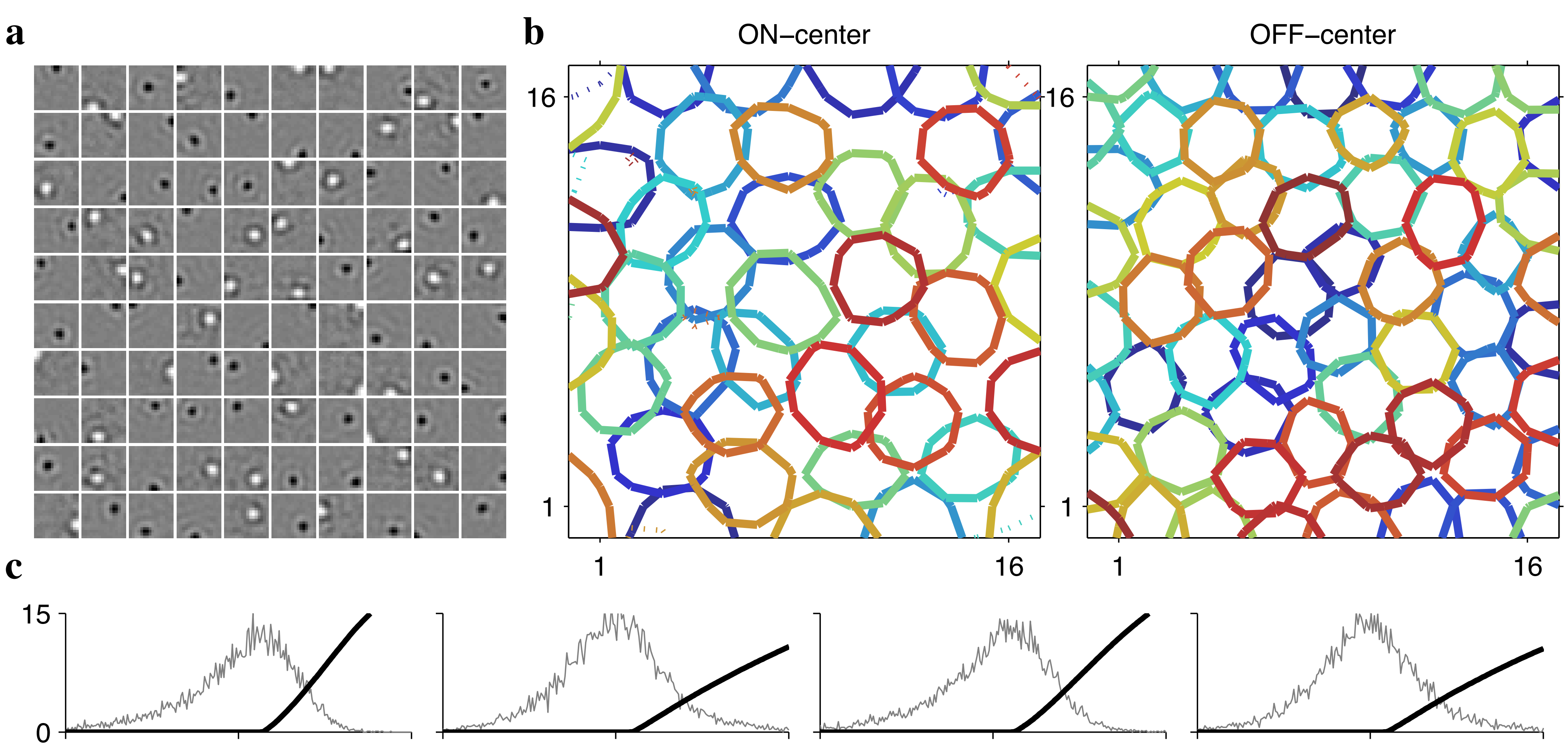


# ‘Robust Coding’

Karklin & Simoncelli (2011) - *Efficient coding of natural images with a population of noisy Linear-Nonlinear neurons*



# Beyond efficient coding

RR is appropriate when there is a bottleneck.

But V1 *expands* dimensionality - many more neurons than inputs

The real goal of sensory representation is to *model* the redundancy in images, not necessarily to reduce it (Barlow 2001)

What we desire is a *meaningful* representation.

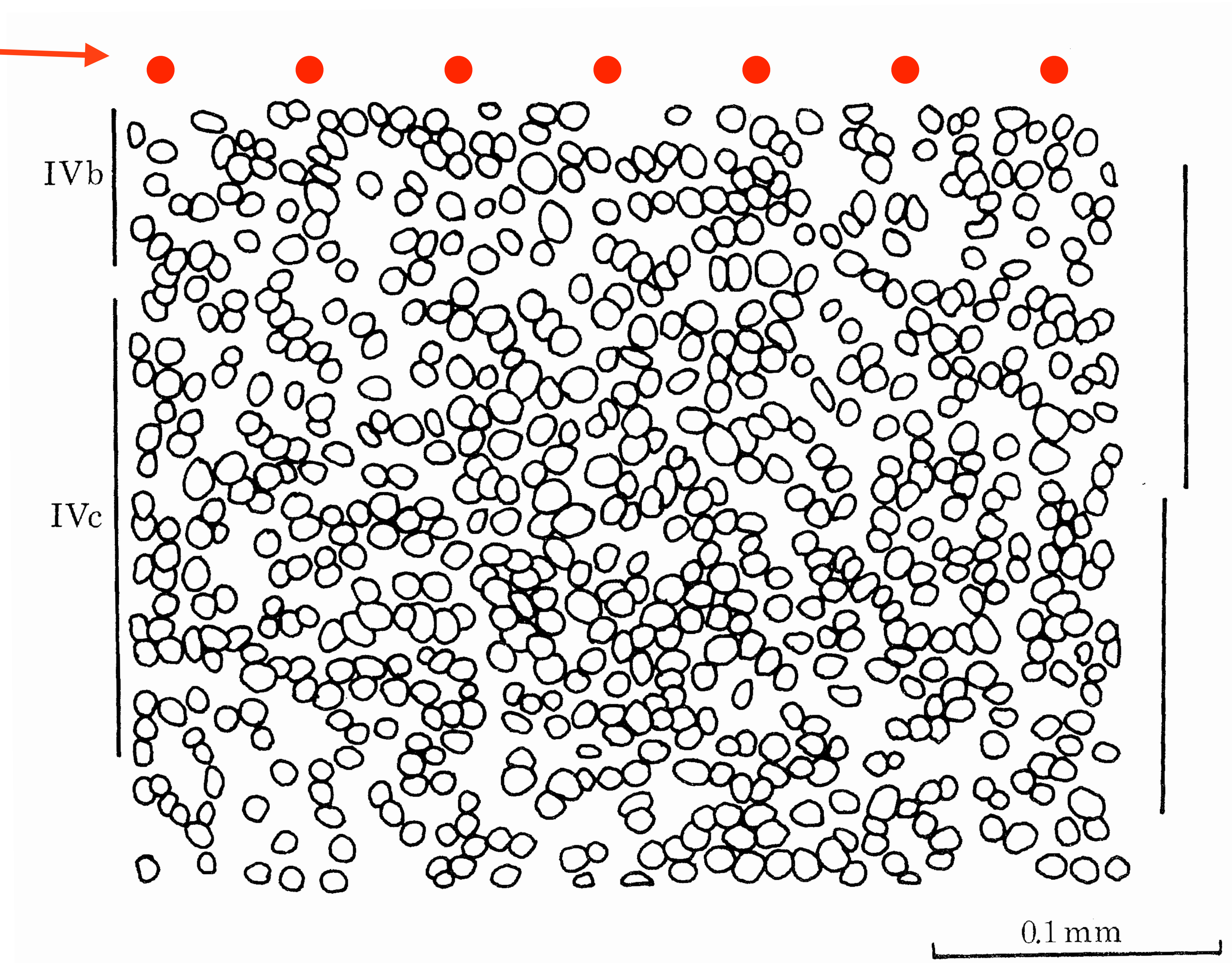
RR provides a valid probabilistic model only when the world can be described in terms of statistically independent components.

To understand cortical representation we must appeal to a different principle.

# V1 is highly overcomplete

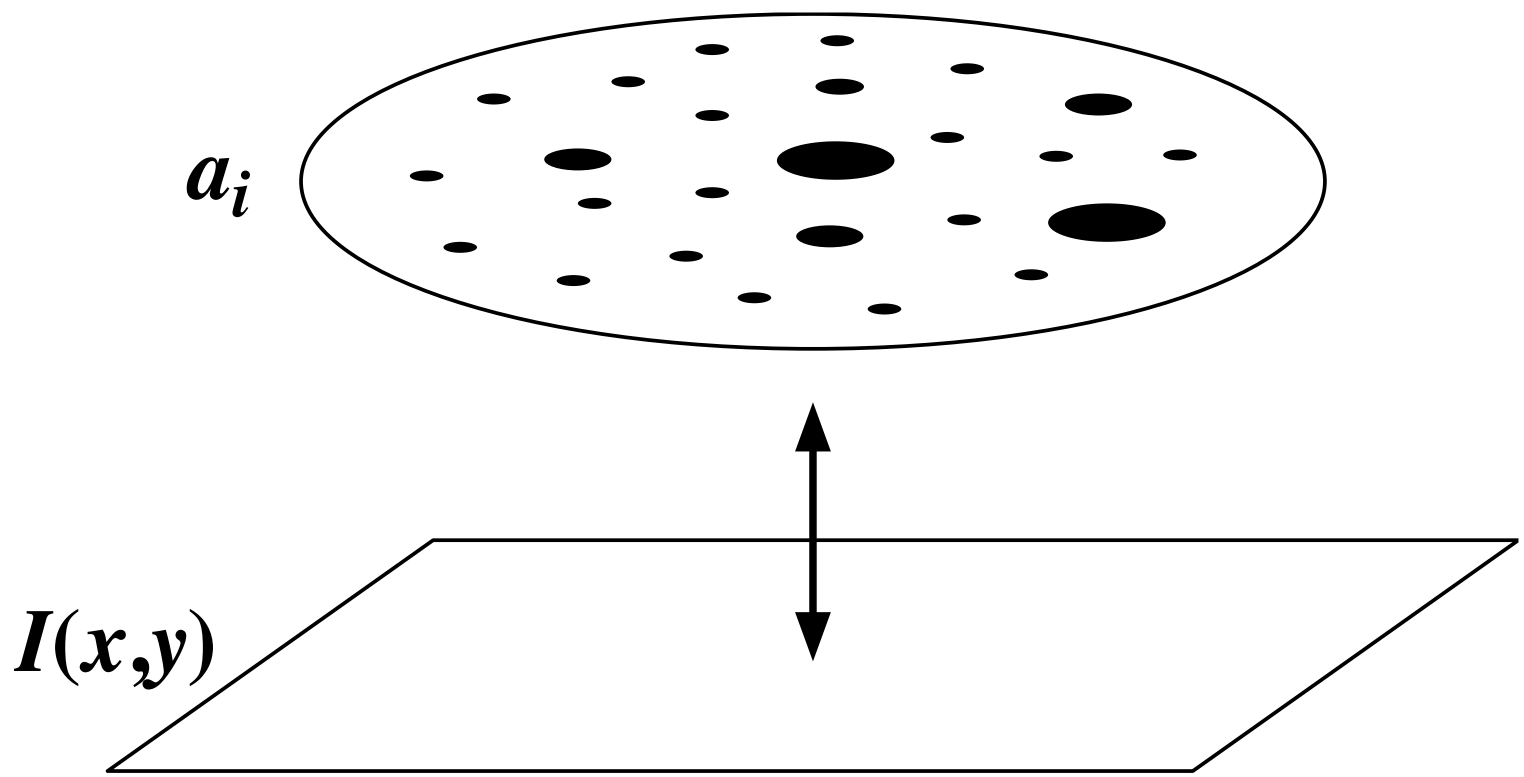
LGN afferents

layer 4 cortex




Barlow (1981)

# Sparse, distributed representation

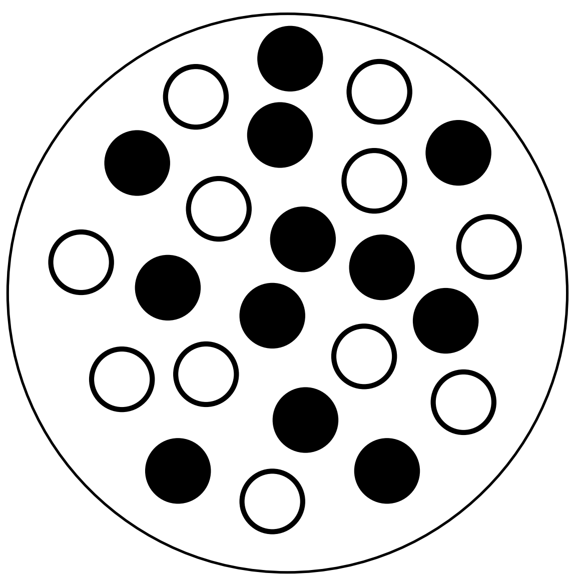



- Provides a way to group things together so that the world can be described in terms of a small number of events at any given moment.

- Converts higher-order redundancy in images into a simple form of redundancy.

# Sparse vs. dense vs. 'grandmother cell' codes

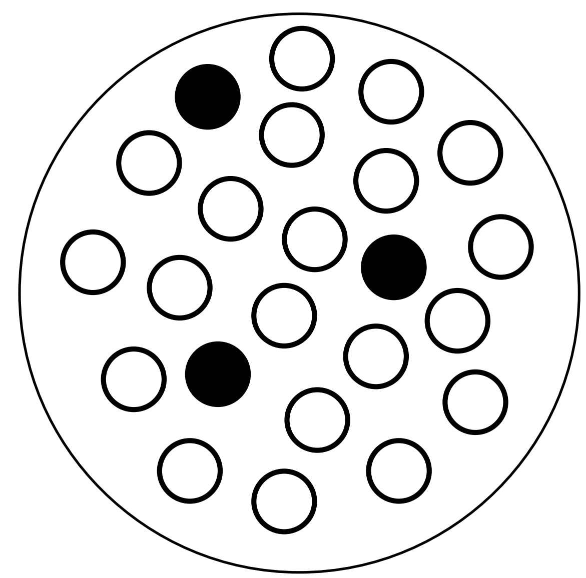
Dense codes
(ascii)



+ High combinatorial capacity ($2^N$)
- Difficult to read out

. . .

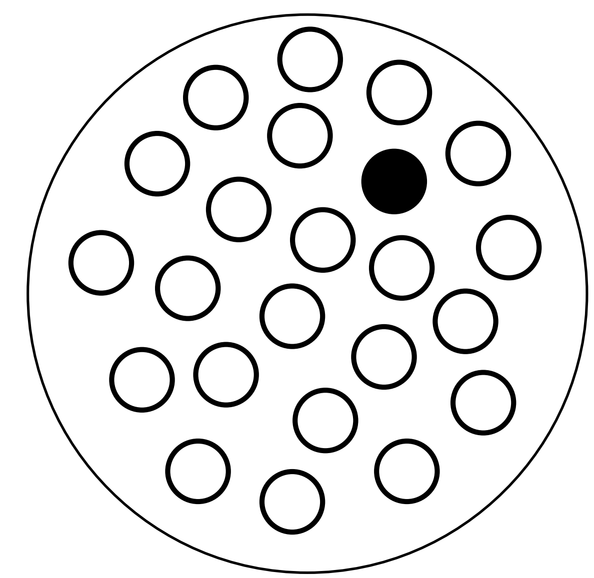
Sparse, distributed codes



+ Decent combinatorial capacity ($\sim N^K$)
+ Still easy to read out

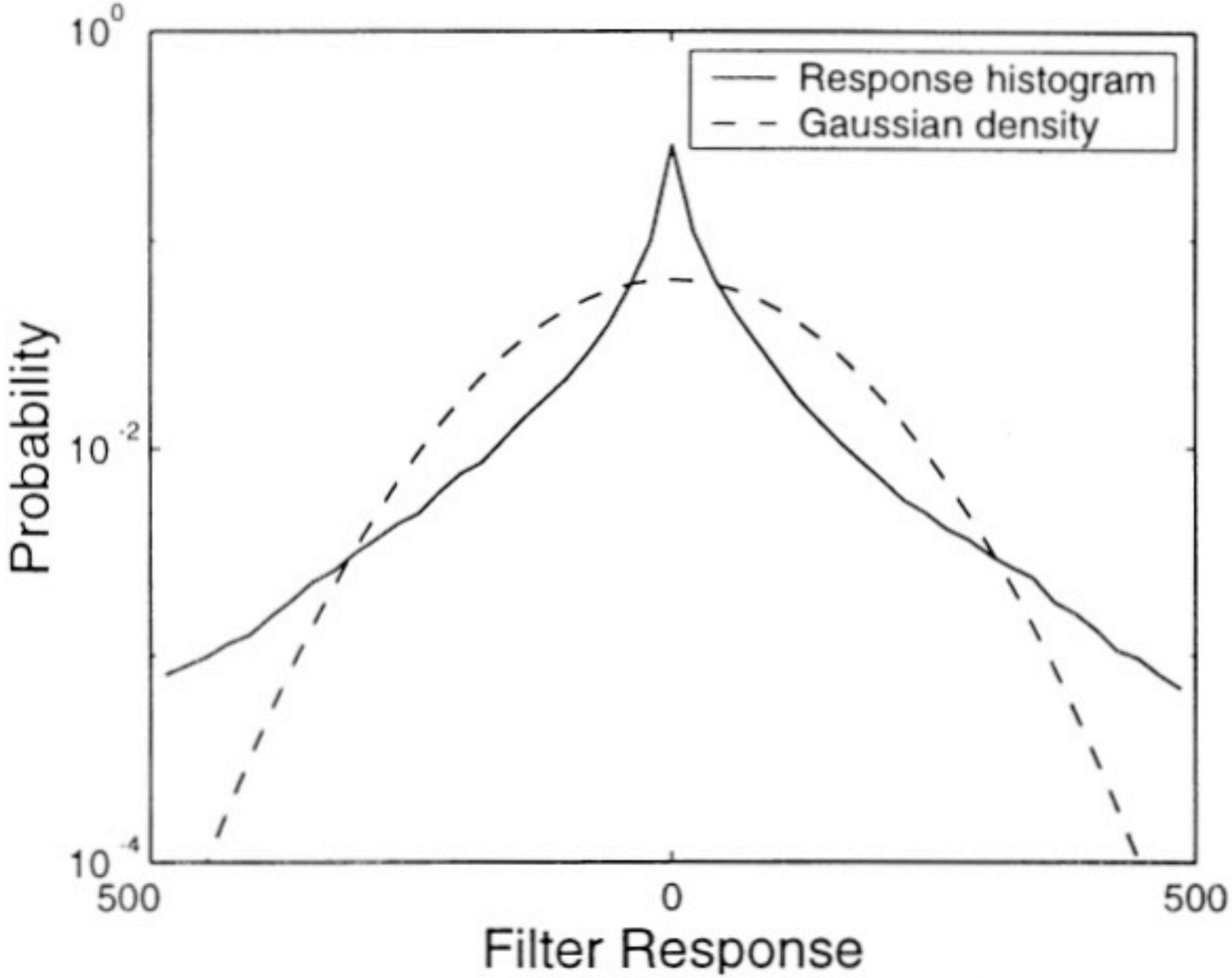
. . .

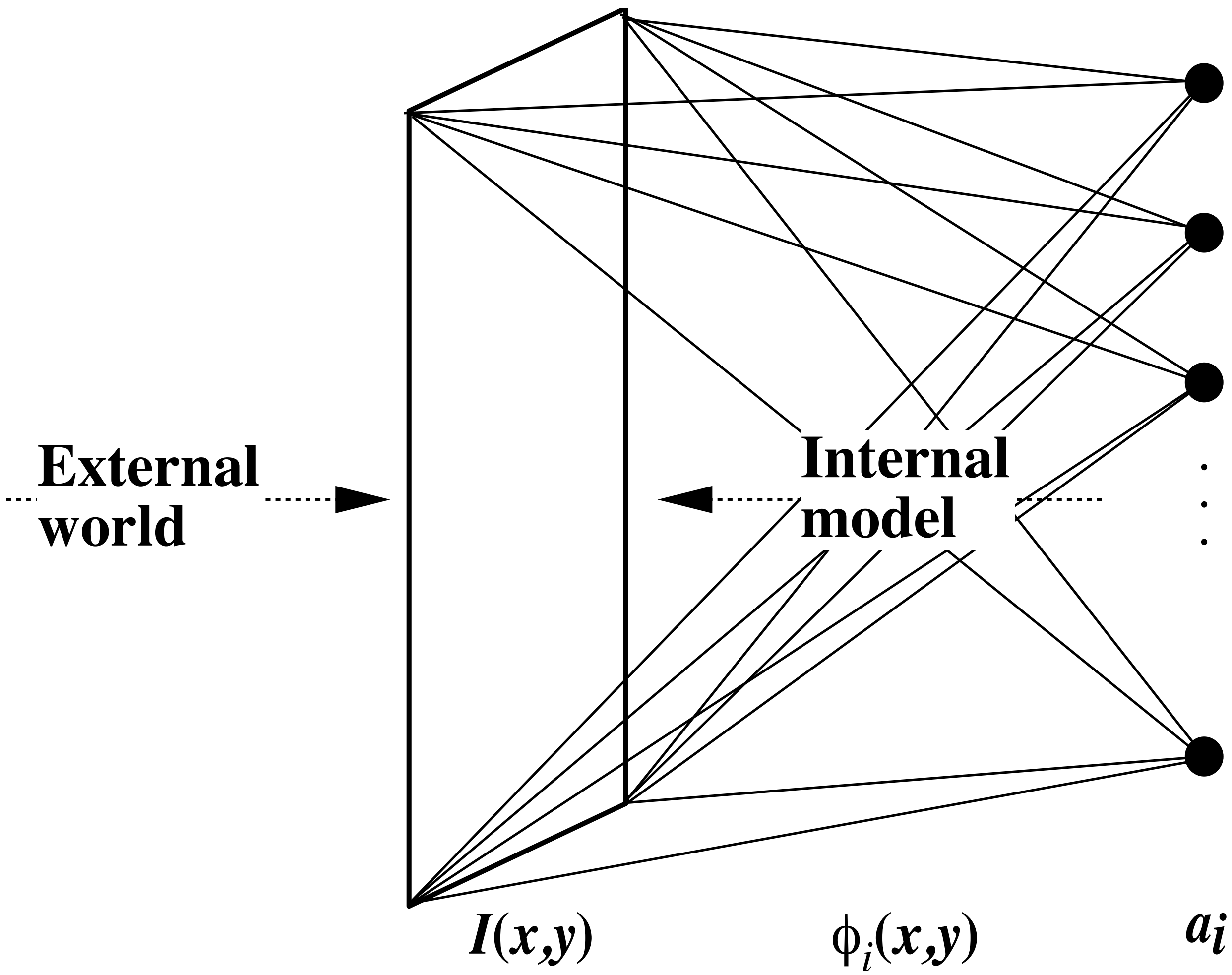
Local codes
(grandmother cells)



- Low combinatorial capacity (N)
+ Easy to read out

# Gabor-filter response histogram

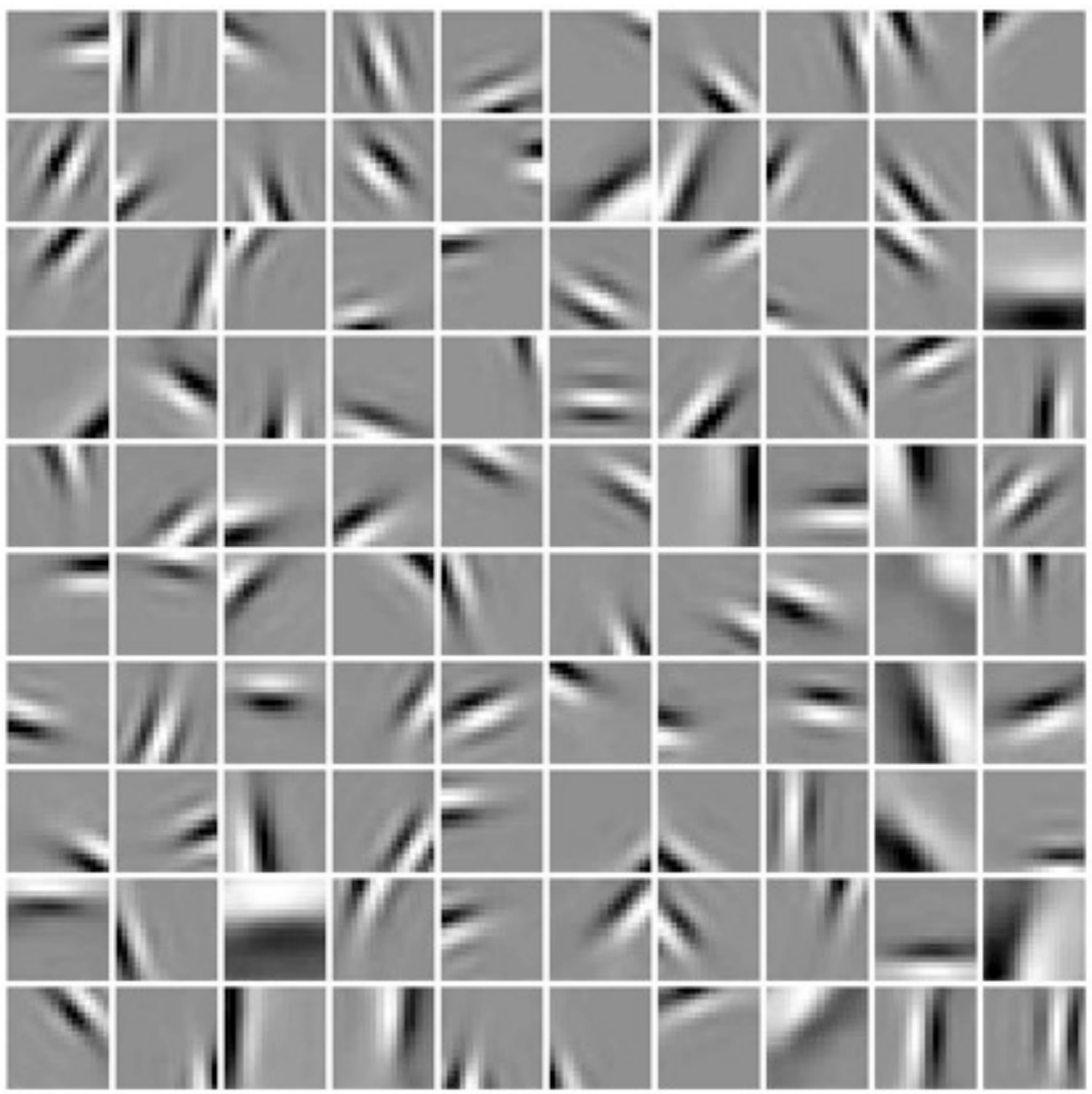


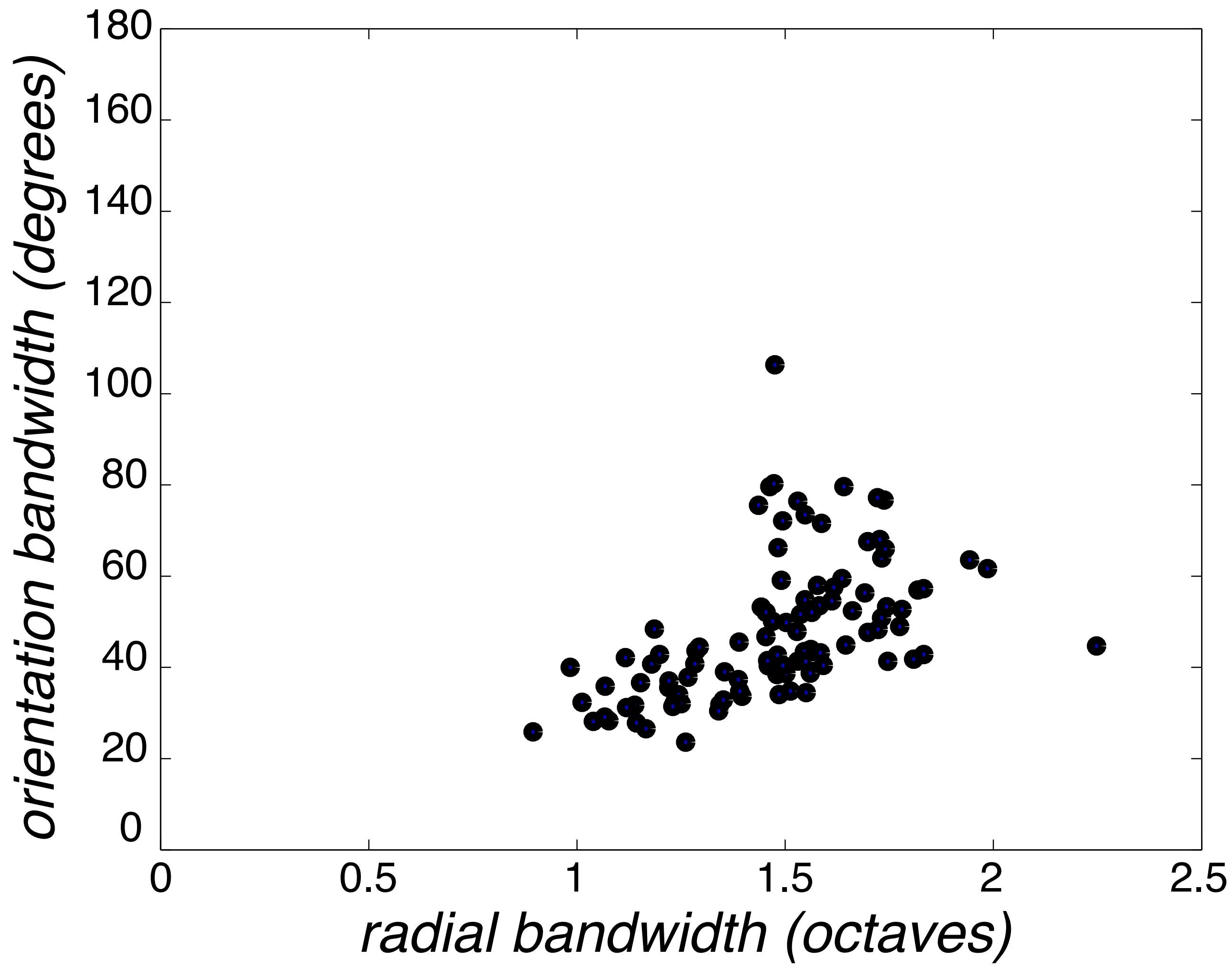



# Sparse coding image model

$$\underbrace{I(x,y)}_{\text{image}} = \sum_i \underbrace{a_i}_{\text{neural activities (sparse)}} \underbrace{\phi_i(x,y)}_{\text{features}} + \underbrace{\epsilon(x,y)}_{\text{other stuff}}$$

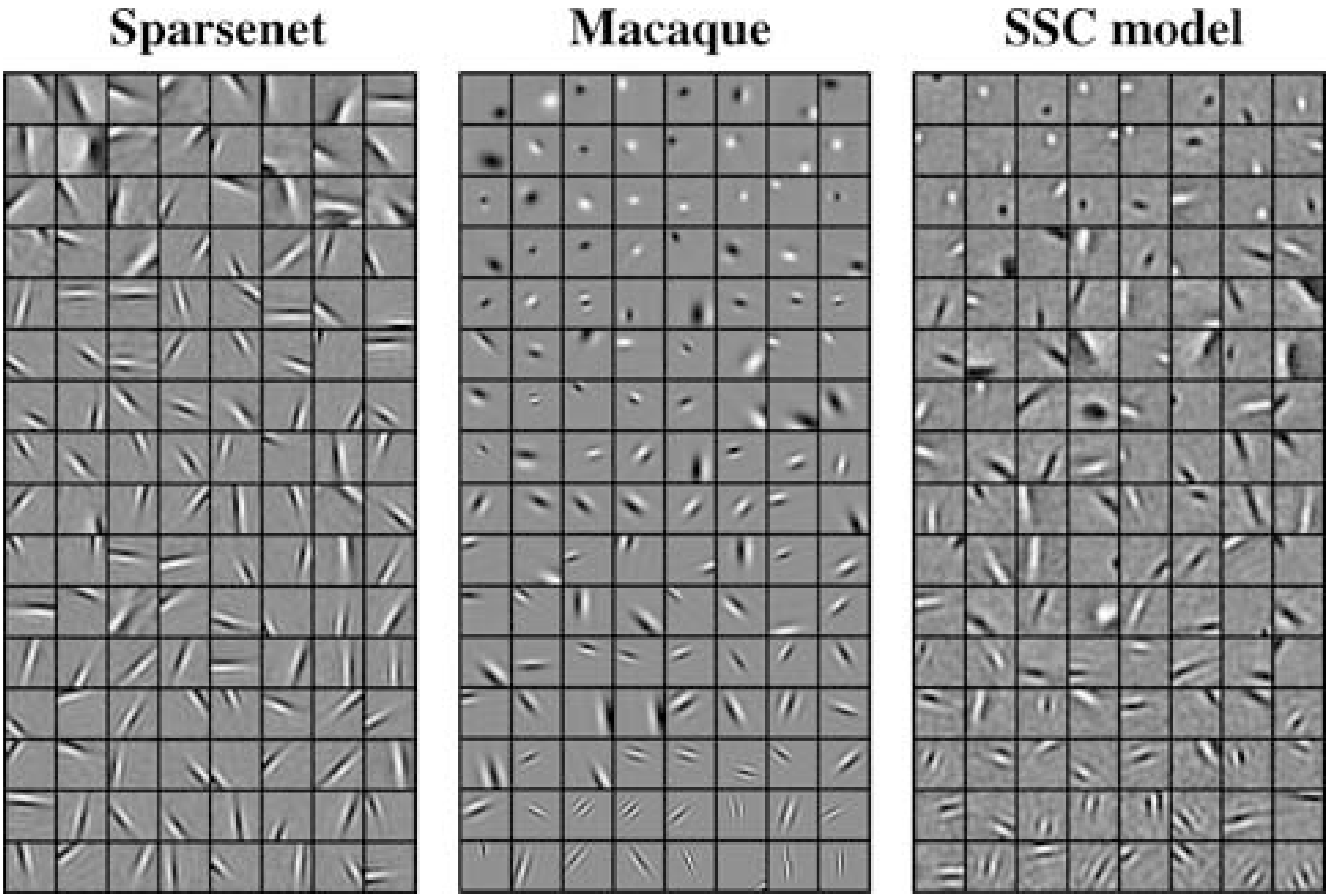
# Learned dictionary Φ
## (critically sampled)





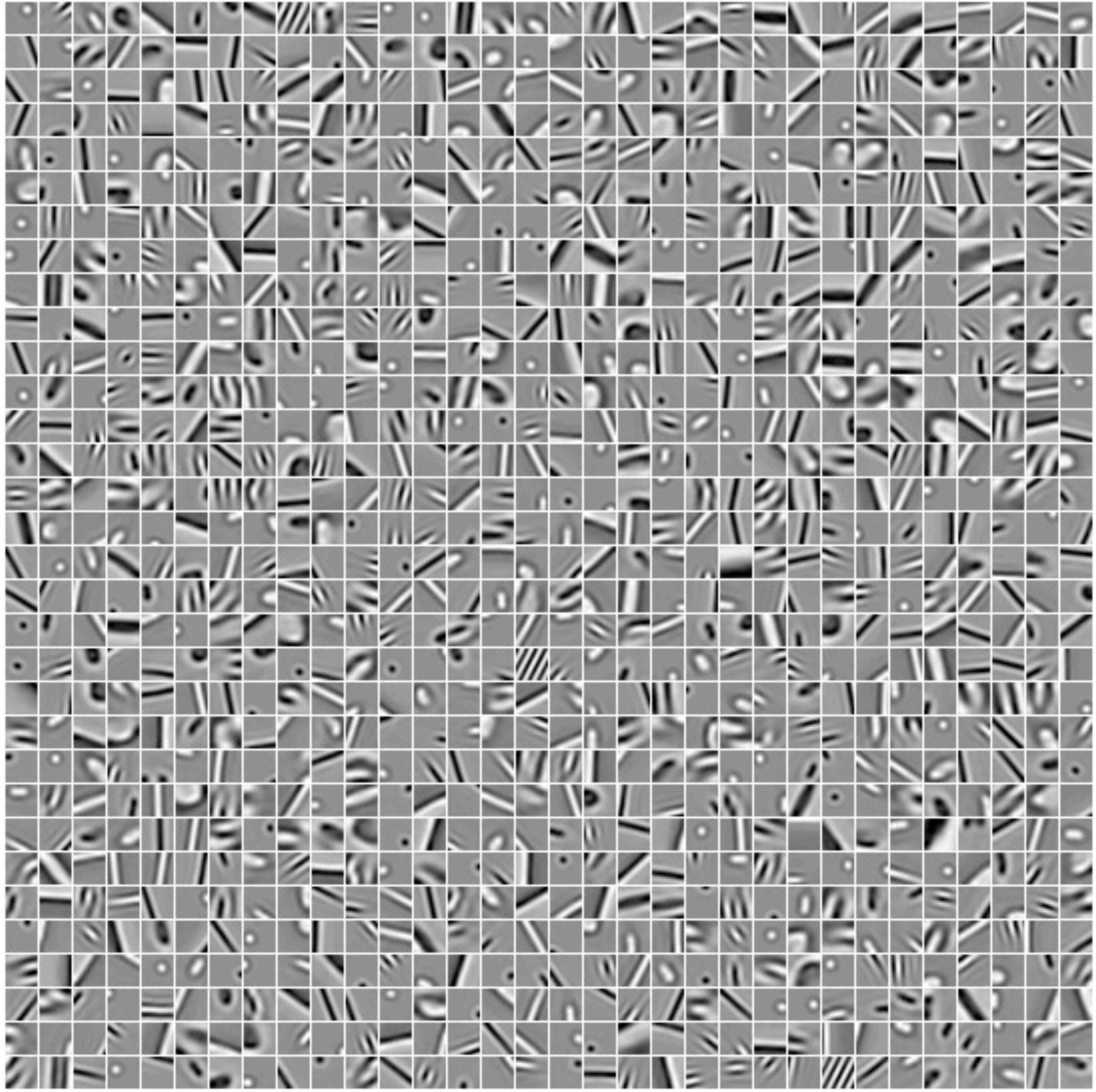
# Effect of overcompleteness and ‘hard sparsity’

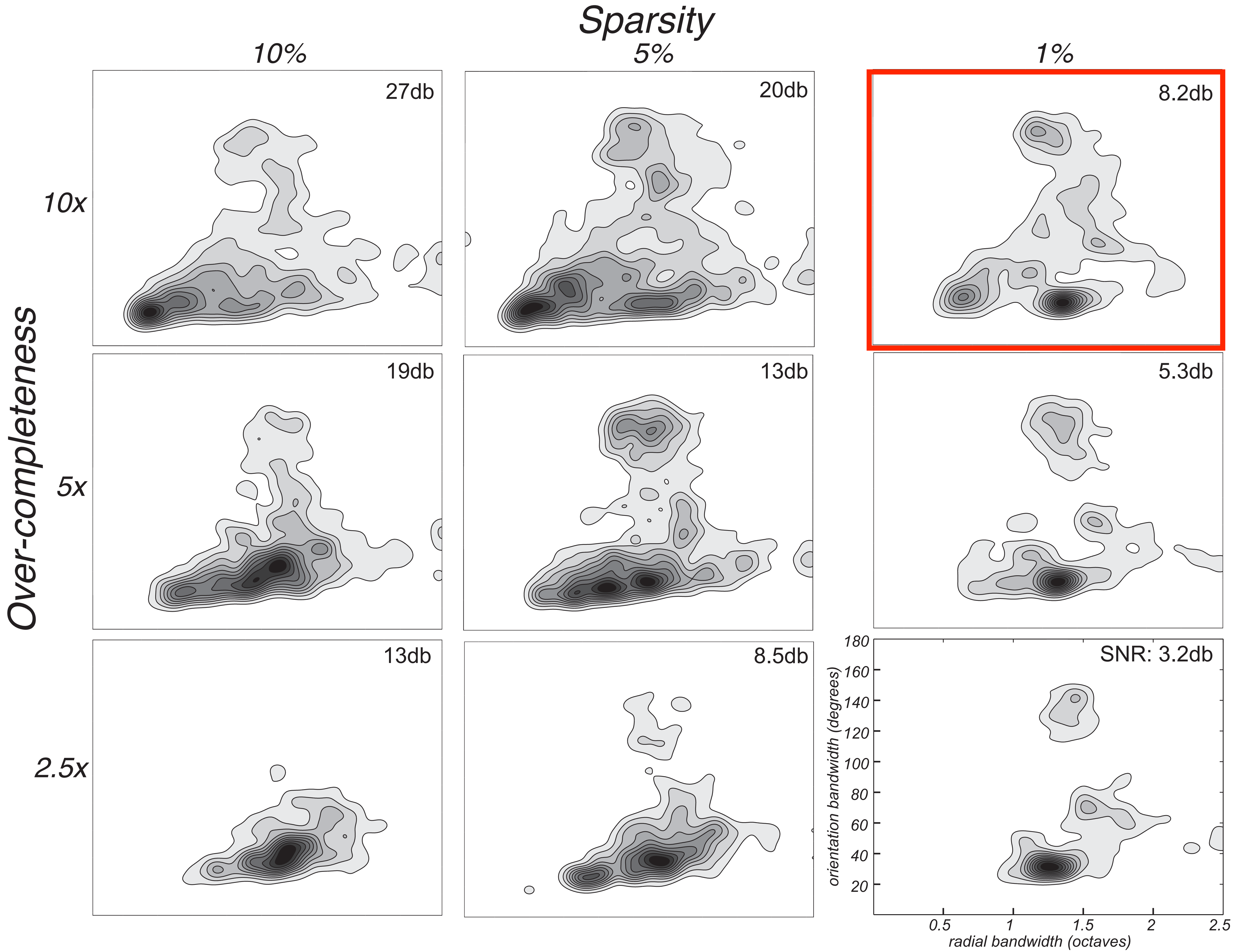
(Rehn and Sommer 2006)



Learned dictionary 10x overcomplete

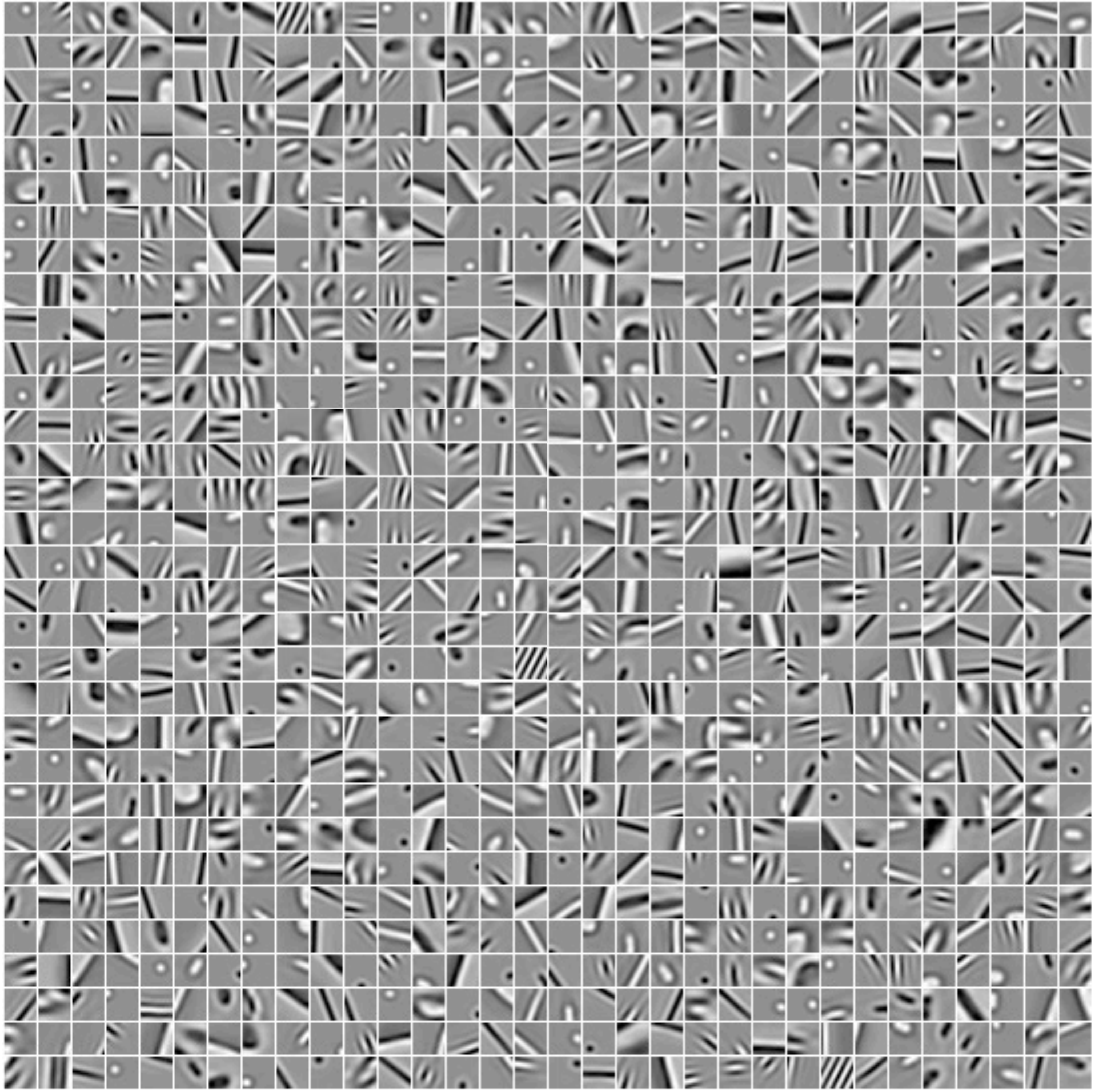
(joint work with David Warland, UC Davis)

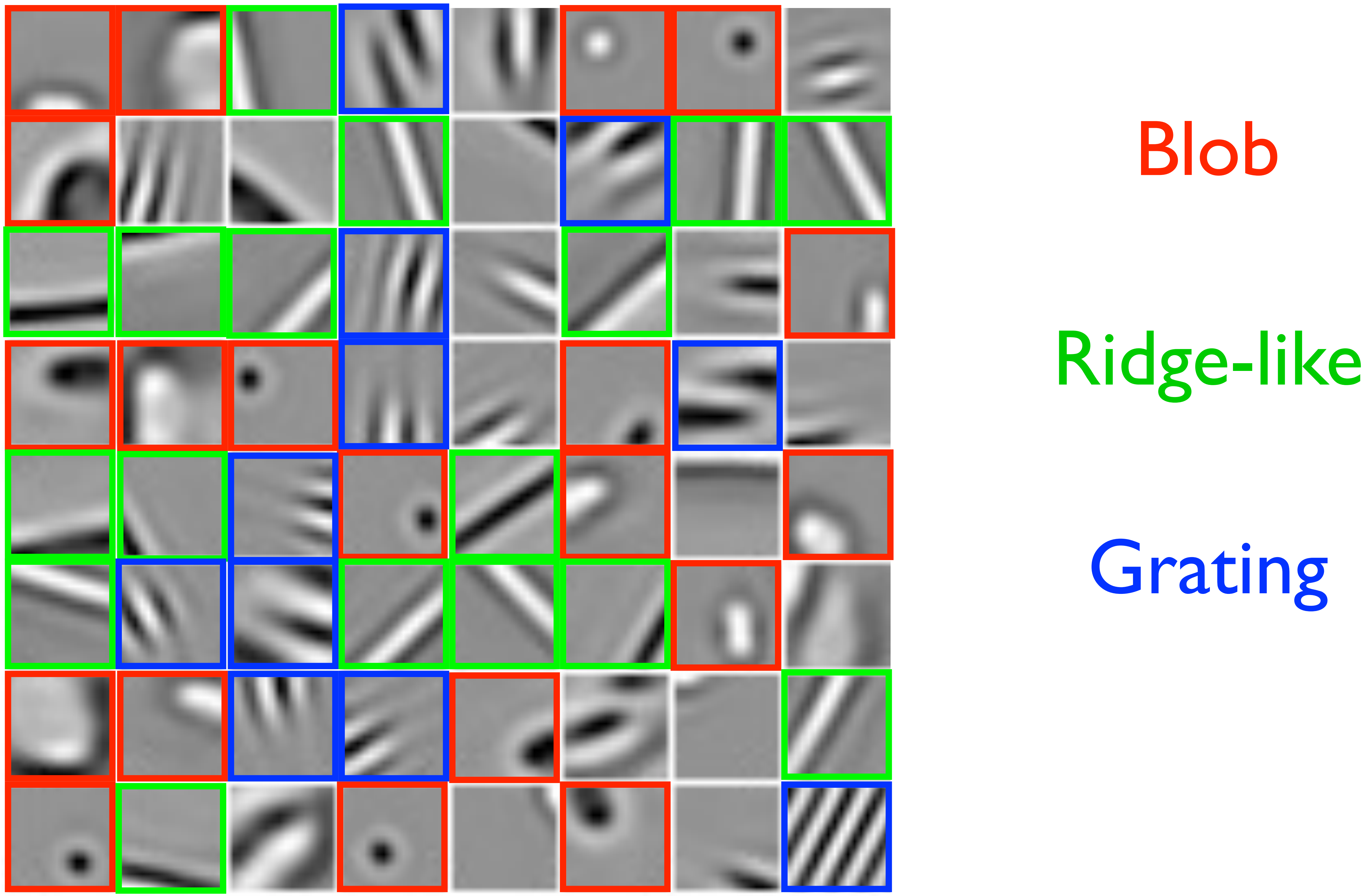


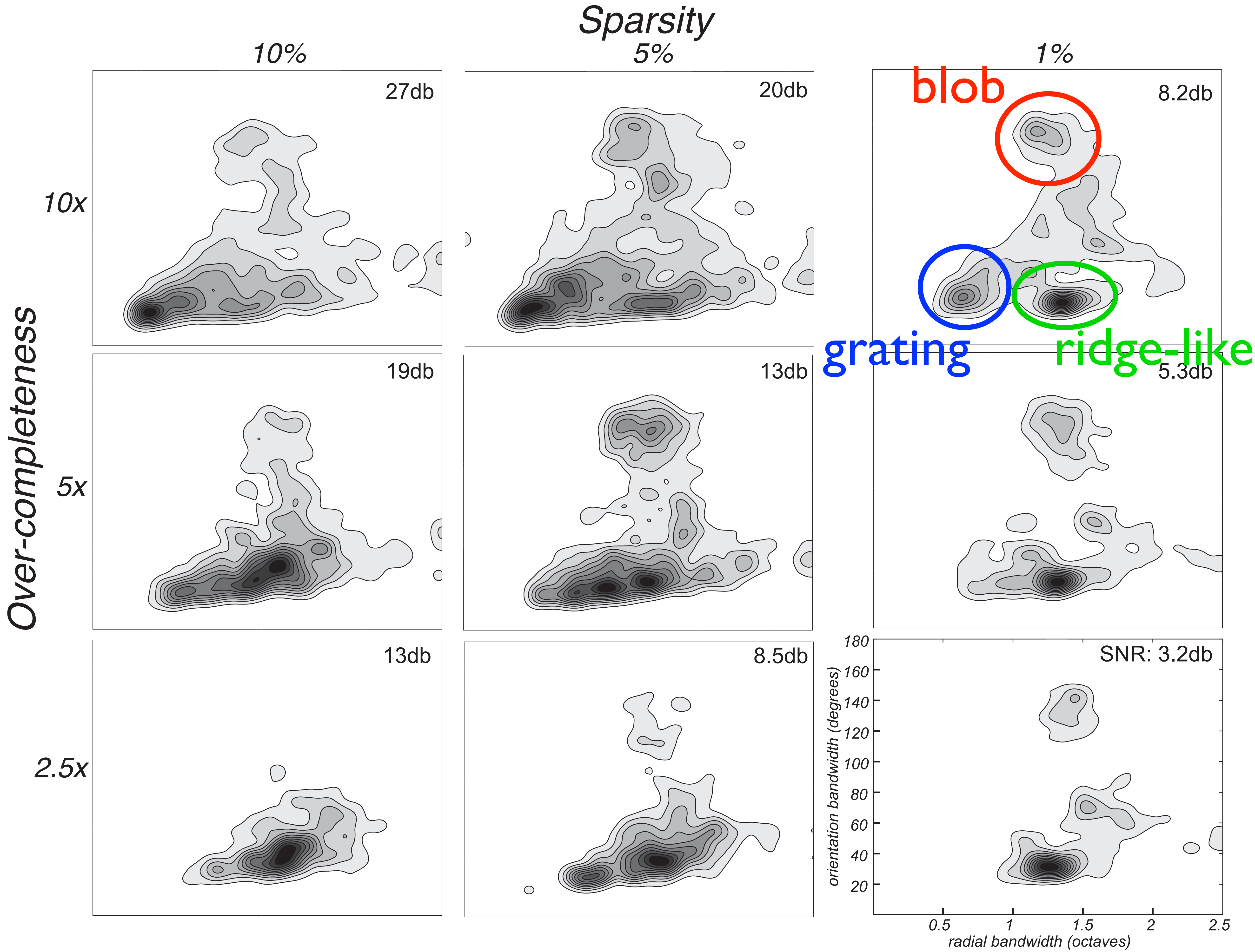


Sparsity
10%
5%
1%
Over-completeness
10x
5x
2.5x
27db
20db
8.2db
19db
13db
5.3db
13db
8.5db
SNR: 3.2db
orientation bandwidth (degrees)
180
160
140
120
100
80
60
40
20
radial bandwidth (octaves)
0.5
1
1.5
2
2.5

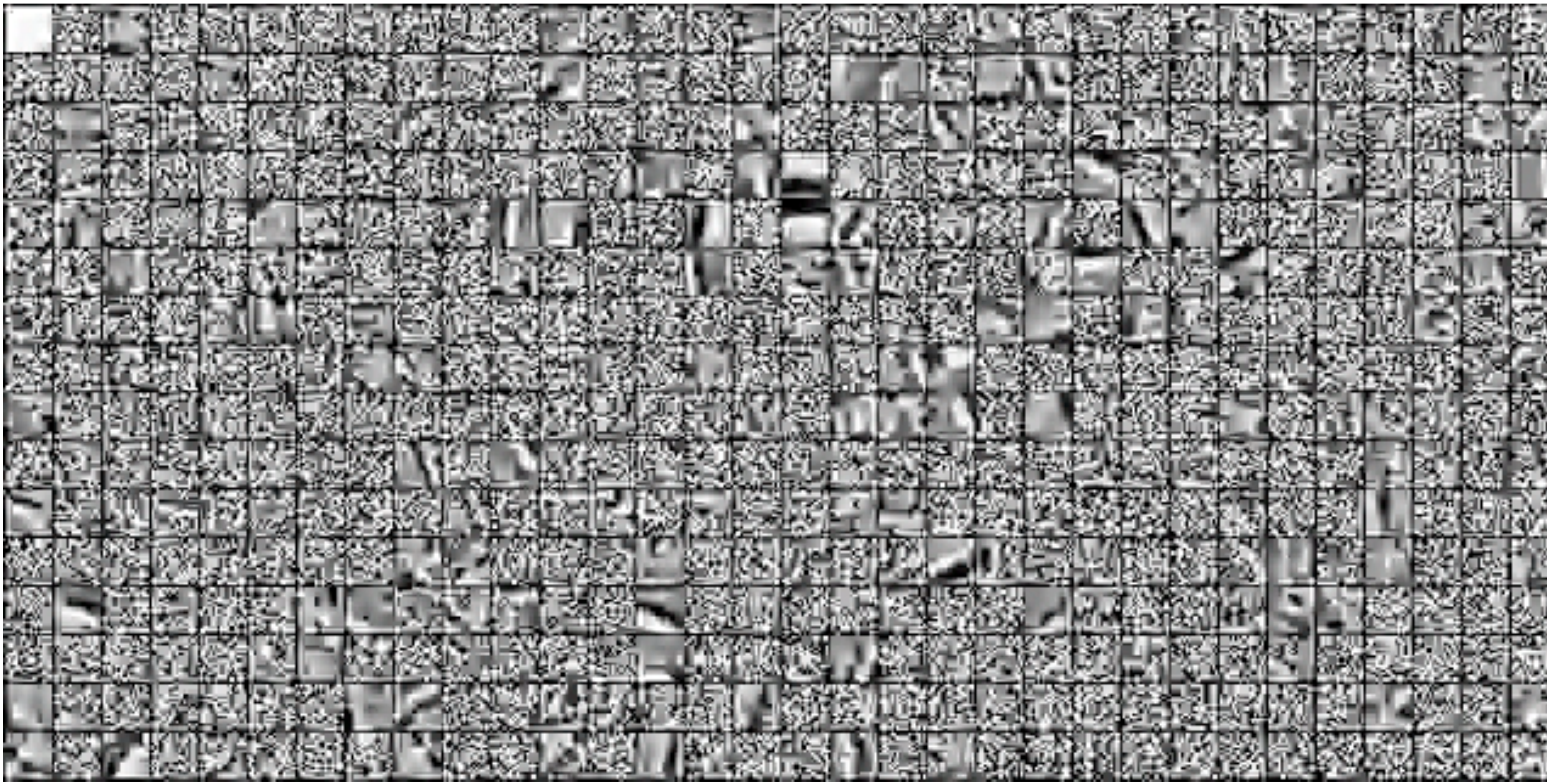
10x, 1%




Blob
Ridge-like
Grating

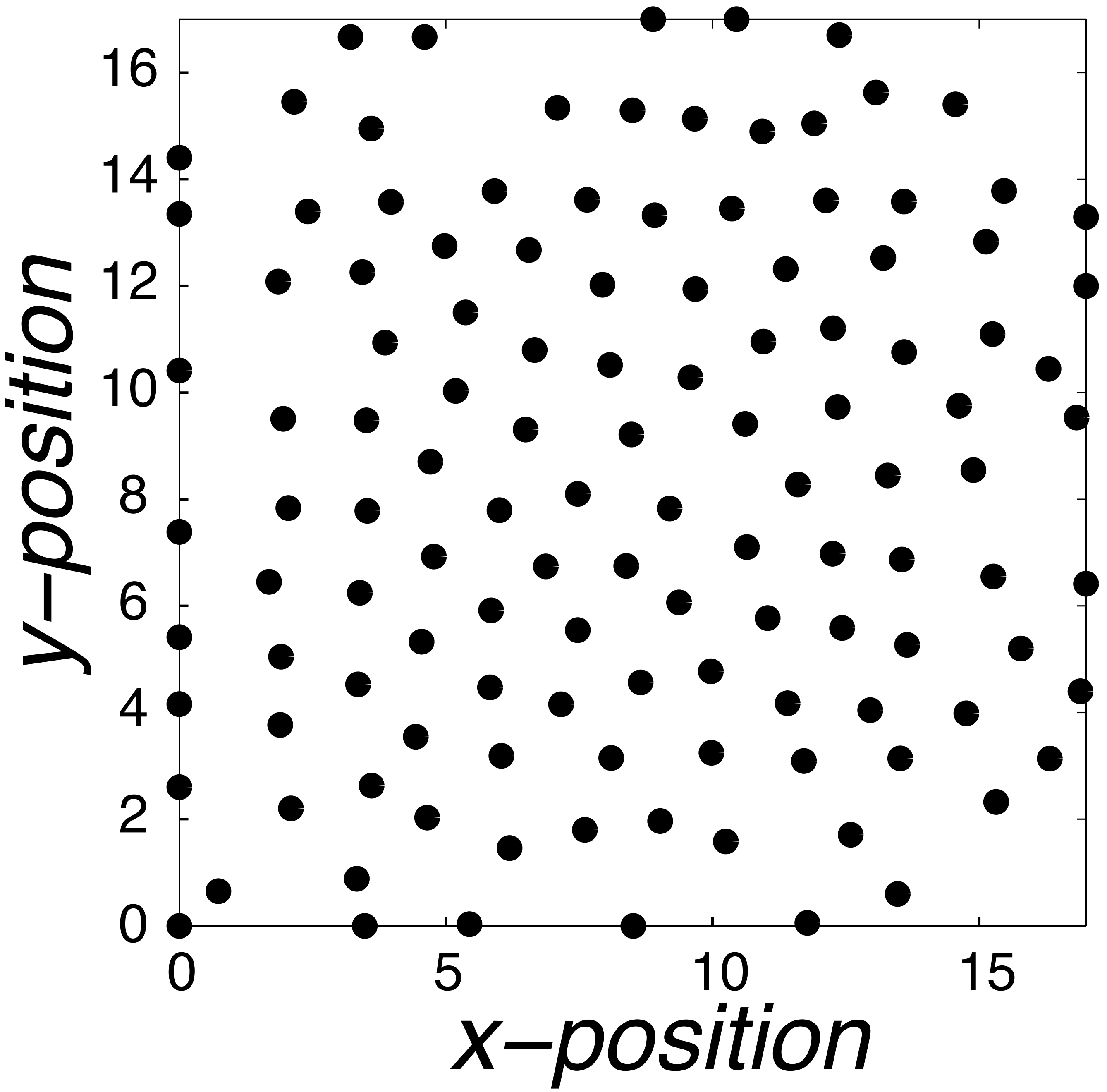

Sparsity
10%
5%
1%
Over-completeness
10x
5x
2.5x
27db
20db
8.2db
blob
grating
ridge-like
19db
13db
5.3db
13db
8.5db
SNR: 3.2db
180
160
140
120
100
80
60
40
20
orientation bandwidth (degrees)
0.5
1
1.5
2
2.5
radial bandwidth (octaves)

# Solutions are stable

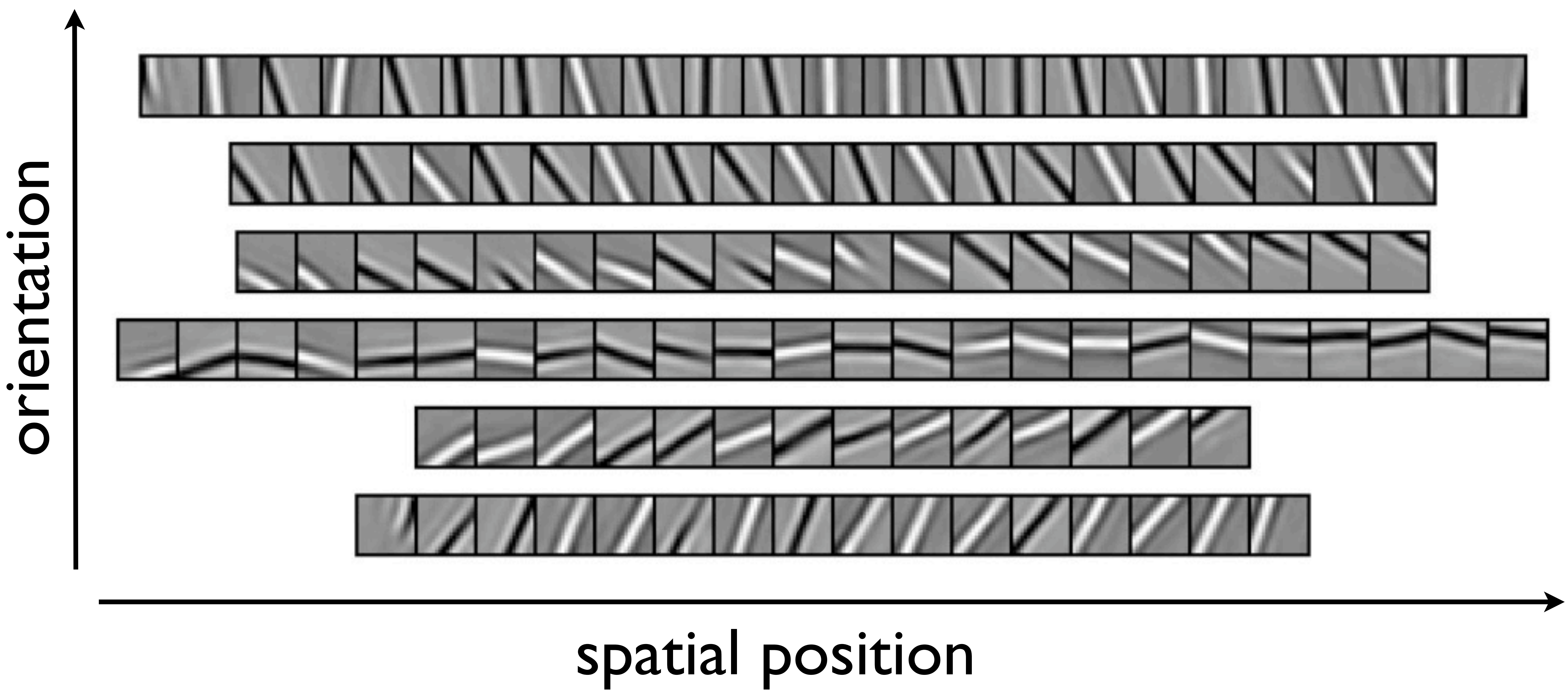


# Tiling properties: Blobs

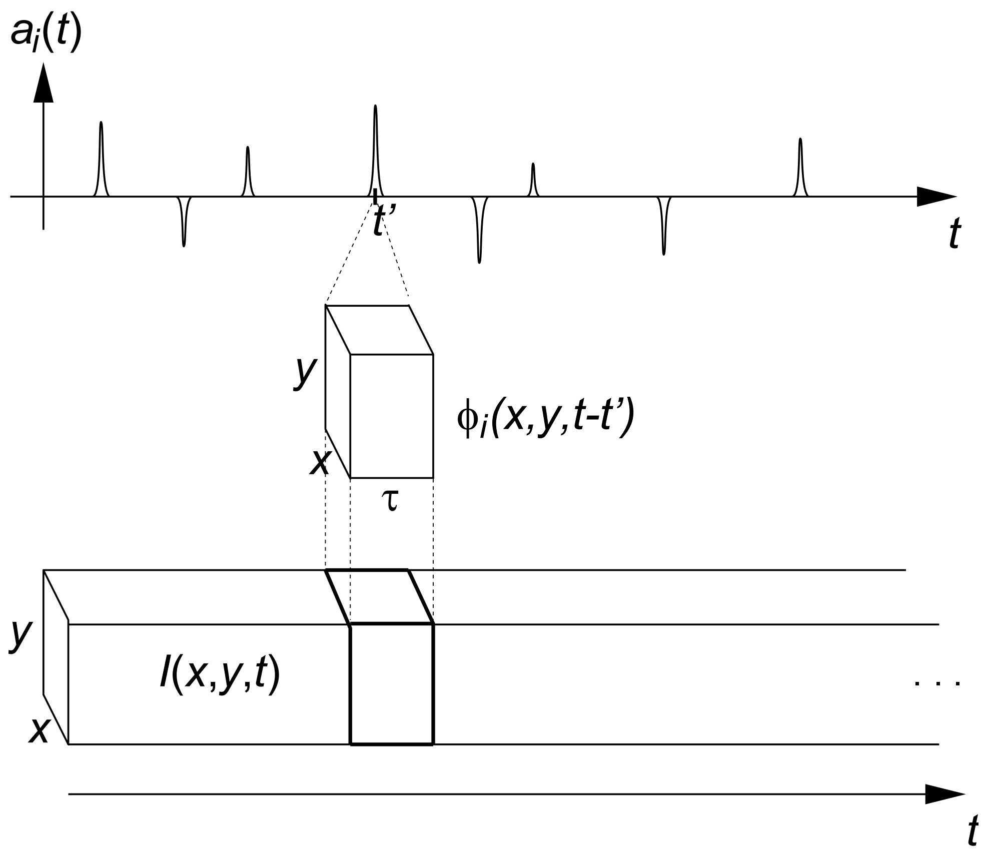
## (highest spatial-frequency band only)



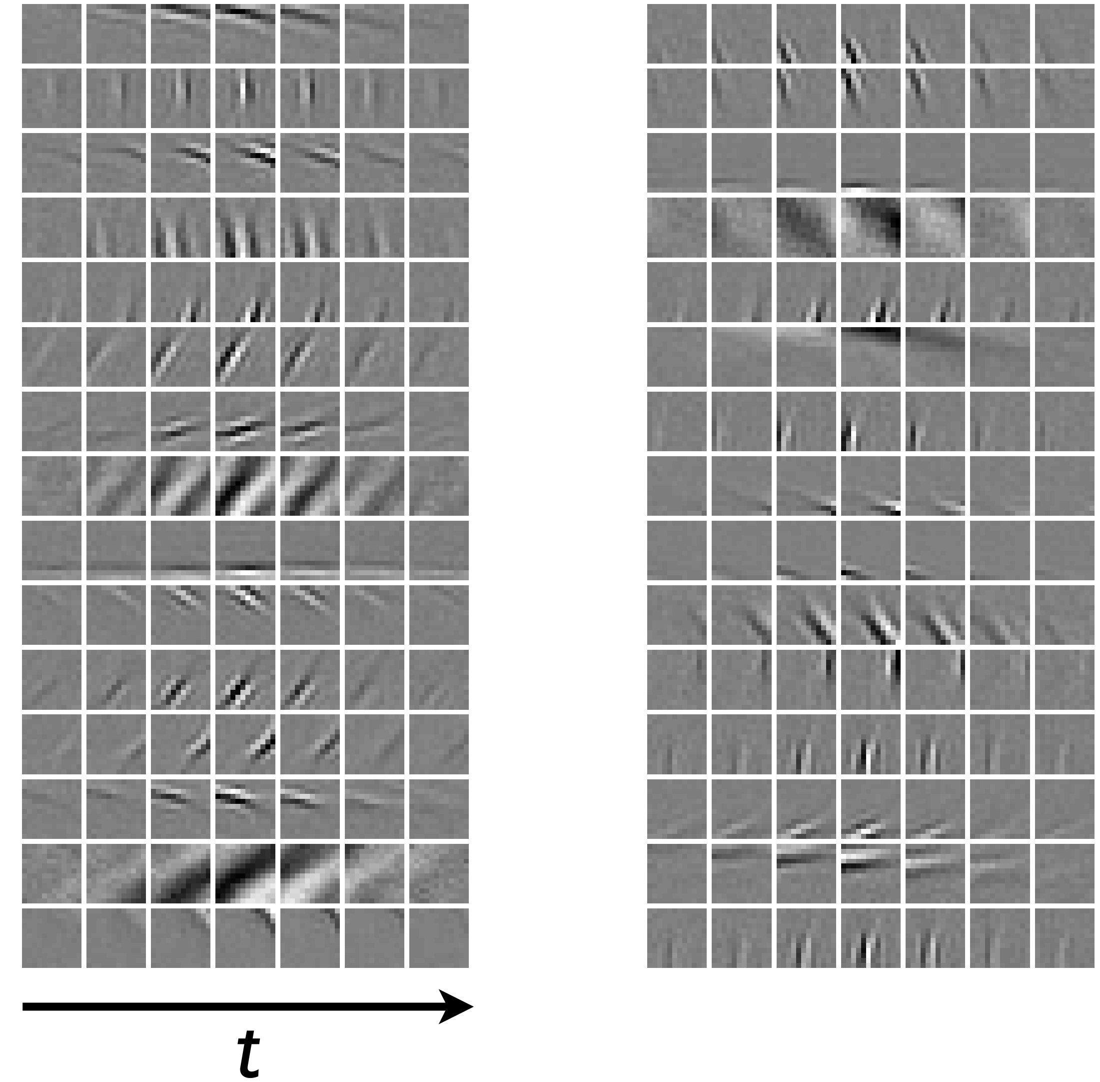
# Tiling properties: Ridge-like

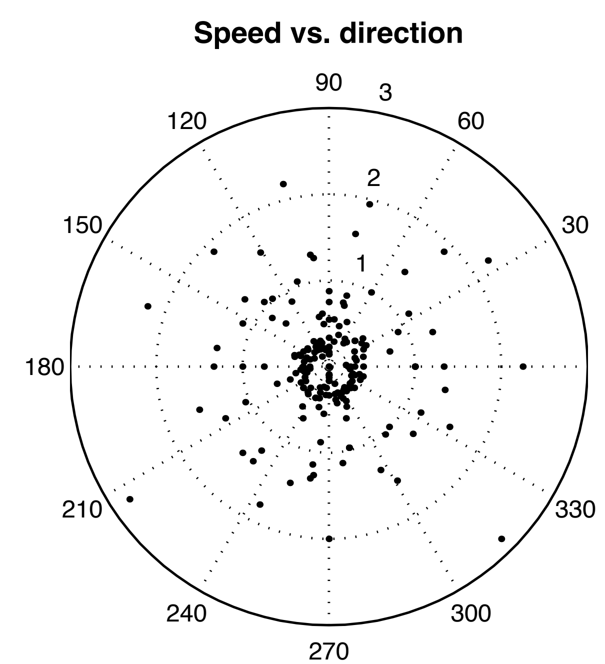


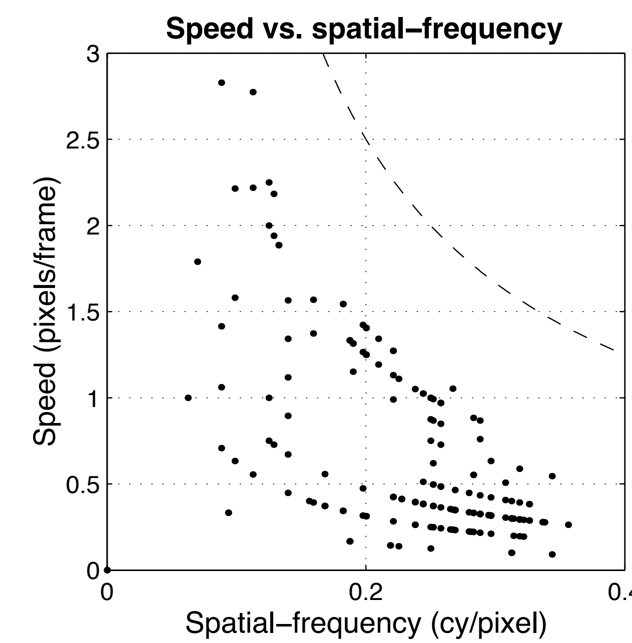
# Sparse coding of time-varying images



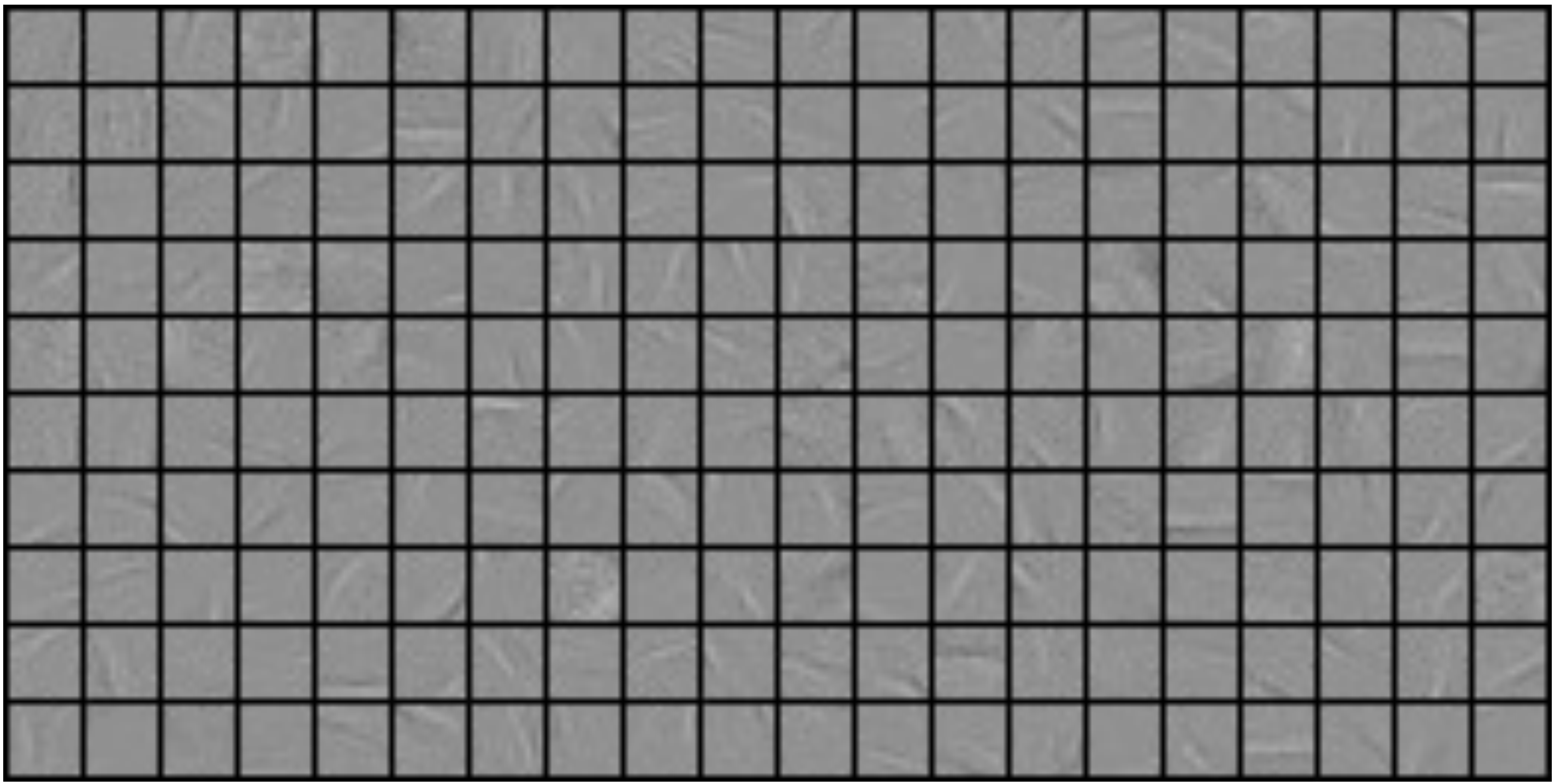

$$I(\vec{x},t) = \sum_i a_i(t) * \phi_i(\vec{x},t) + \epsilon(\vec{x},t)$$



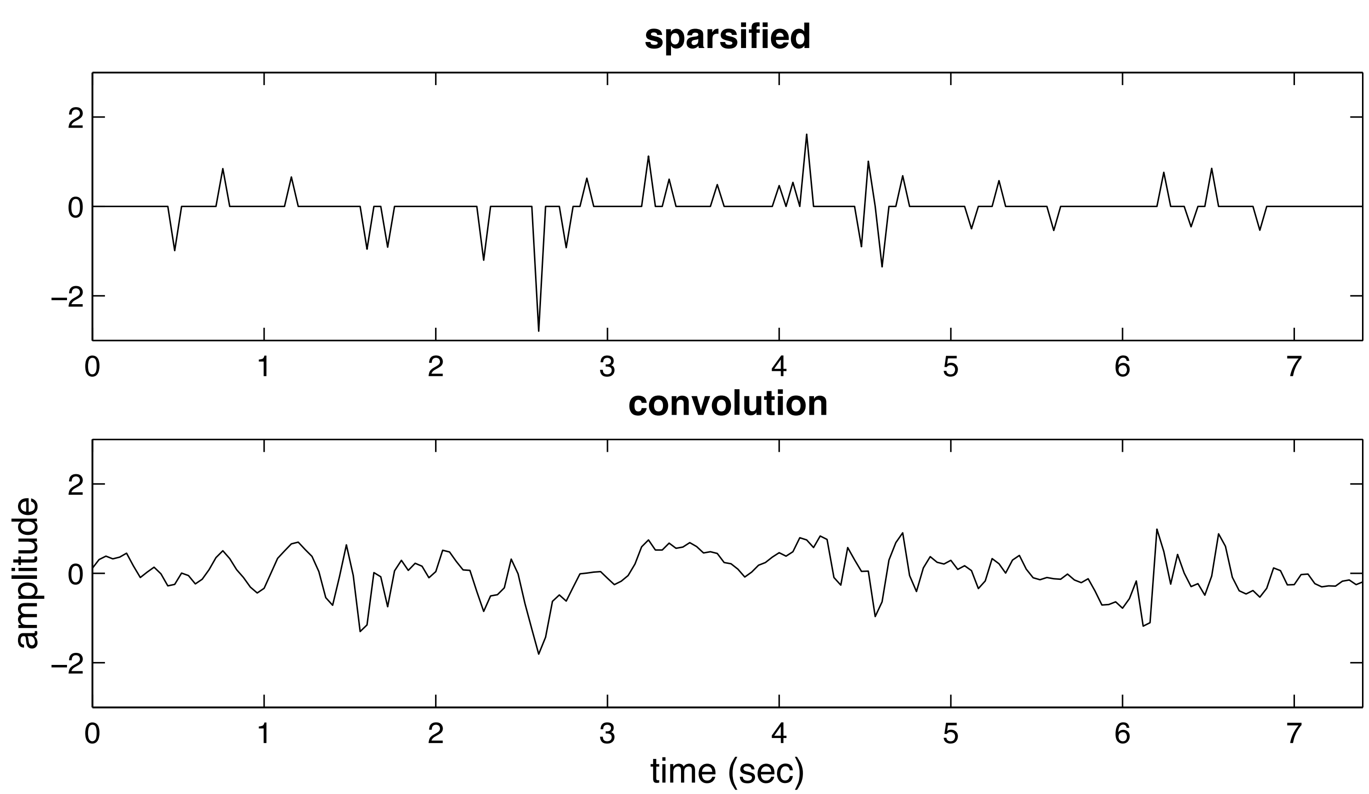



# Learned basis space-time basis functions
(200 bfs, 12 x12 x 7)

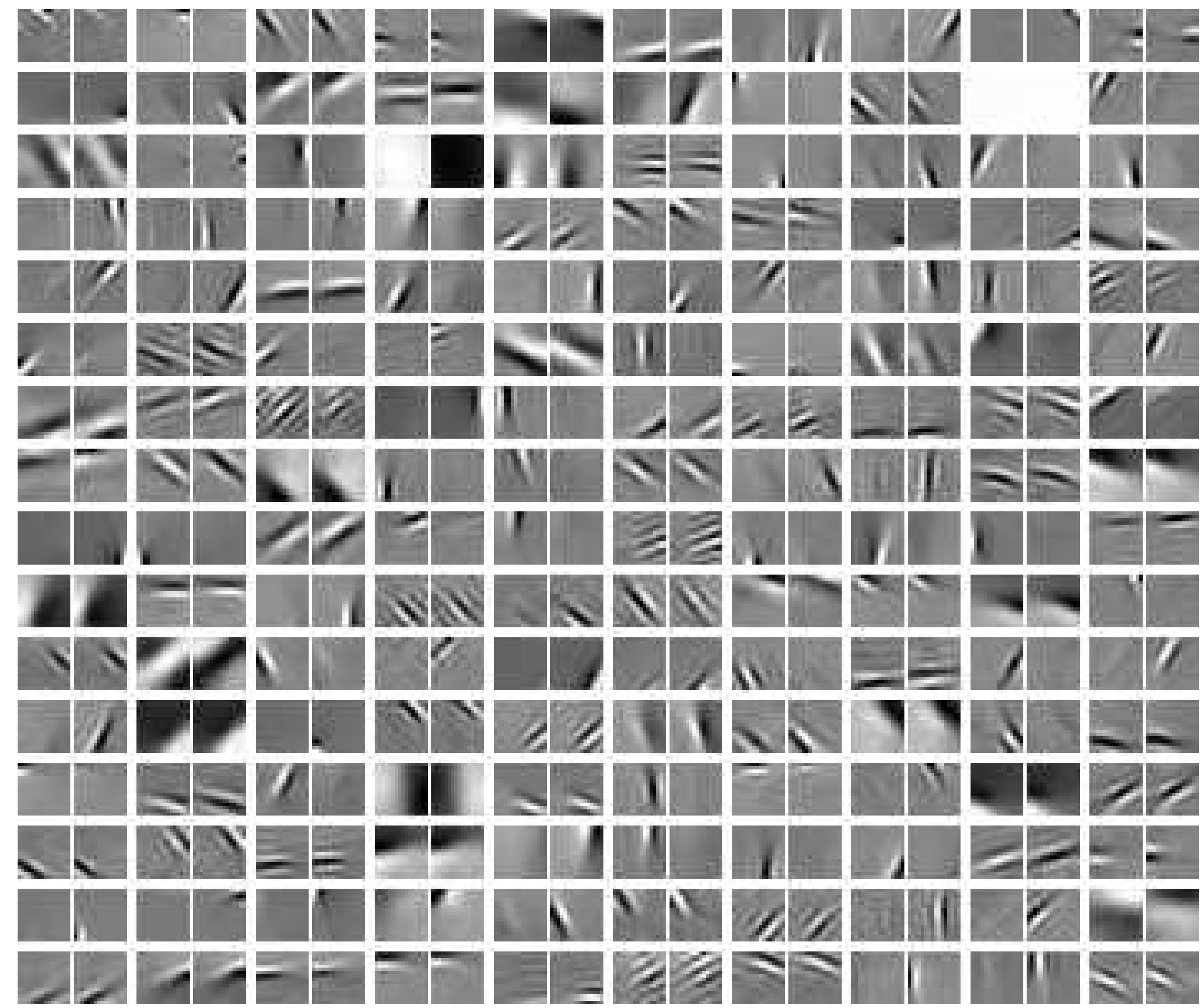


# Sparse coding and reconstruction



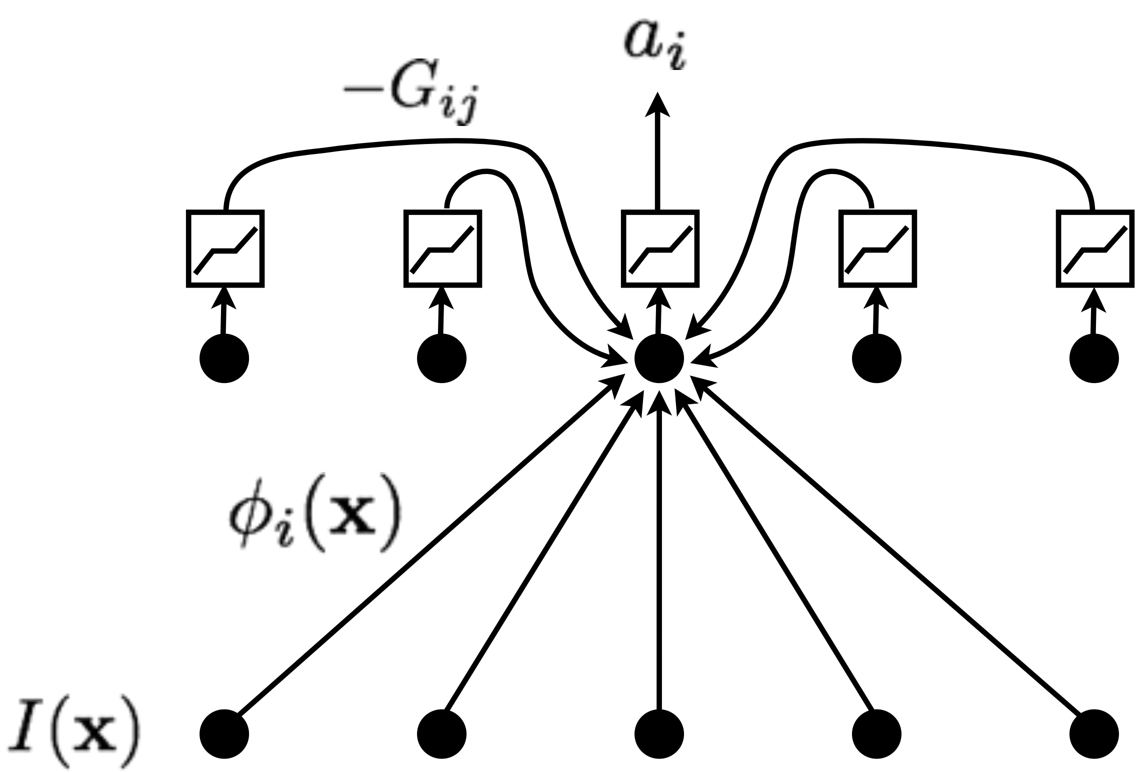
# Extensions to color, disparity

Wachtler, Lee and Sejnowski (2001), Hoyer & Hyvarinen (2000)

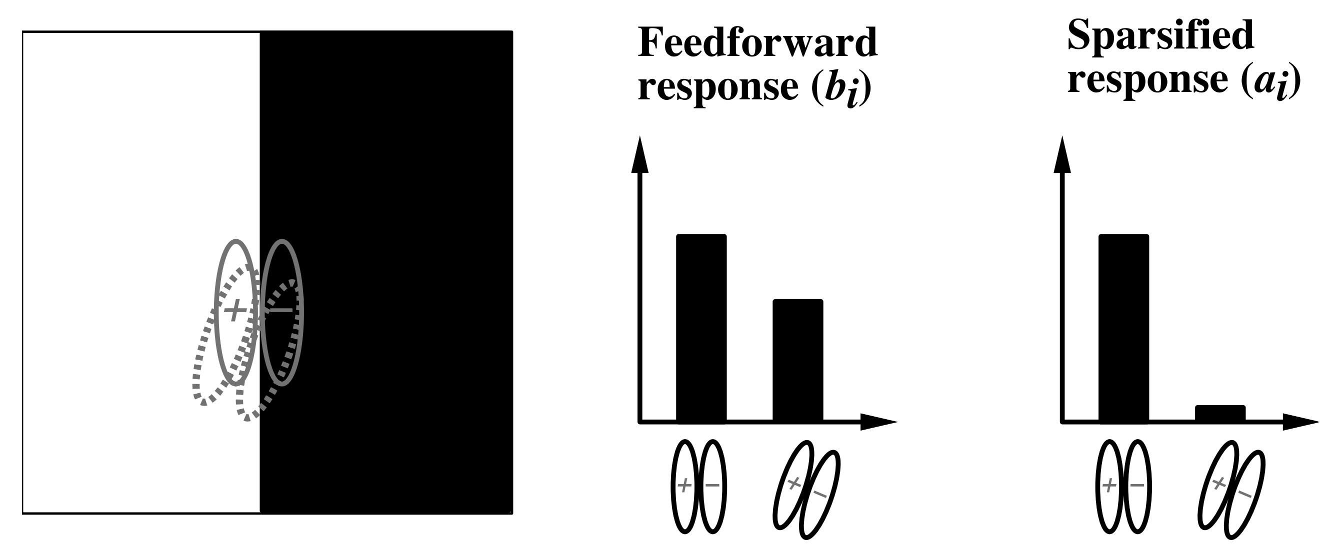



# Nonlinear encoding

Solutions may be computed by a network of leaky integrators and threshold units (Rozell et al. 2008)




‘Explaining away’

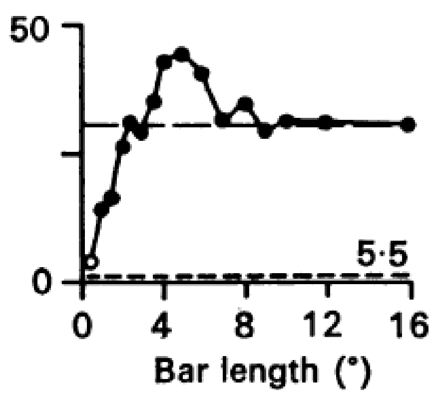


# Nonlinear encoding

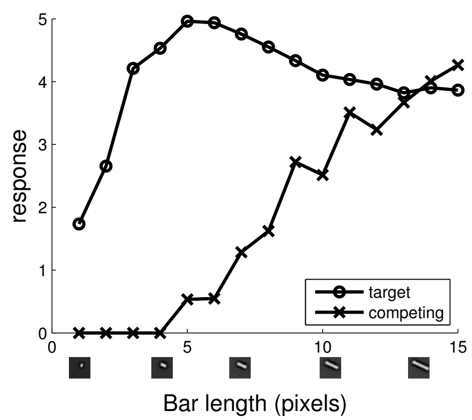
## ‘explaining away’ explains nCRF effects

## Lee et al. (2007), Zhu & Rozell (2010)

### ‘end-stopping’

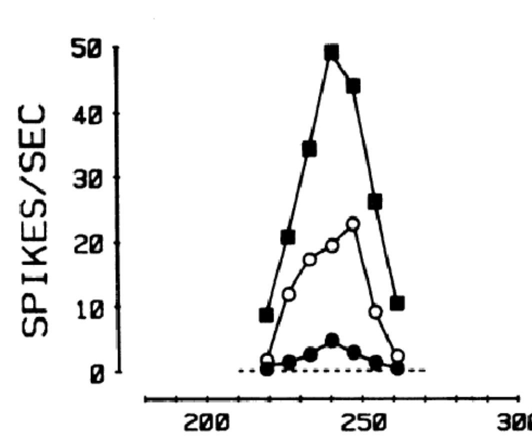



(a)

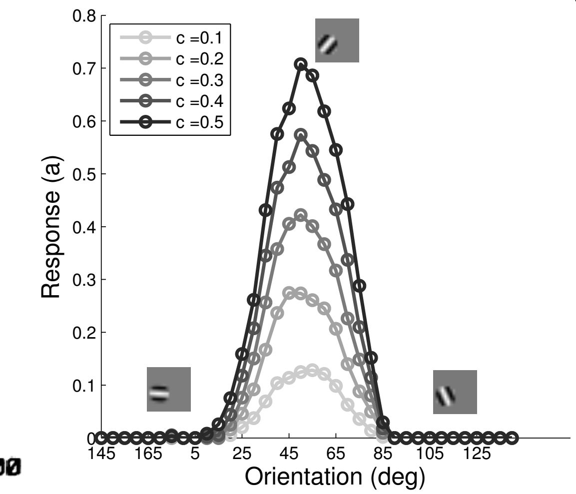



(b)

### contrast-invariant tuning

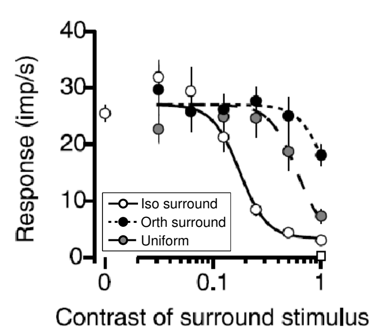



(a)

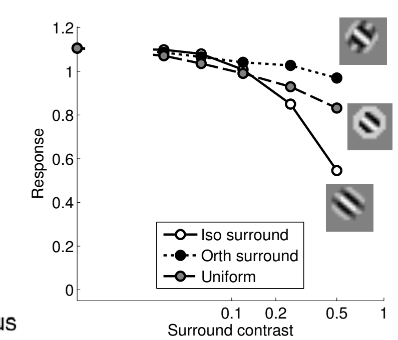



(b)

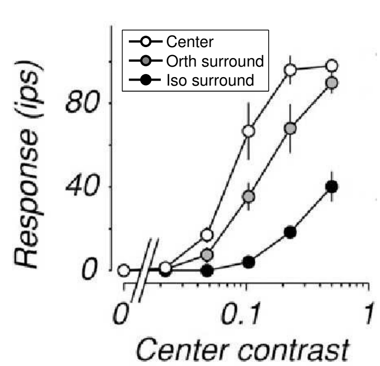
### ‘surround suppression’



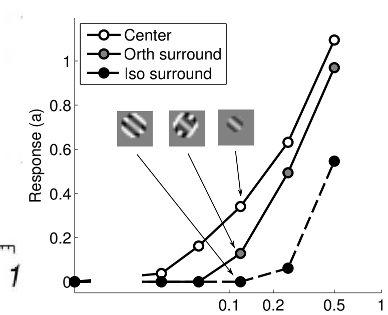

(a)




(b)




(c)




(d)

# Energy-based models

Osindero, Welling and Hinton (2005)

$$P(\mathbf{x}) = \frac{1}{Z} e^{-E(\mathbf{x})}$$

$$E(\mathbf{x}) = \sum_{i=1}^{M} \alpha_i \log \left( 1 + \frac{(\mathbf{J}_i\, \mathbf{x})^2}{2} \right)$$