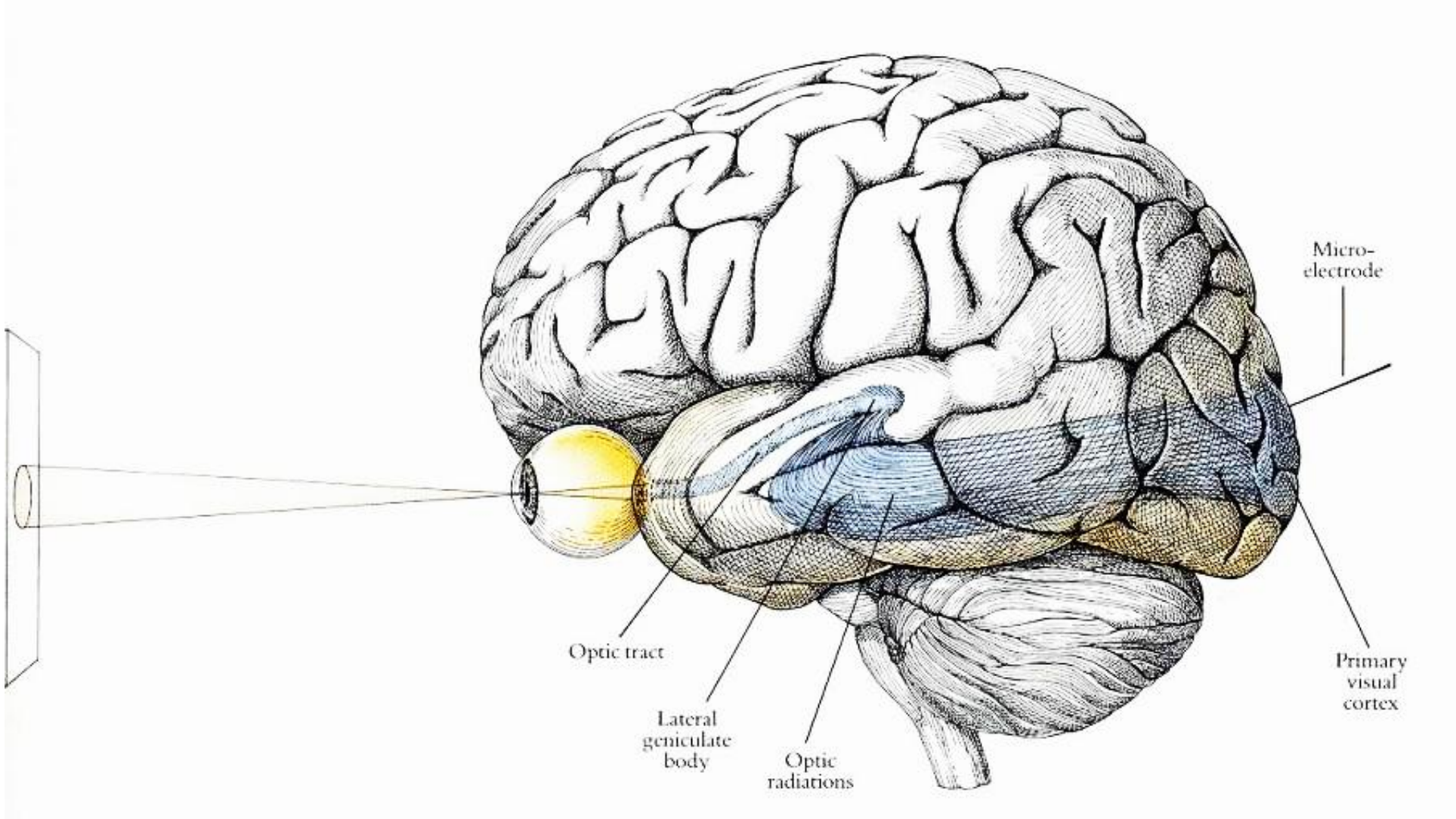
Natural image statistics and biological vision

Bruno A. Olshausen

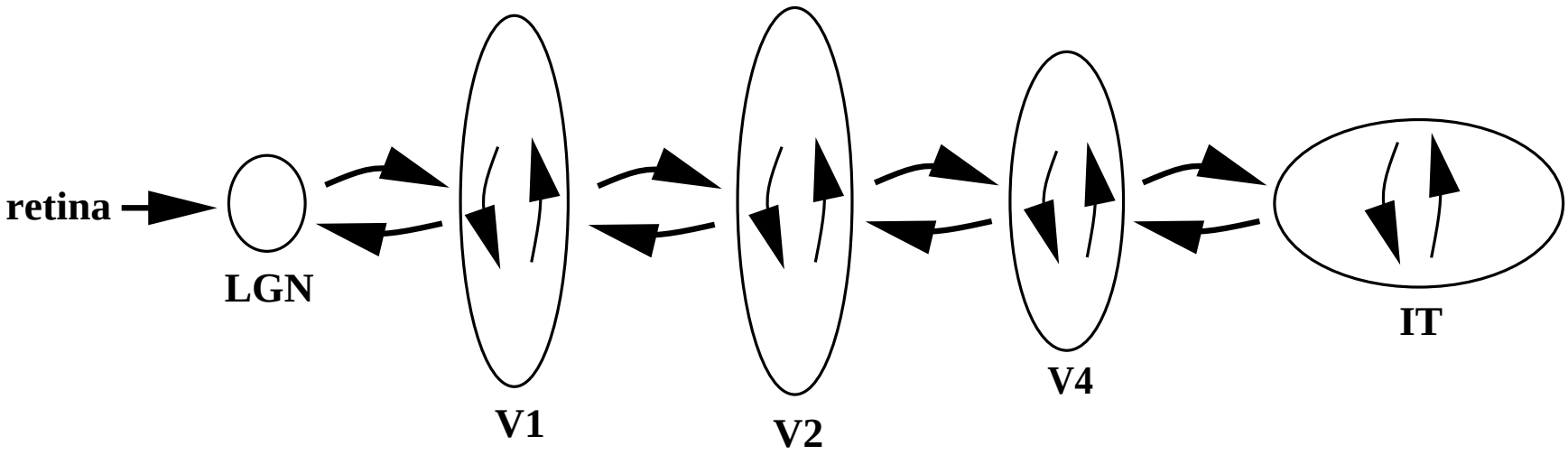
Center for Neuroscience, U.C. Davis

&

REDWOOD
Neuroscience Institute


Micro-
electrode
Optic tract
Lateral
geniculate
body
Optic
radiations
Primary
visual
cortex

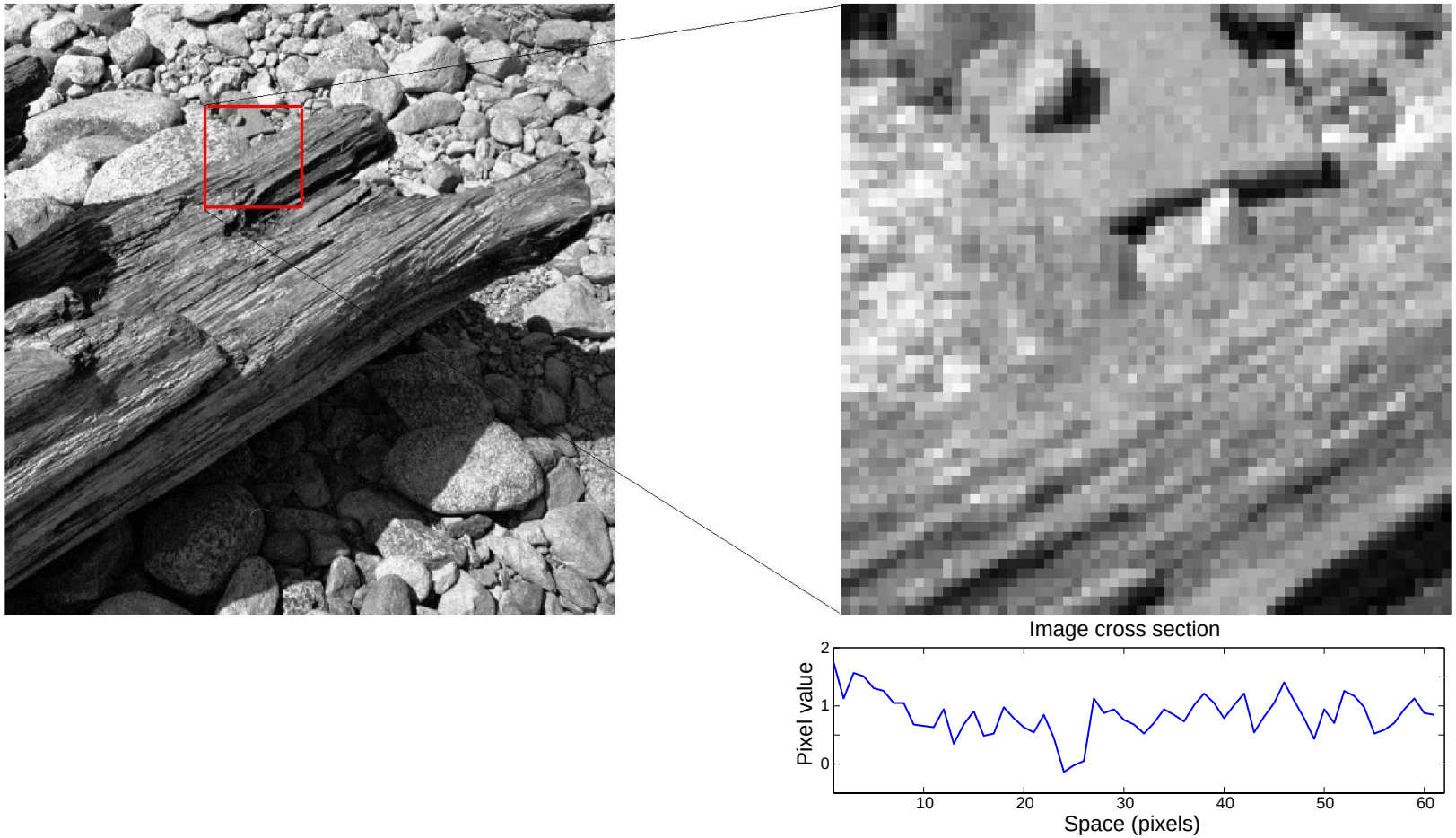
Recurrent computation is pervasive throughout cortex



Main Points

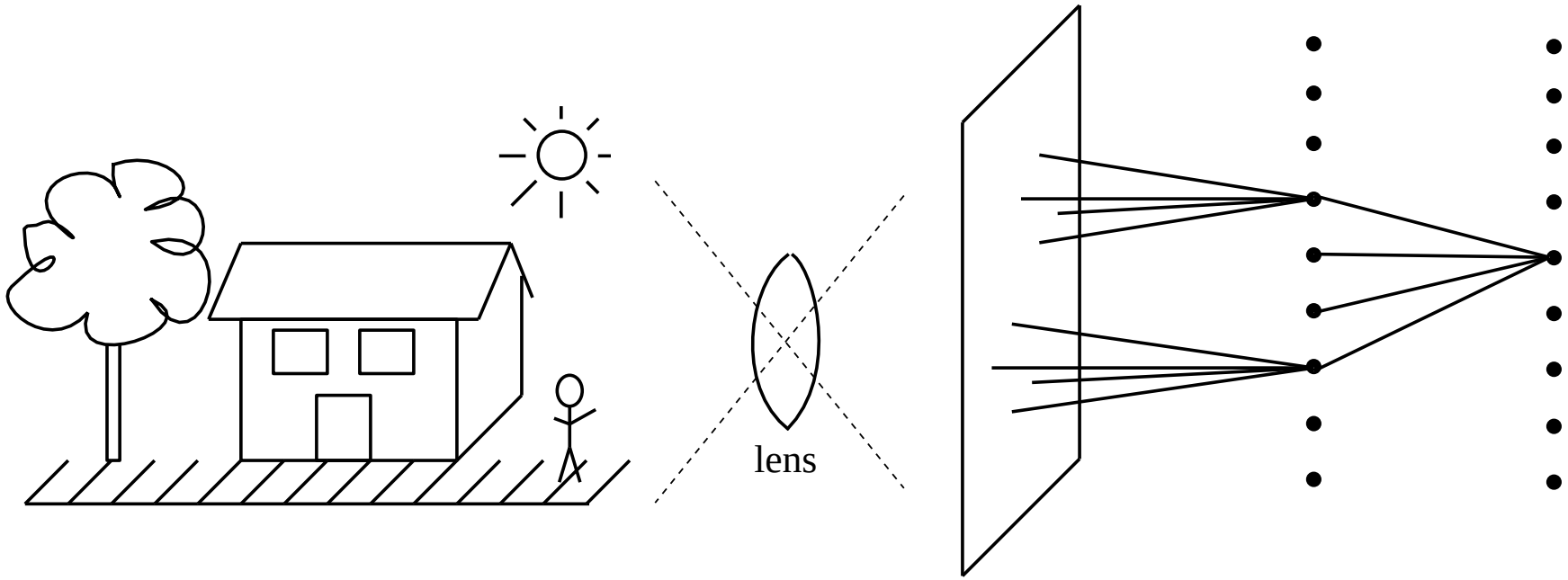
- Vision as inference
- Sparse, overcomplete representations
- Sparse coding in V1
- Learning invariances

The problem: scene analysis



How do you interpret an edge?

Vision as inference



Bayes' rule

$$P(E|D) \propto \underbrace{P(D|E)}_{\substack{\text{how data is} \\ \text{generated by} \\ \text{the environment}}} \times \underbrace{P(E)}_{\substack{\text{prior beliefs} \\ \text{about the} \\ \text{environment}}}$$

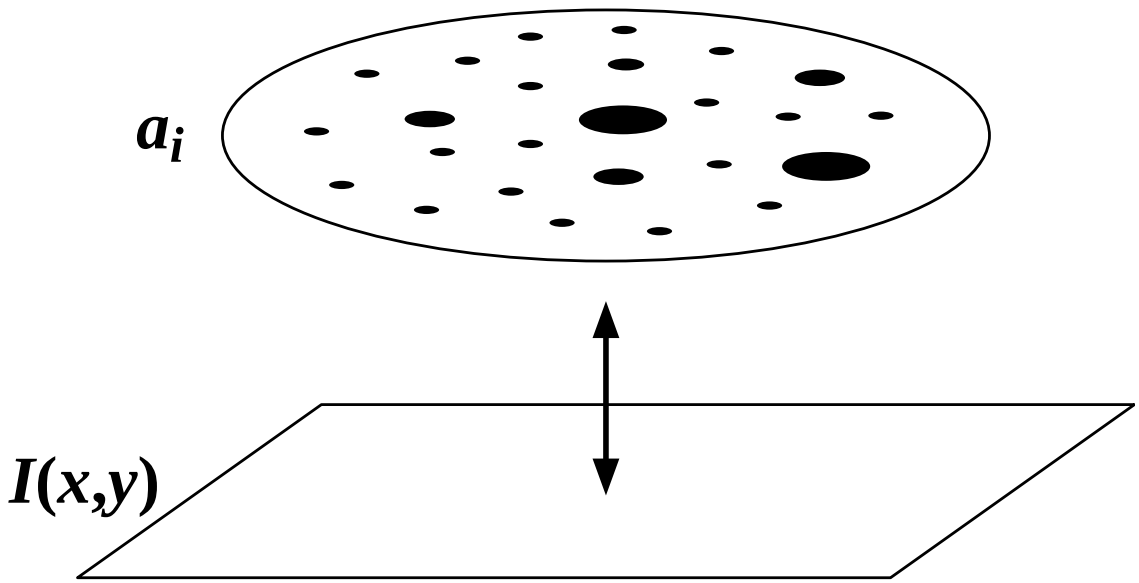
E = the actual state of the environment

D = data about the environment

Principles of cortical representation

- Sparseness
- Invariance
- Hierarchy and feedback

Sparse coding

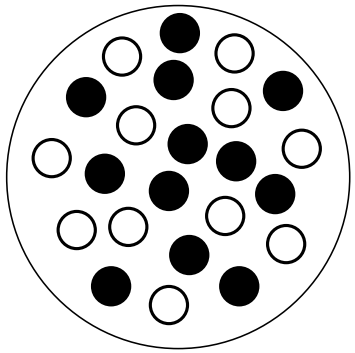


Why sparseness?

- Makes structure in data explicit.
- Makes it easier to do pattern matching.
- Increases storage capacity in associative memory models.
- Sparse codes are energy efficient.

Dense codes vs. local codes

Dense codes
(ascii)

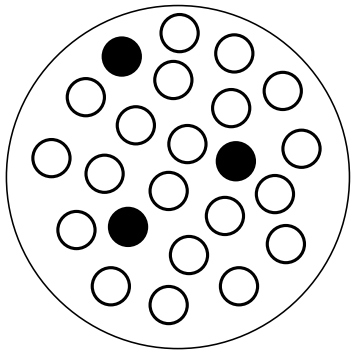


+ High combinatorial capacity (2^N)

- Difficult to read out

. . .

Sparse, distributed codes

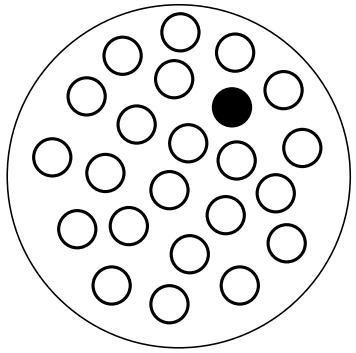


+ Decent combinatorial capacity ($\sim N^K$)

+ Still easy to read out

. . .

Local codes
(grandmother cells)



- Low combinatorial capacity (N)

+ Easy to read out

Evidence for sparse coding

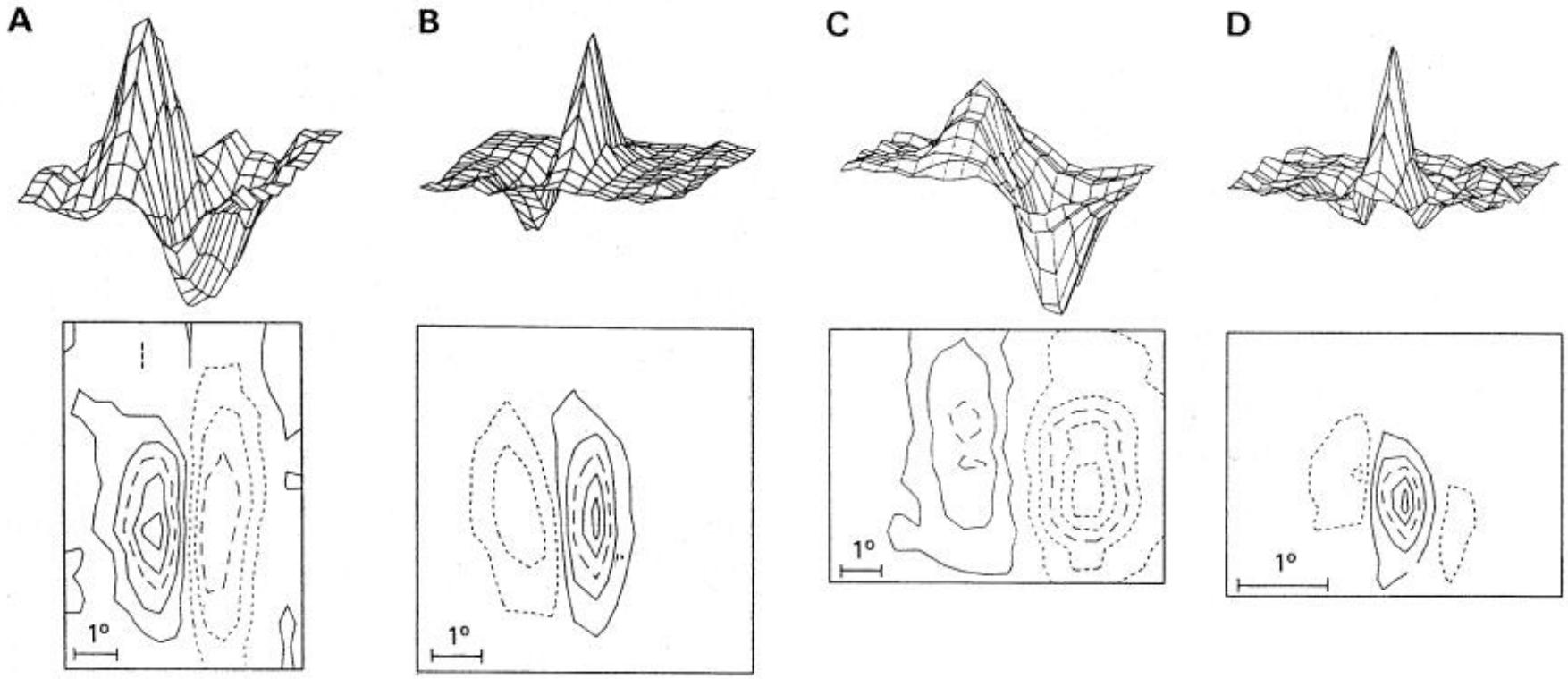
- Gilles Laurent - mushroom body, insect
- Michael Fee - HVC, zebra finch
- Tony Zador - auditory cortex, mouse
- Bill Skaggs - hippocampus, primate
- Harvey Swadow - motor cortex, rabbit
- Michael Brecht - barrel cortex, rat
- Jack Gallant - visual cortex, macaque monkey
- Christof Koch - inferotemportal cortex, human

See:

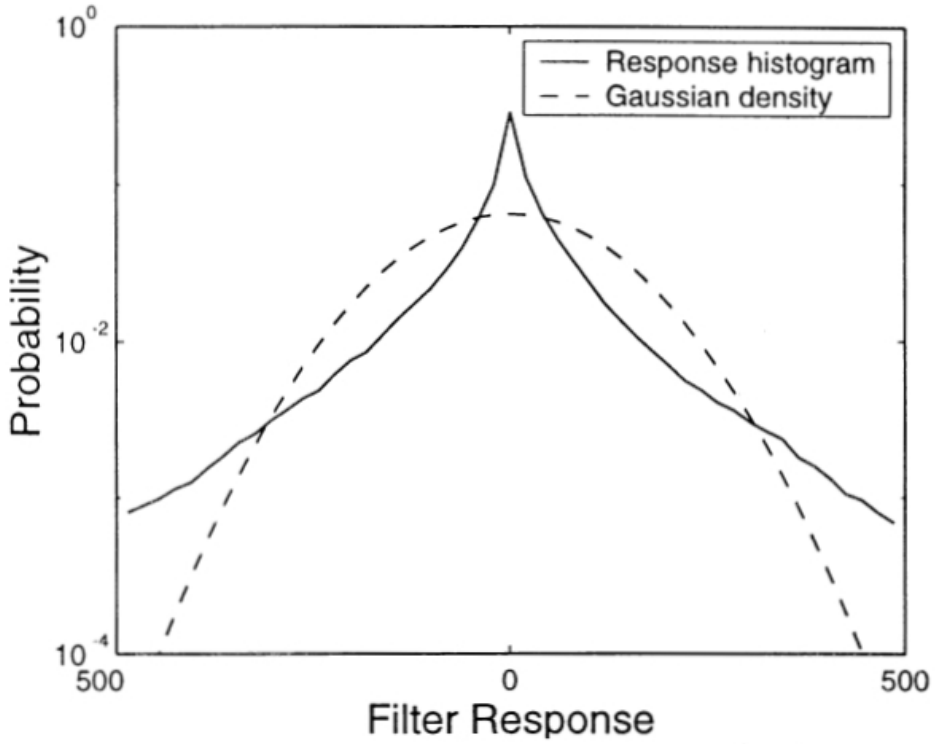
Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. *Current Opinion in Neurobiology, 14*, 481-487.


California
SPARSE
SESQUICENTENNIAL - 150 YEARS

Simple cell receptive fields (Jones & Palmer, 1987)



Gabor-filter histogram



Overcomplete representations

- In oriented, multiscale pyramids, overcompleteness is necessary to ascribe meaning to coefficients (Simoncelli, Freeman, Adelson, and Heeger, 1992).

- Overcomplete time-frequency dictionaries are best able to reveal time-frequency structure embedded in signals (Chen, Donoho, Saunders, 2001).

- Area V1 is highly overcomplete, by approximately 25:1 (in cat).

Image model

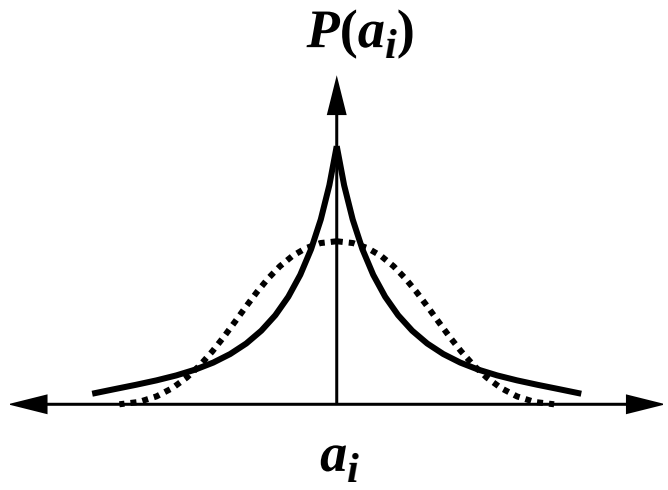
$$I(x,y) = \sum_i a_i \, \phi_i(x,y) + \nu(x,y) \; .$$

Goal: Find a set of basis functions $\{\phi_i\}$ for representing natural images such that the coefficients a_i are as sparse and statistically independent as possible.

Prior

- Factorial: $P(\mathbf{a}|\theta) = \prod_i P(a_i|\theta)$

- Sparse: $P(a_i|\theta) = \frac{1}{Z_S} e^{-S(a_i)}$



Inference (perception)

MAP estimate:

$$\hat{\mathbf{a}} = \arg\max_{\mathbf{a}} P(\mathbf{a}|\mathbf{I}, \theta)$$

$$P(\mathbf{a}|\mathbf{I}, \theta) \propto P(\mathbf{I}|\mathbf{a}, \theta)\, P(\mathbf{a}|\theta)$$

Energy function:

$$\begin{aligned} E(\mathbf{I}, \mathbf{a}) &= -\log P(\mathbf{a}|\mathbf{I}, \theta) \\ &= \frac{\lambda_N}{2}|\mathbf{I} - \Phi\mathbf{a}|^2 + \sum_i S(a_i) + \text{const.} \end{aligned}$$

Dynamics:

$$\begin{aligned} \dot{\mathbf{a}} &\propto -\frac{\partial E}{\partial \mathbf{a}} \\ &= \lambda_N\, \Phi^T\, \mathbf{I} - \lambda_N \Phi^T \Phi\, \mathbf{a} - S'(\mathbf{a}) \end{aligned}$$

Learning

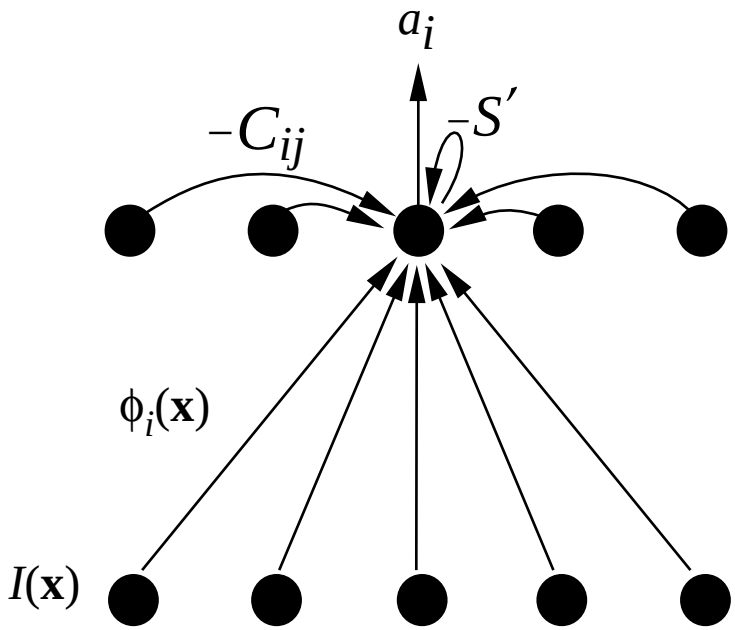
Objective function:

$$\begin{aligned} \mathcal{L} &= \langle \log P(\mathbf{I}|\theta) \rangle \\ P(\mathbf{I}|\theta) &= \int P(\mathbf{I}|\mathbf{a},\theta)\, P(\mathbf{a}|\theta)\, d\mathbf{a} \end{aligned}$$

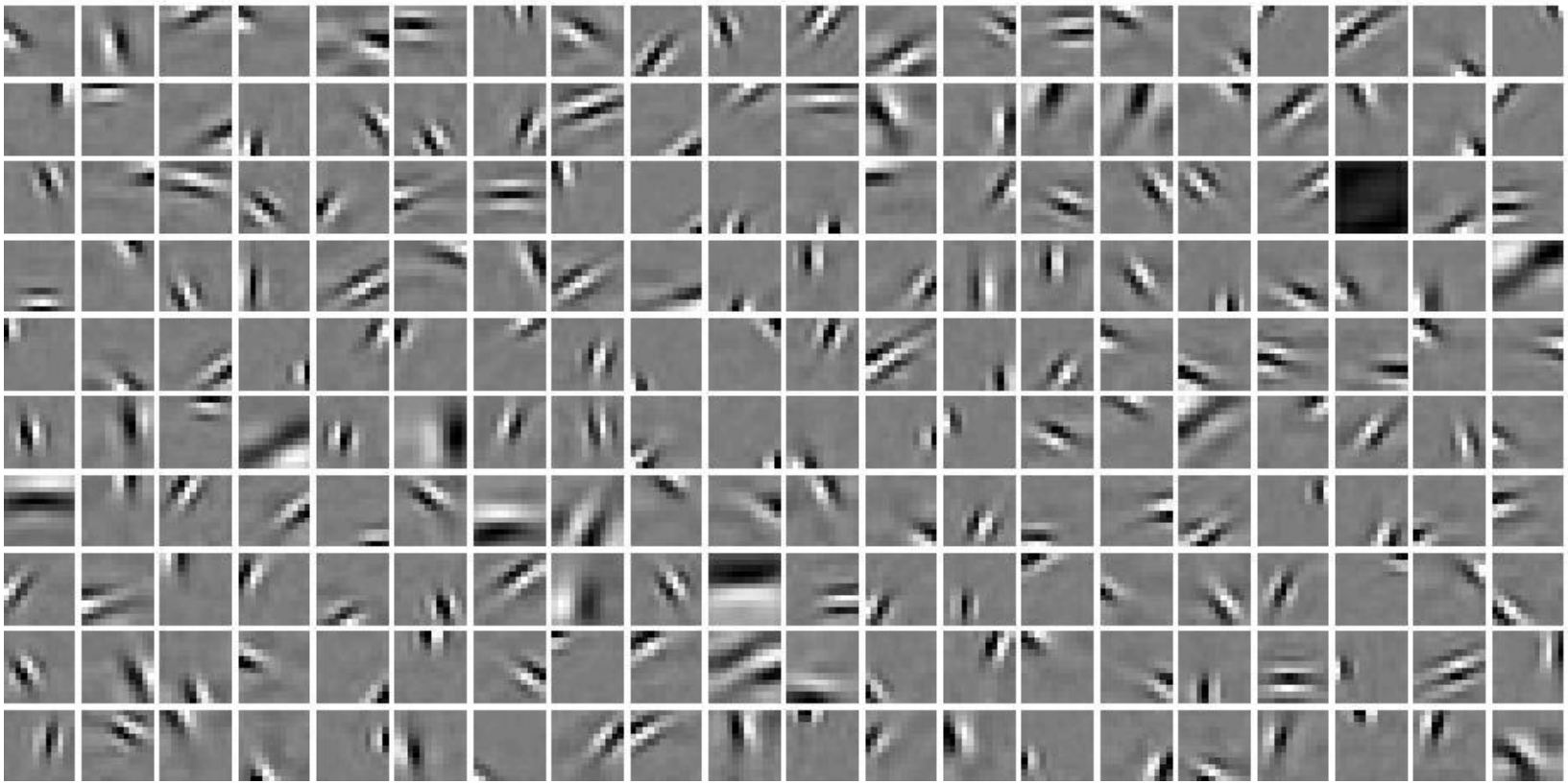
Learning rule:

$$\begin{aligned} \Delta\Phi &\propto \frac{\partial \mathcal{L}}{\partial \Phi} \\ &= \lambda_N \int [I - \Phi\, \mathbf{a}]\; P(\mathbf{a}|\mathbf{I},\theta)\, d\mathbf{a} \end{aligned}$$

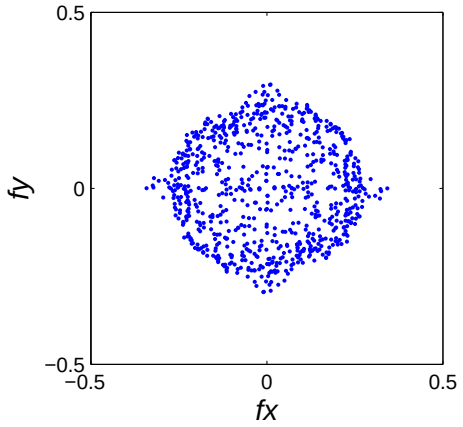
Network implementation

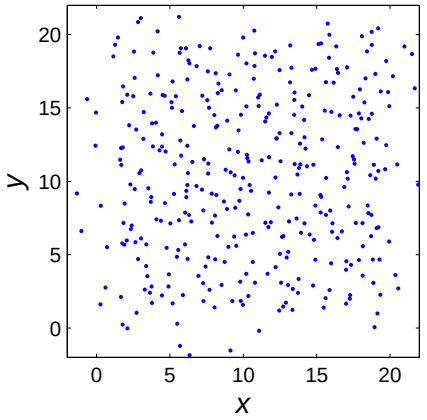


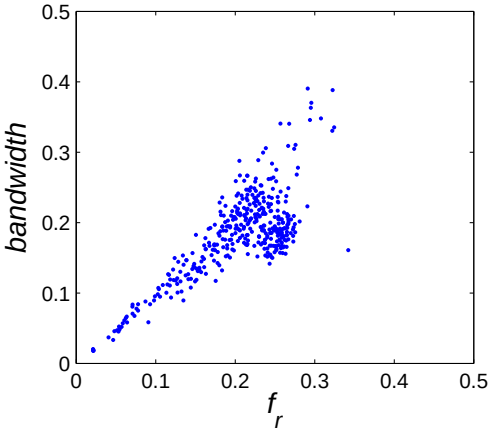
Learned basis functions (200, 12x12)

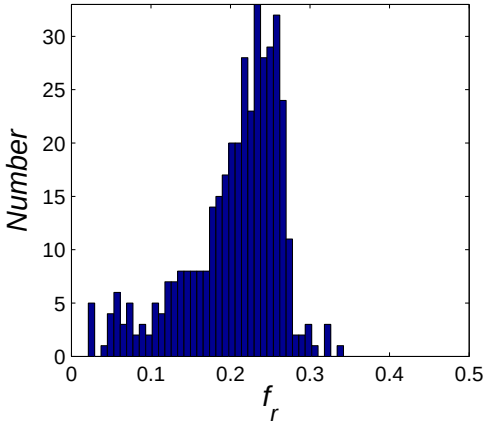


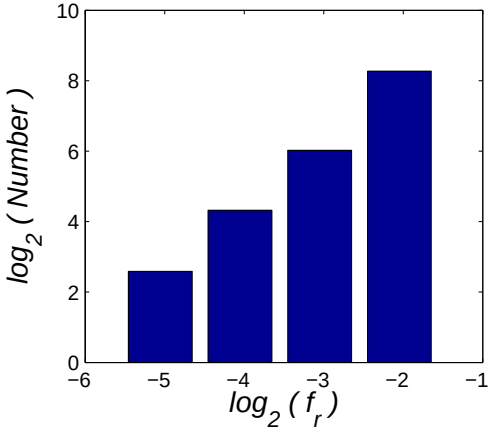
Tiling properties

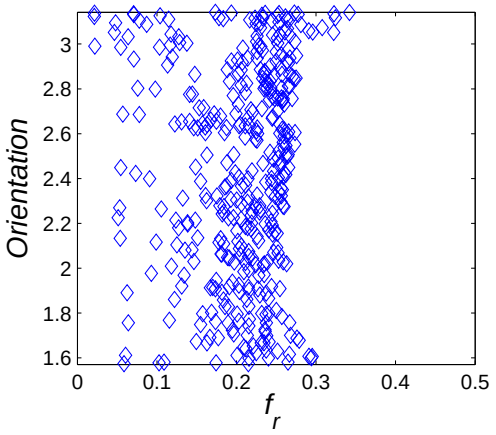






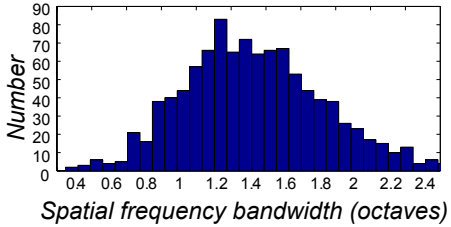




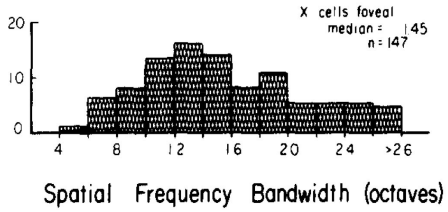


Spatial-frequency bandwidth

Model:

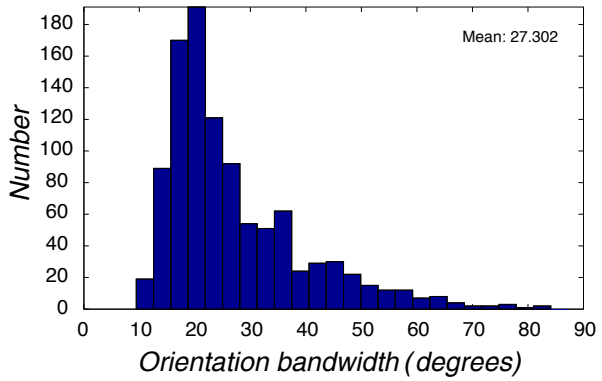


Physiology (DeValois lab):

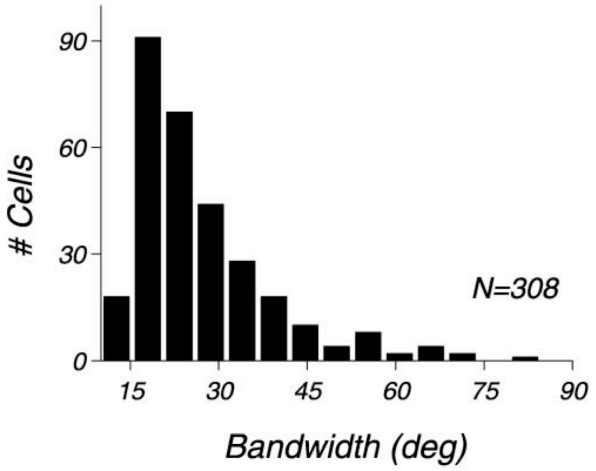


Orientation bandwidth

Model:

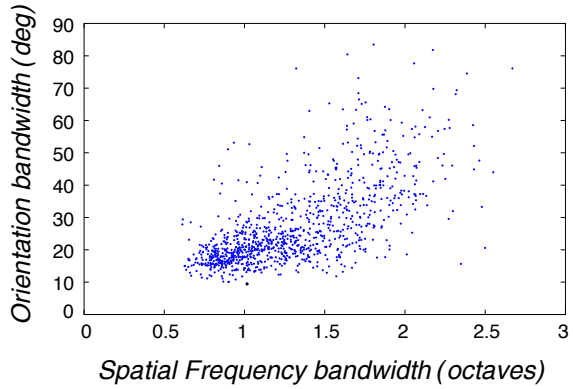


Physiology (Shapley lab):

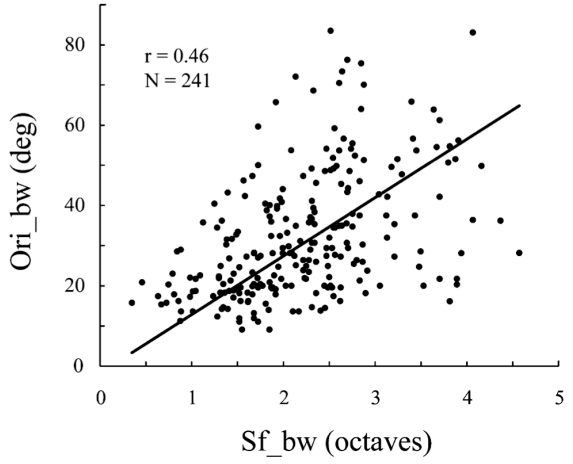


Orientation bandwidth vs. spatial-frequency bandwidth

Model:

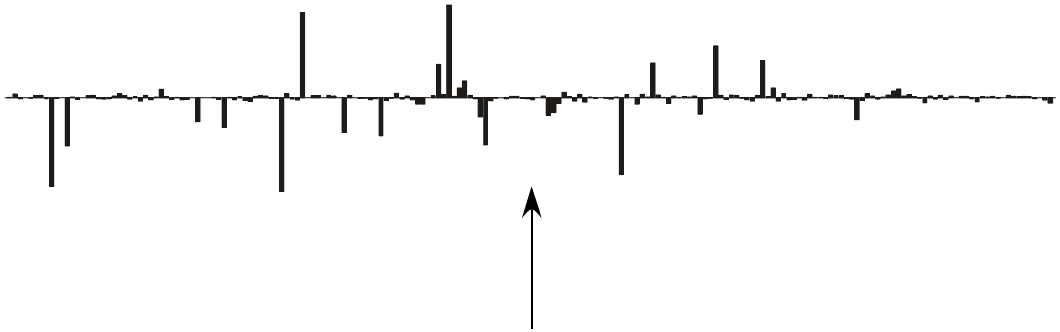


Physiology (Shapley lab):



Sparsification

Outputs of sparse coding network (a_i)



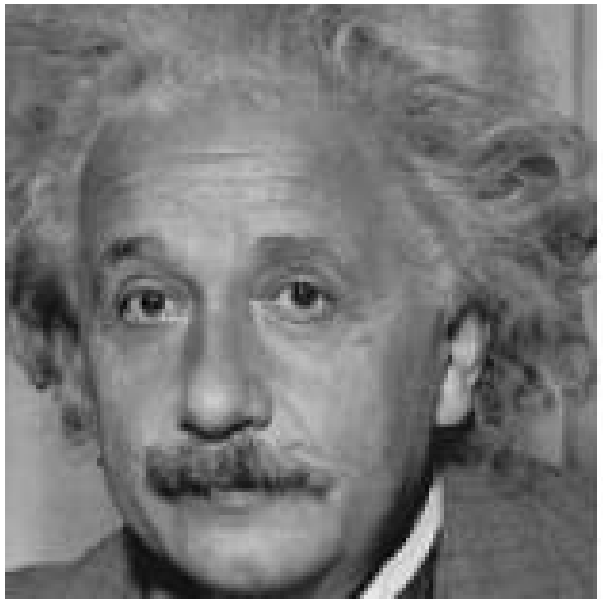
Pixel values

Image $I(x,y)$

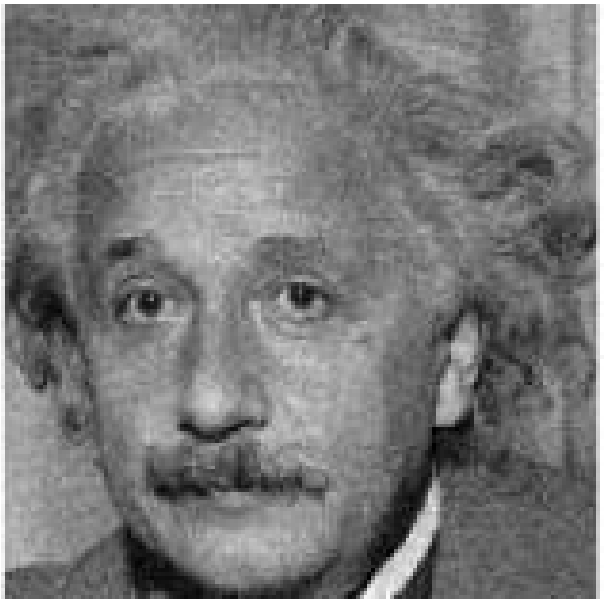
Denoising

Aoccdrnig to rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a total mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.

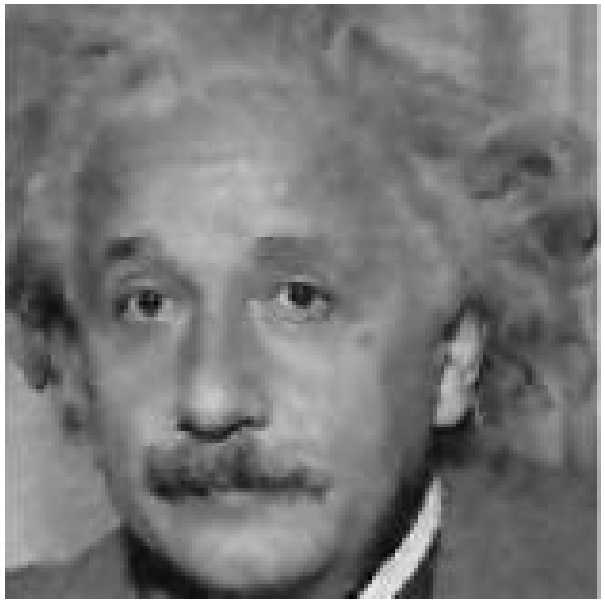
original



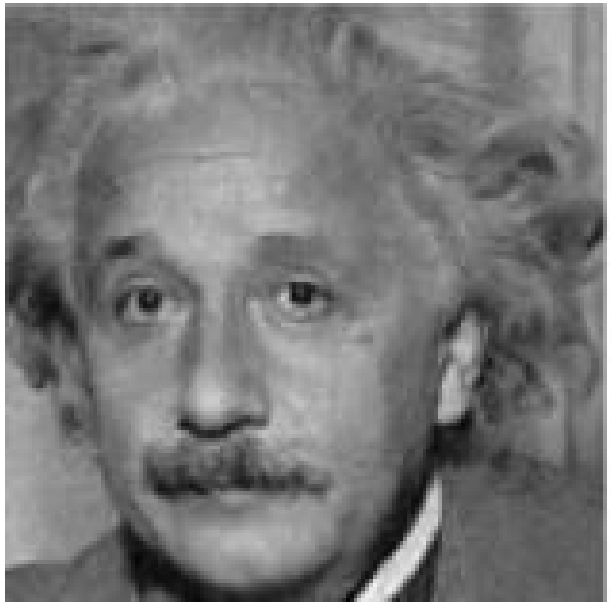
noisy (σ=10) SNR=12.3983



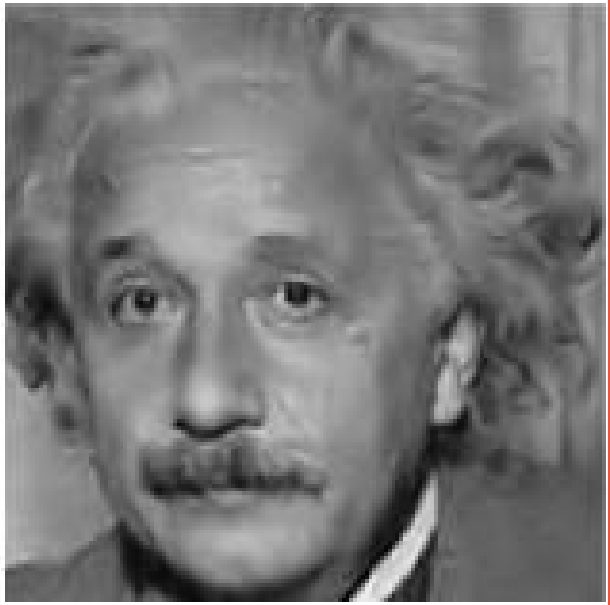
wiener2 SNR=15.8033



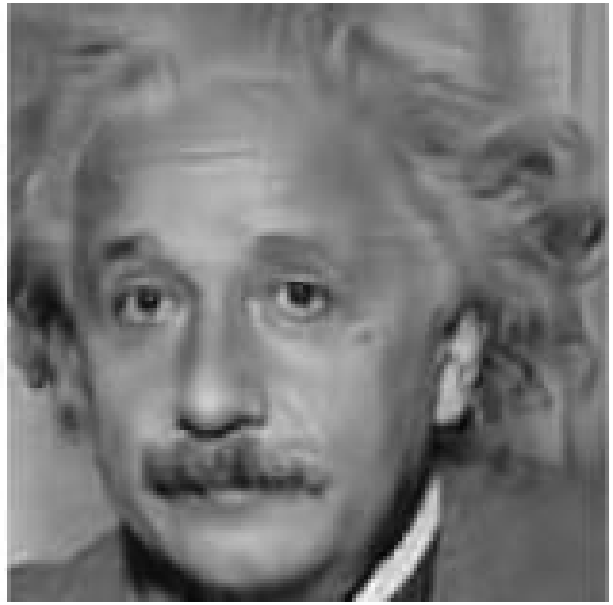
BayesCore steer6 SNR=16.3591



D+G steer6 SNR=16.4714



D+G learned6 SNR=16.1939



Applications

Denoising:

$$I(x,y) = \sum_i a_i \, \phi_i(x,y) + \nu(x,y)$$

Deblurring:

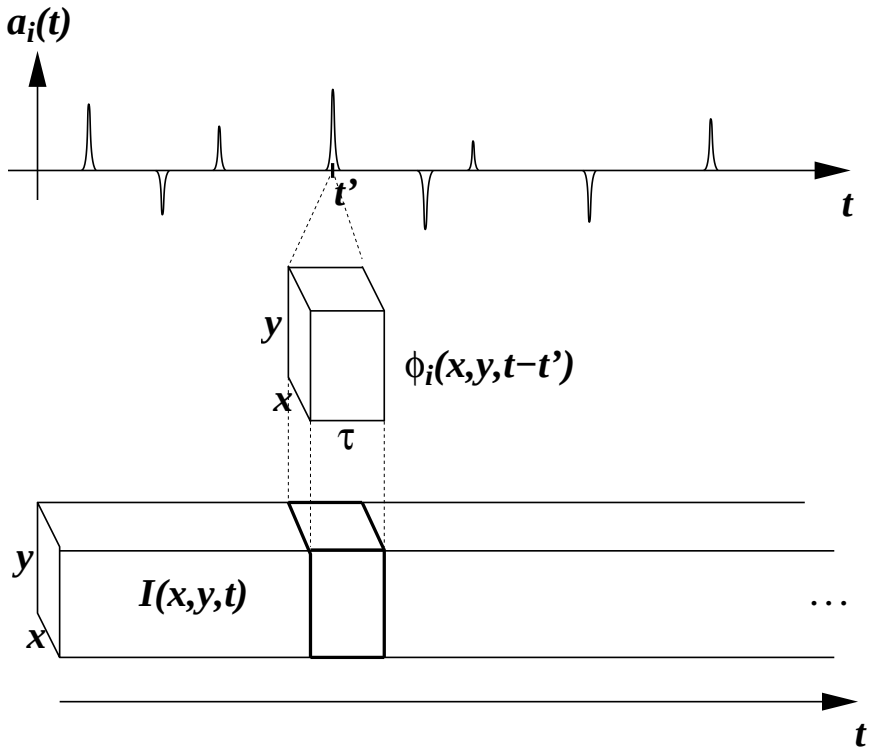
$$I(x,y) = B(x,y) * \sum_i a_i \, \phi_i(x,y) + \nu(x,y)$$

Filling-in:

$$\nu(x,y) \sim \begin{cases} \mathcal{N}(0, 1/\lambda_N) & x, y \ \epsilon \ \text{valid data} \\ \mathcal{N}(0, \infty) & x, y \ \epsilon \ \text{missing data} \end{cases}$$

Space-time image model

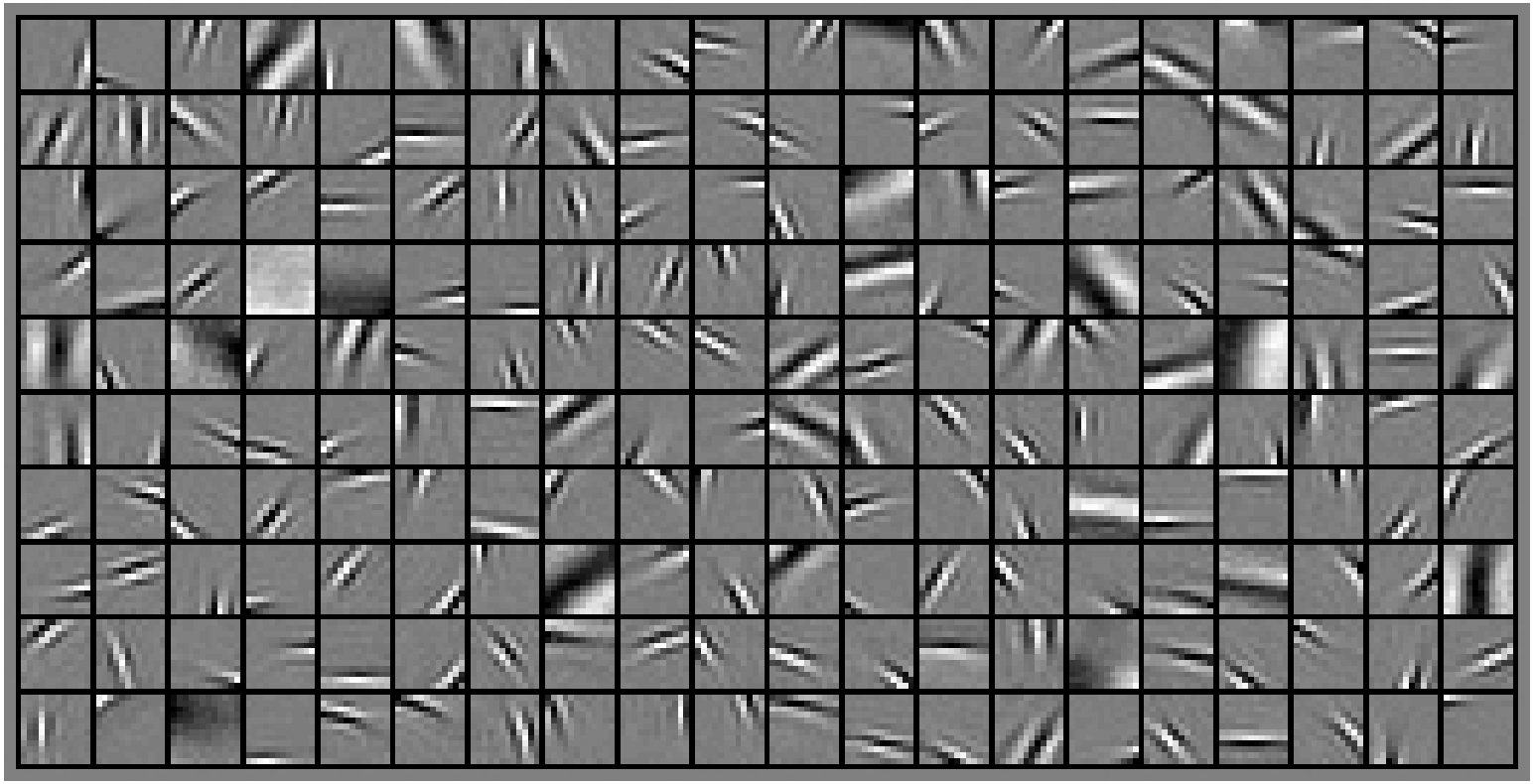
$$I(x,y,t) = \sum_i a_i(t) * \phi_i(x,y,t) + \nu(x,y,t)$$



Goal: Find a set of space-time basis functions $\{\phi_i\}$ for representing natural images such that the *time-varying* coefficients $a_i(t)$ are as sparse and statistically independent as possible *over both space and time*.

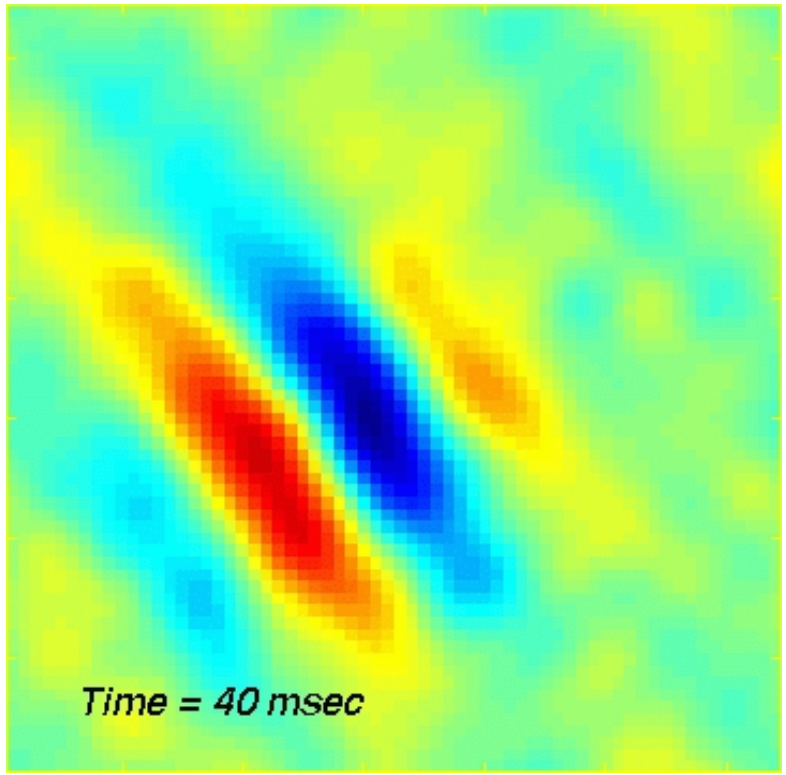
Learned space-time basis functions (200, $12 \times 12 \times 7$)

Training set: nature documentary

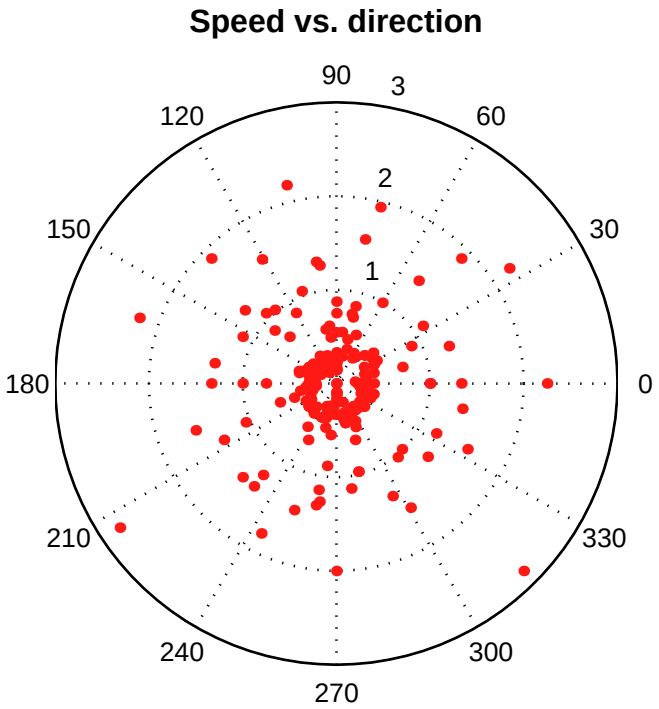


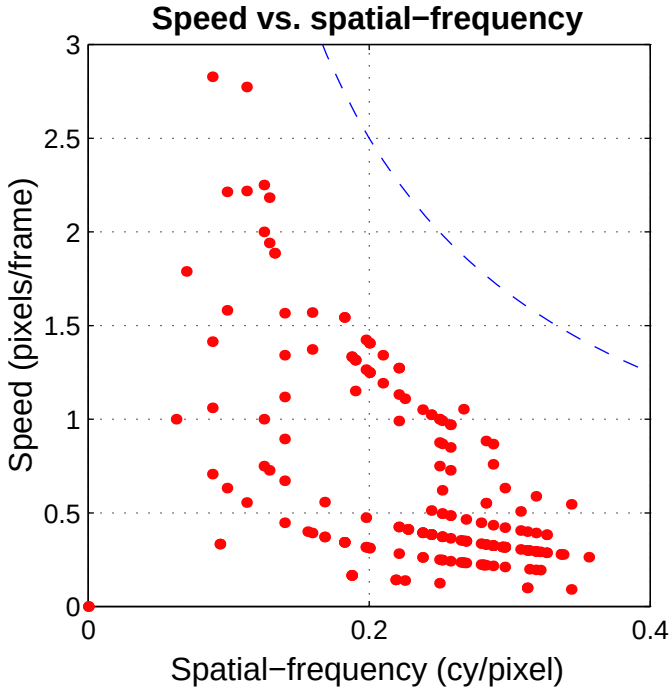
V1 space-time receptive field

(Courtesy of Dario Ringach)

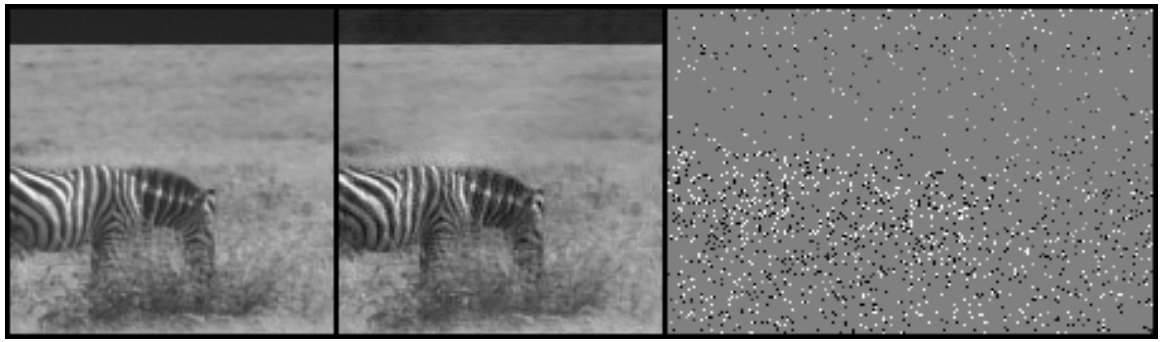


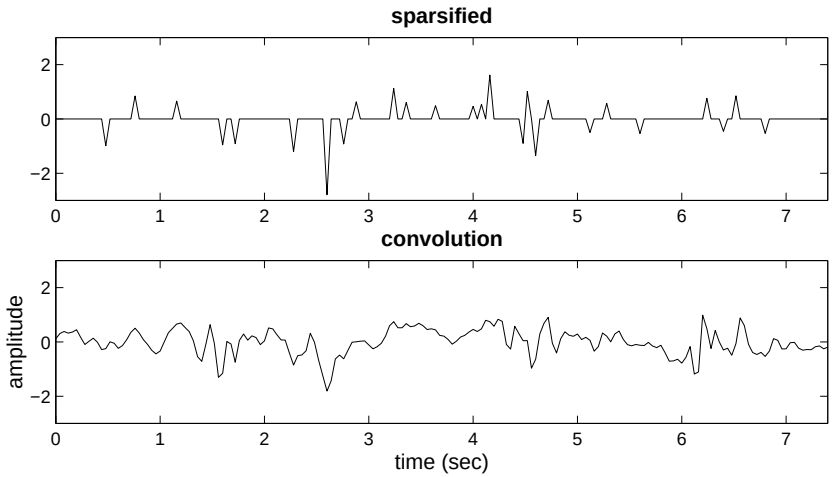
Basis function properties



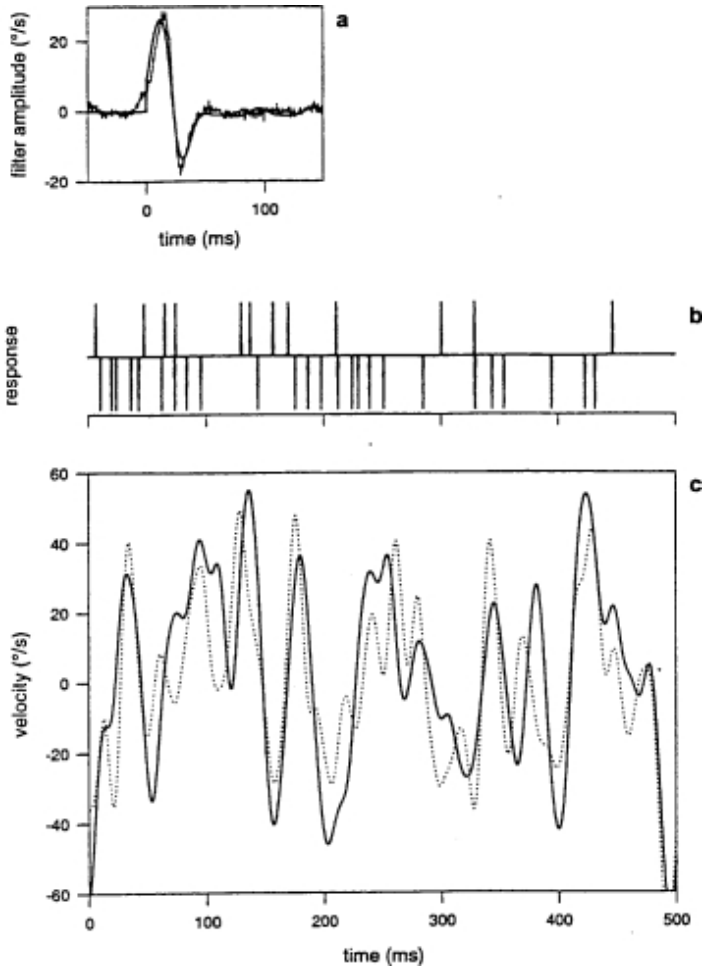


Spike encoding and reconstruction





Sparse codes and spikes

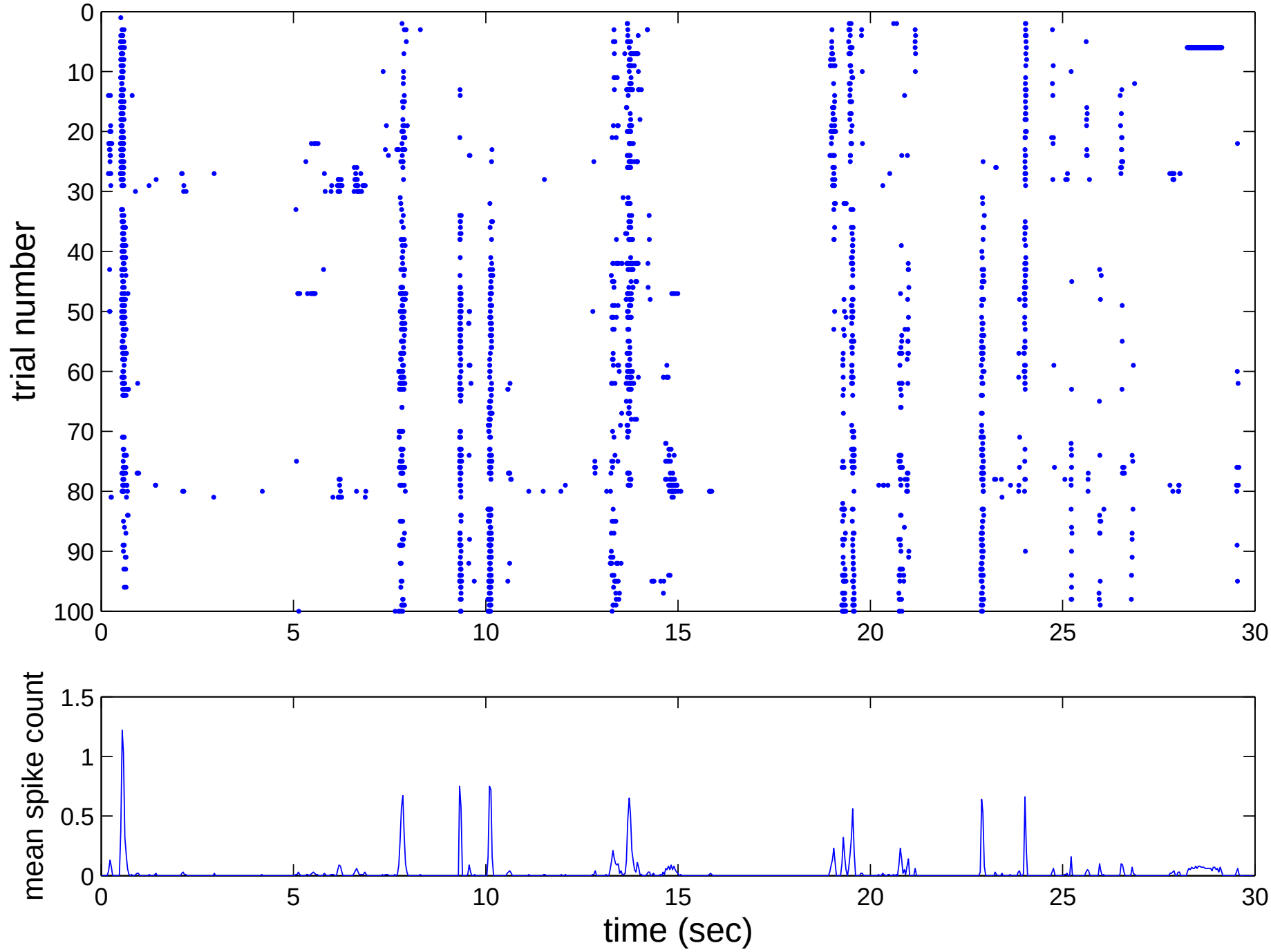


From Rieke et al., "Spikes" (1997)

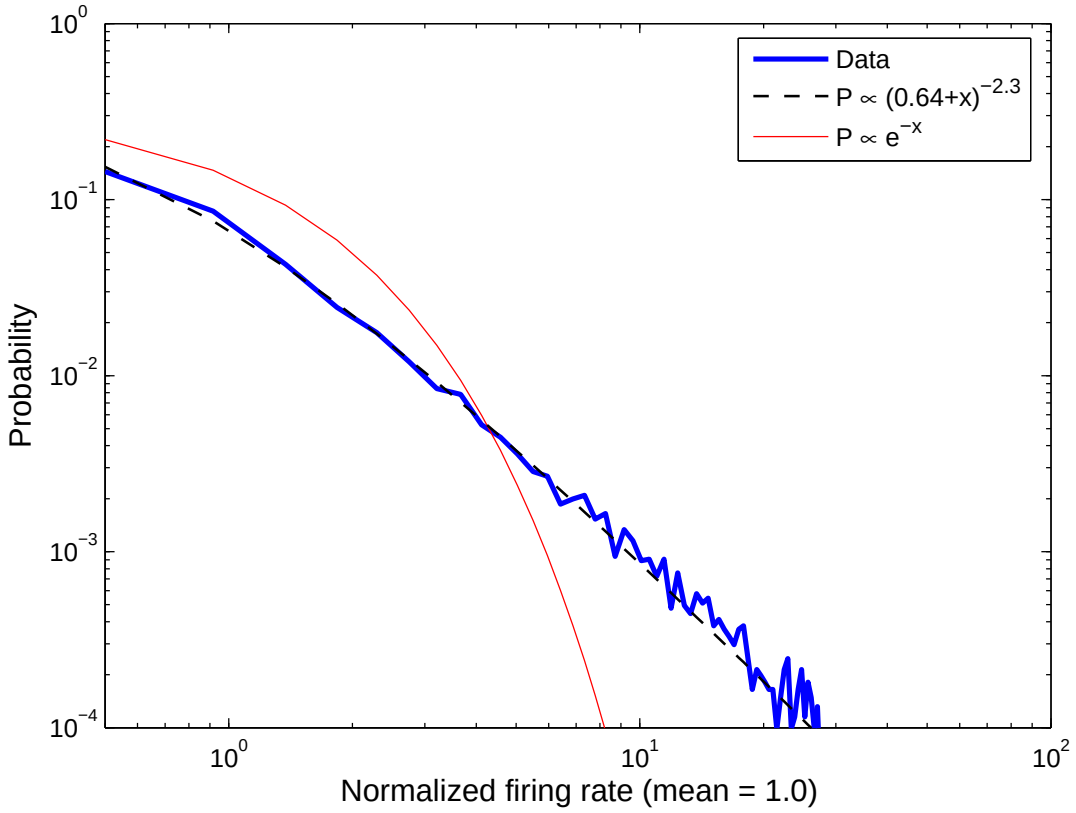
$$S_{\text{est}} = \rho(t) * K(t)$$

"...it seems clear that—at least under some conditions—many neurons make use of sparse coding in the time domain"

Cat V1 - natural movies (J. Baker, S.C. Yen, C.M. Gray, MSU Bozeman)

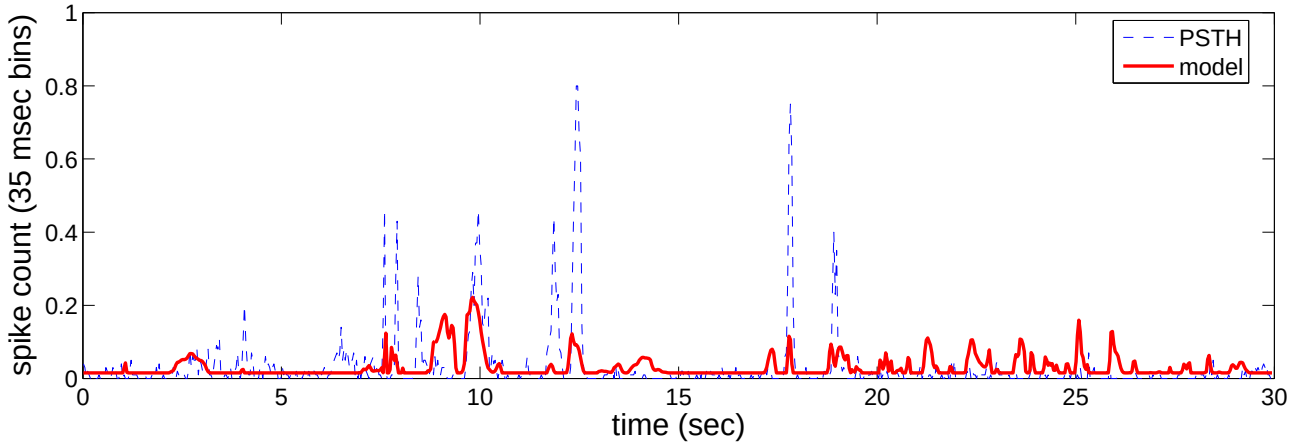


Firing rate distribution is power-law

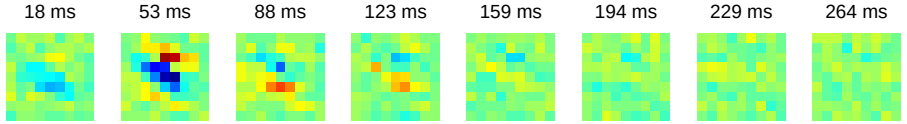


Responses are not well-predicted from RF filter models

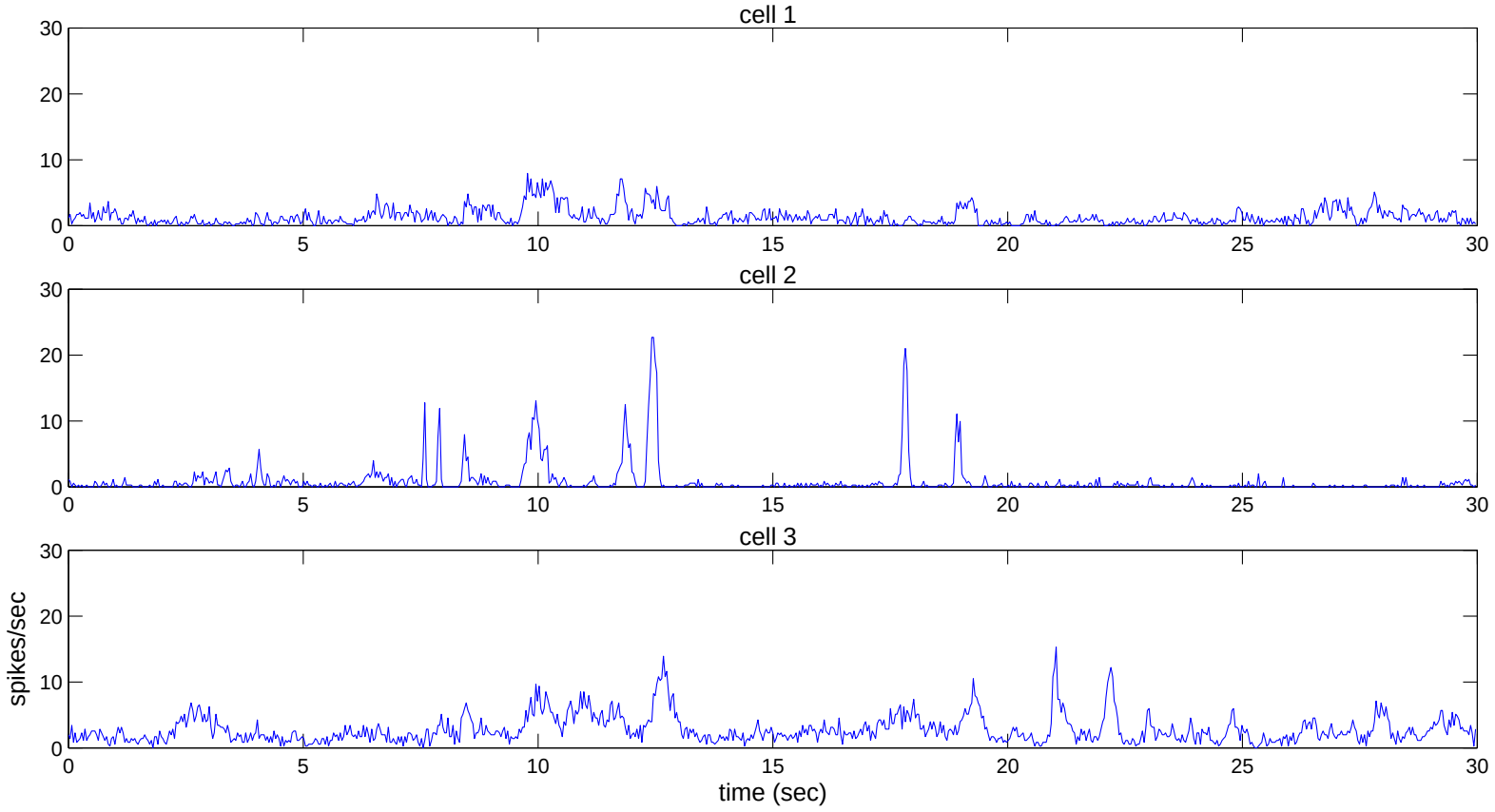
Data from Gray lab (J. Baker and S. Yen)



Receptive field:

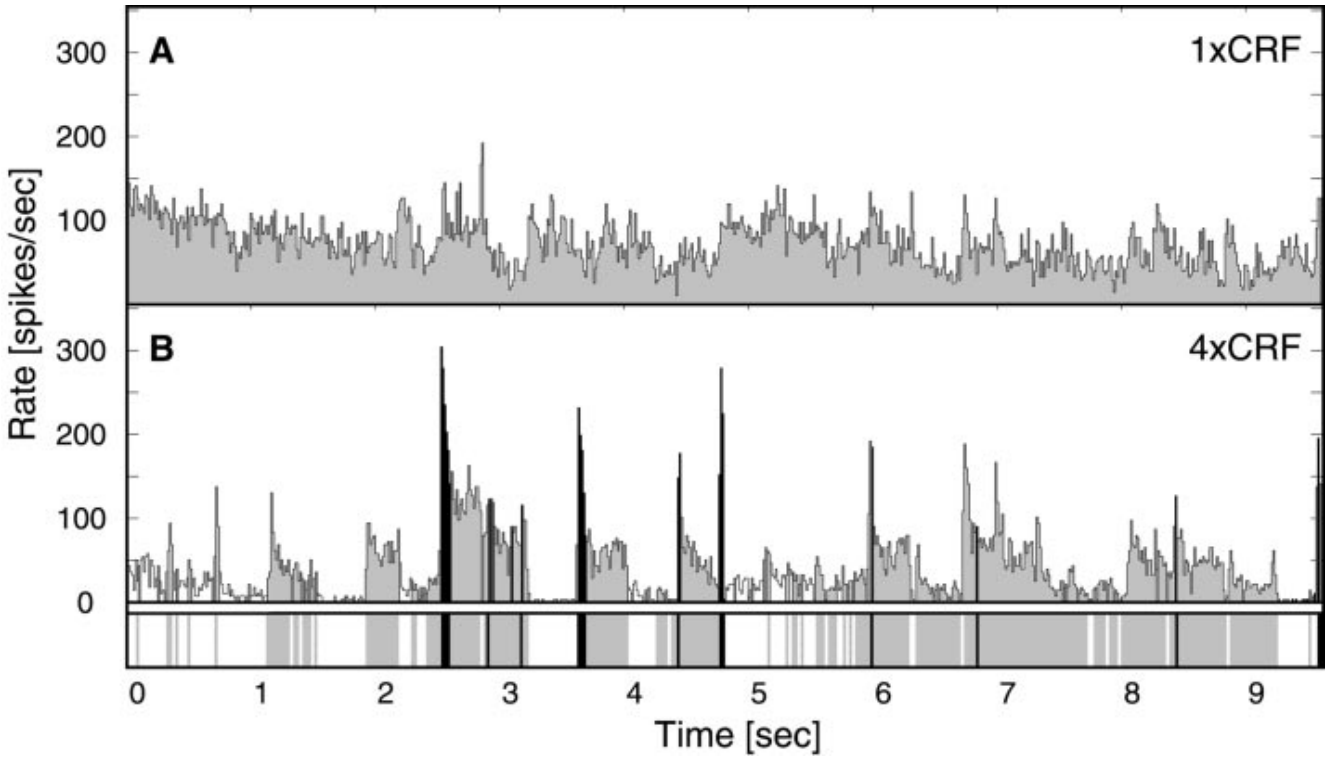


Responses of nearby units are heterogeneous



Context in natural scenes sparsifies responses

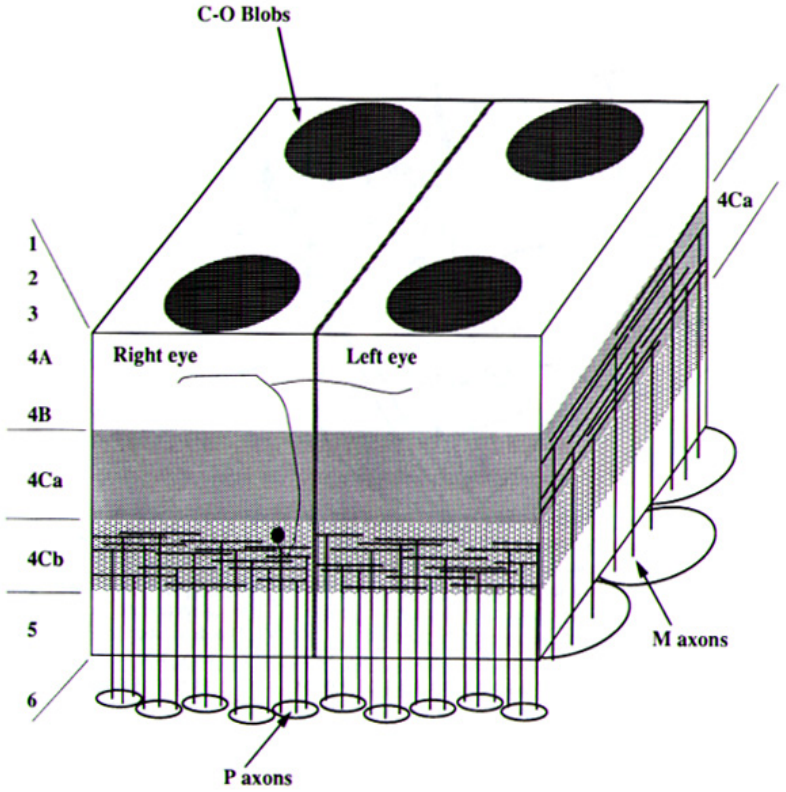
Vinje & Gallant (2000, 2002)

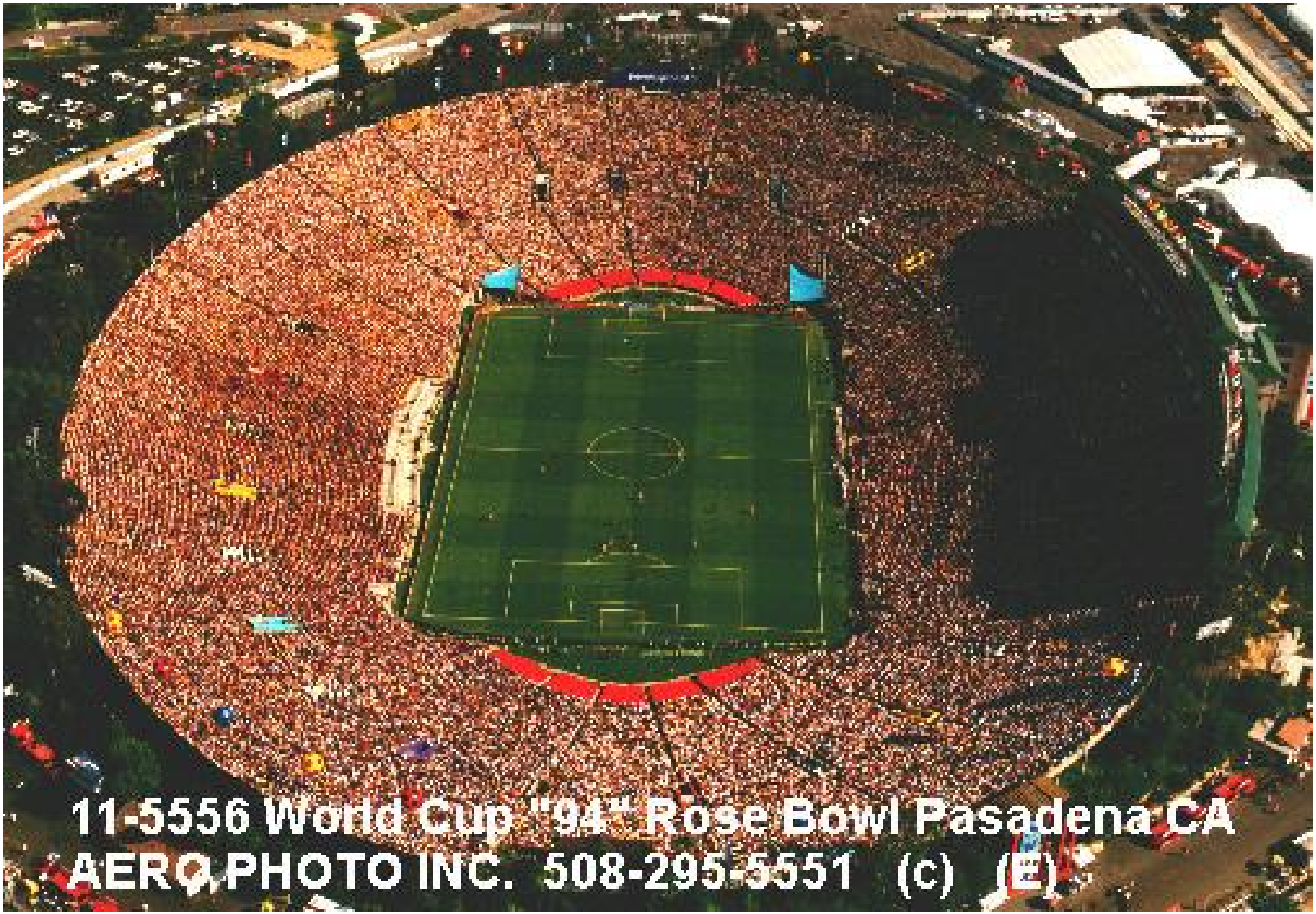


V1 responses to natural movies - summary

- Precise
- Sparse (power-law)
- Non-filter like
- Heterogeneous response
- Context sparsifies response

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




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

How close are we to understanding V1?

Five problems with the current view of V1:

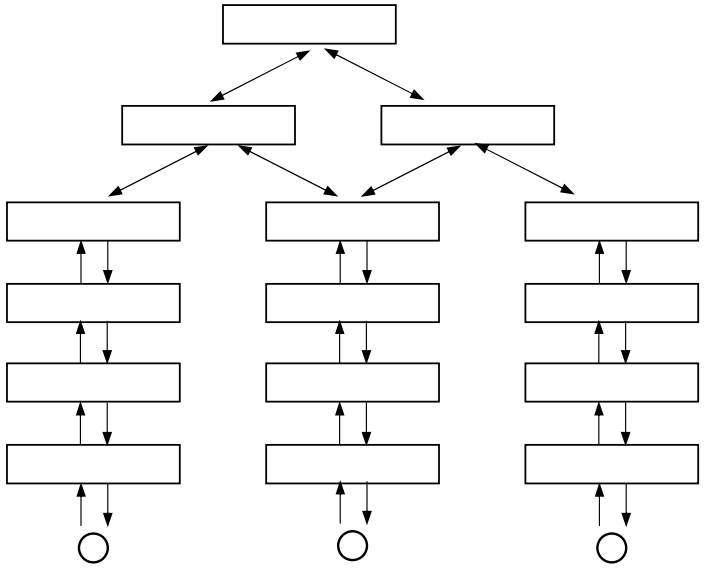
1. Biased sampling (single unit recording)
2. Biased stimuli (bars, spots, gratings)
3. Biased theories (data-driven vs. functional theories)
4. Interdependence and context (effect of intra-cortical inputs)
5. Ecological deviance

See:

Olshausen BA, Field DJ (2005) How close are we to understanding V1? *Neural Computation, 17(8)*, in press.

Hierarchical representation

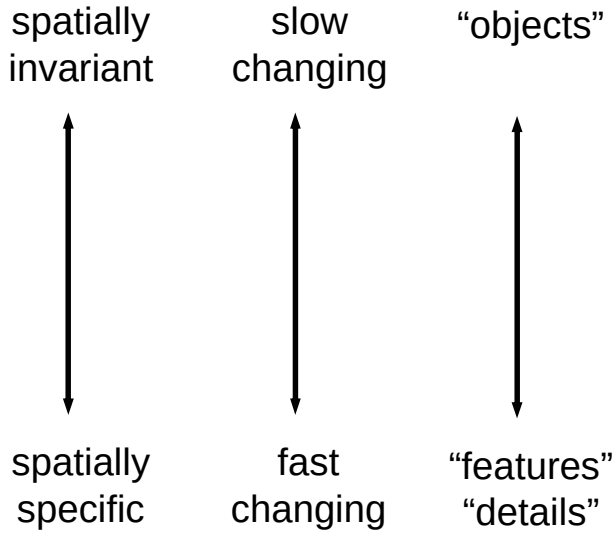
Hawkins & Blakeslee (2004) - "On Intelligence"



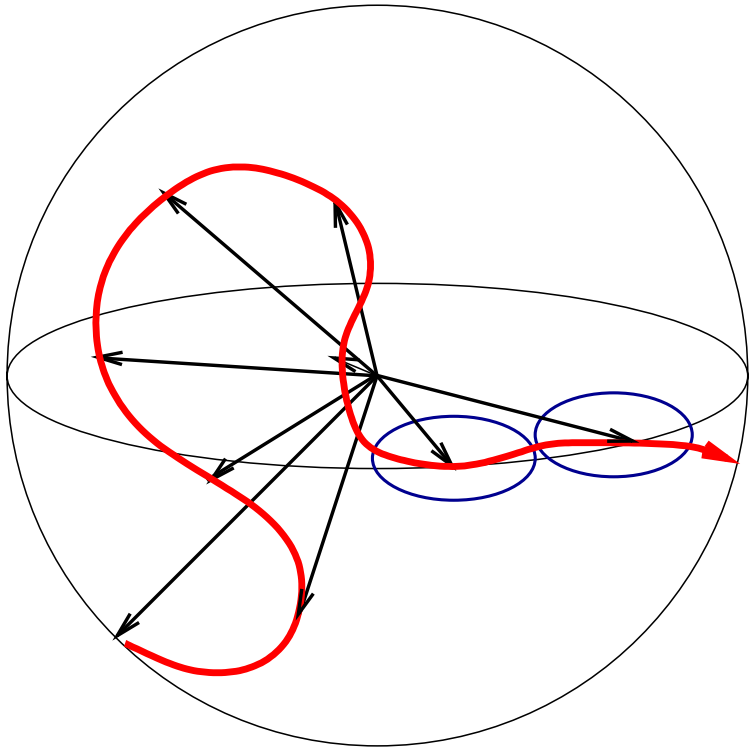
touch
motor

audition

vision



Invariance manifolds



Phase shifting in complex Gabor wavelets

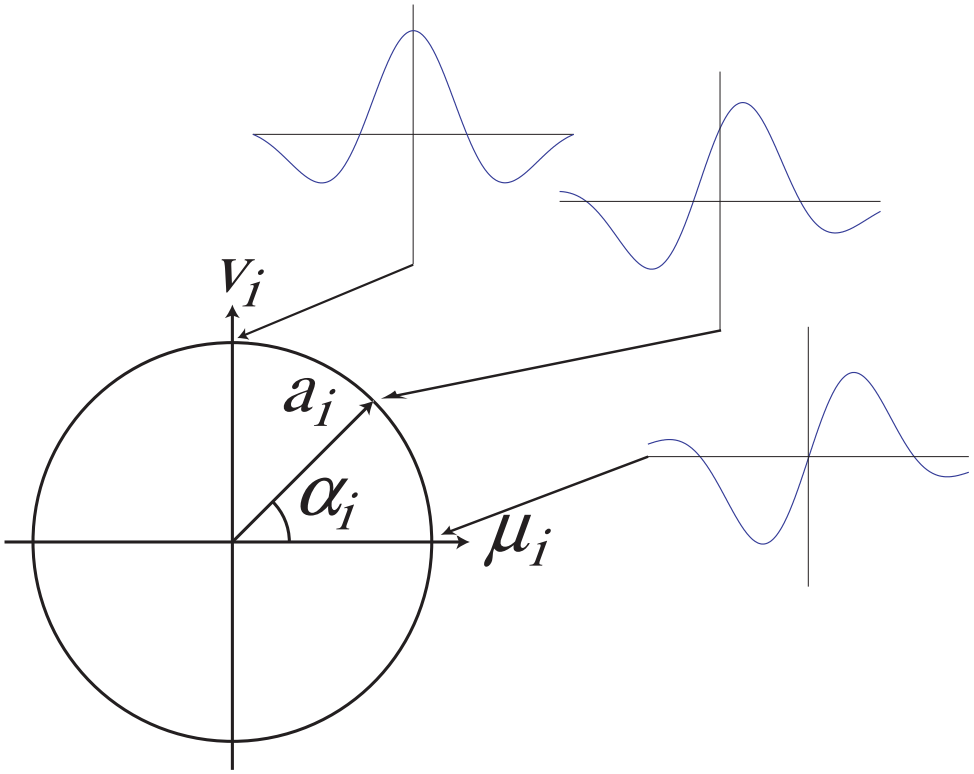
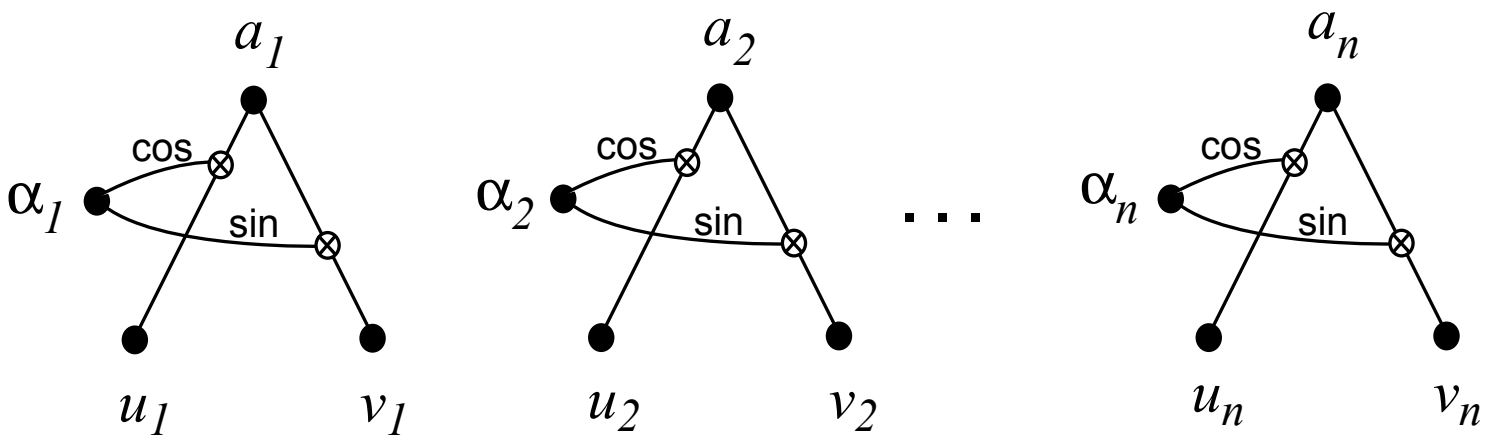


Image model with 'shiftable' basis functions

$$
\begin{aligned}
I(x,y) &= \sum_i \Re\{z_i\, \phi_i(x,y)\} \\
z_i &= a_i\, e^{j\,\alpha_i} \\
\phi_i(x,y) &= \phi_i^R(x,y) + j\,\phi_i^I(x,y)
\end{aligned}
$$

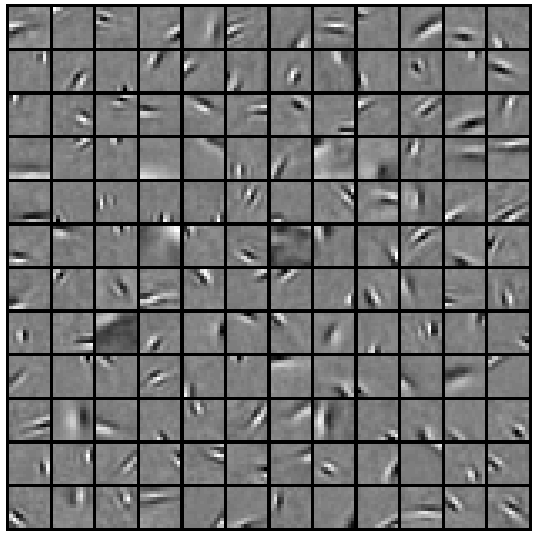
$$
I(x,y) = \sum_i a_i \left[\cos\,\alpha_i\, \phi_i^R(x,y) + \sin\,\alpha_i\, \phi_i^I(x,y)\right]
$$

Image model with 'shiftable' basis functions

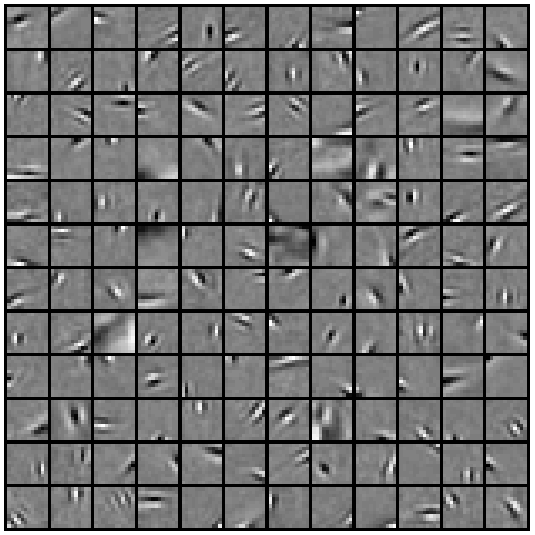


Learned complex basis functions (144, 12 $\times$ 12 patches)

real

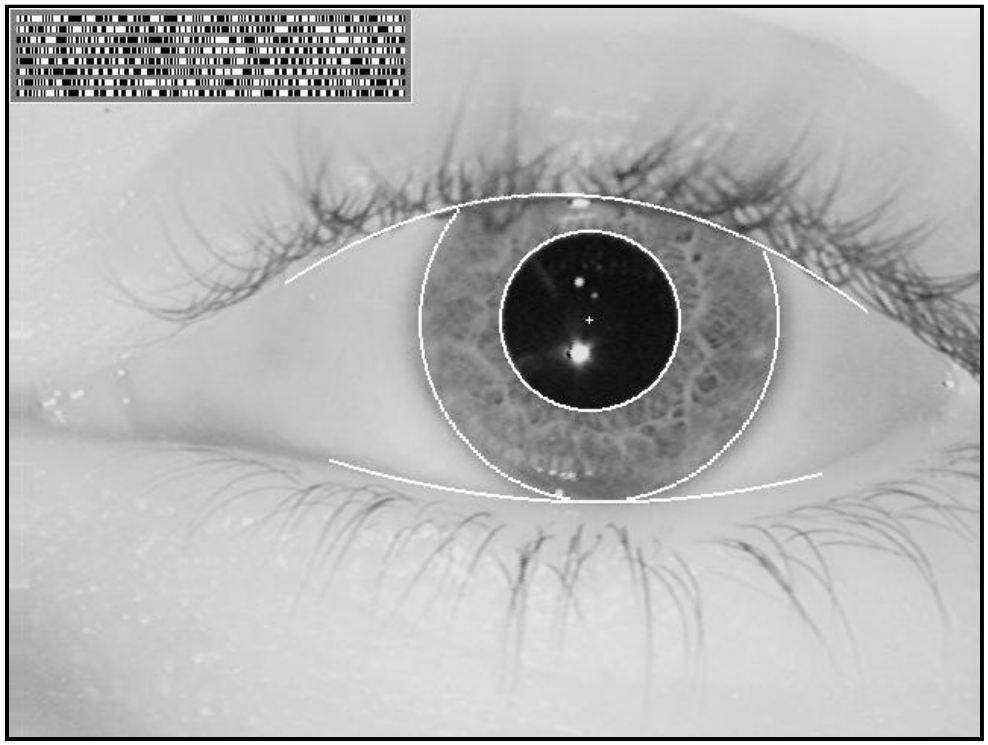


imag



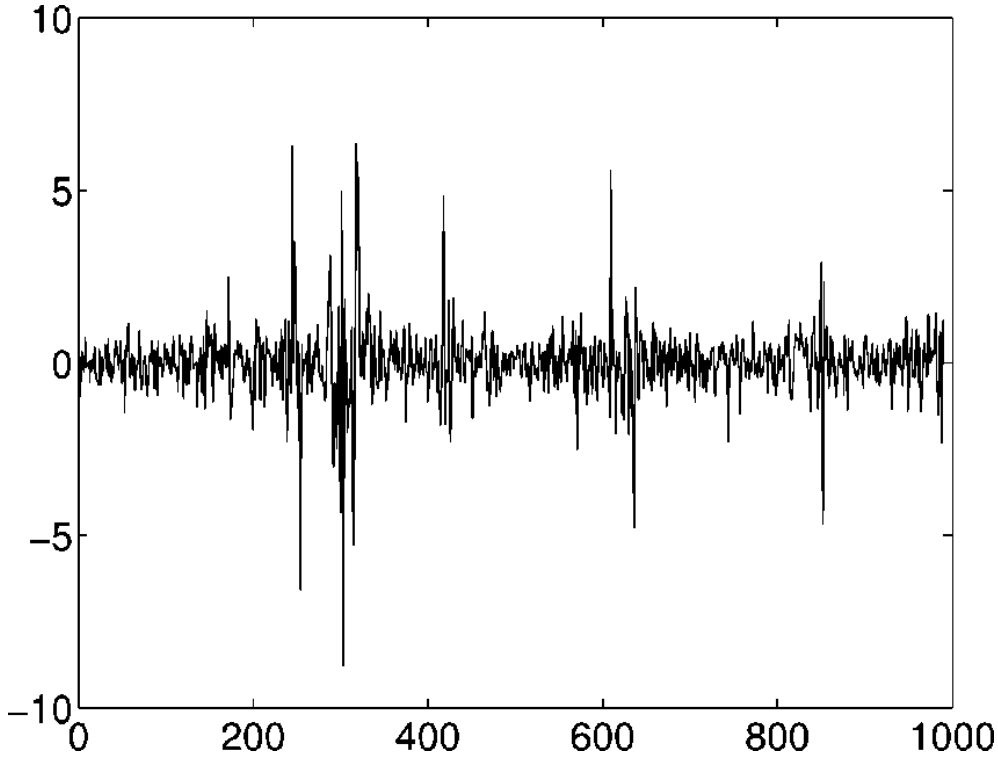
animate!

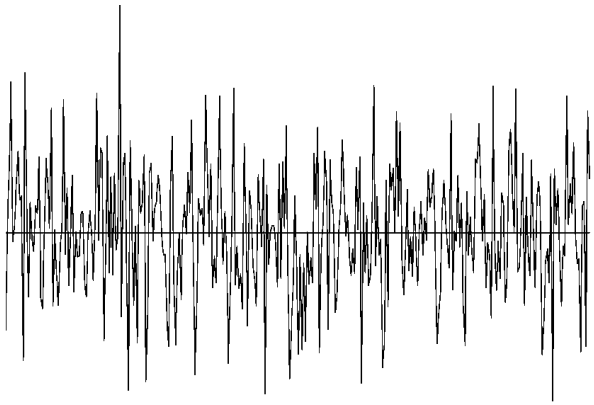
Iris recognition is based on local phase (Daugman)



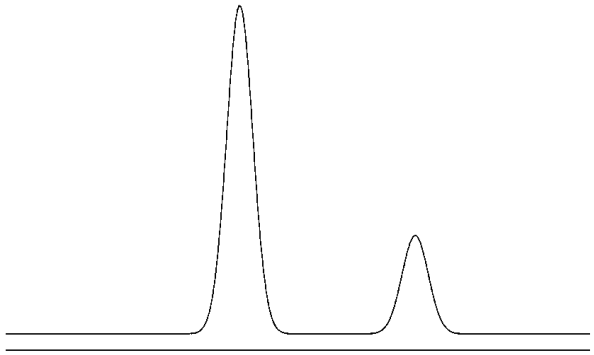
Sparse space-time bubbles

Hyvarinen et al. (2003) JOSA *20*

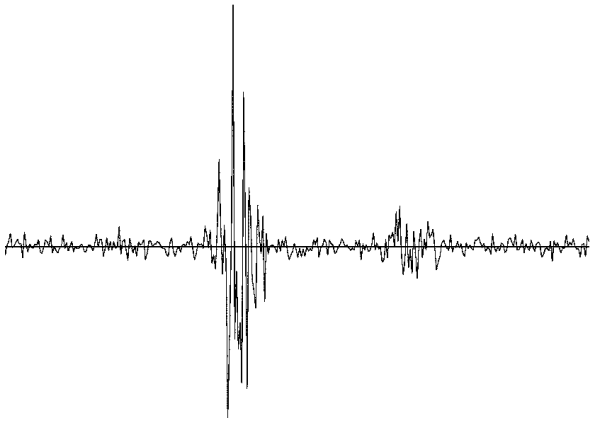




x



=

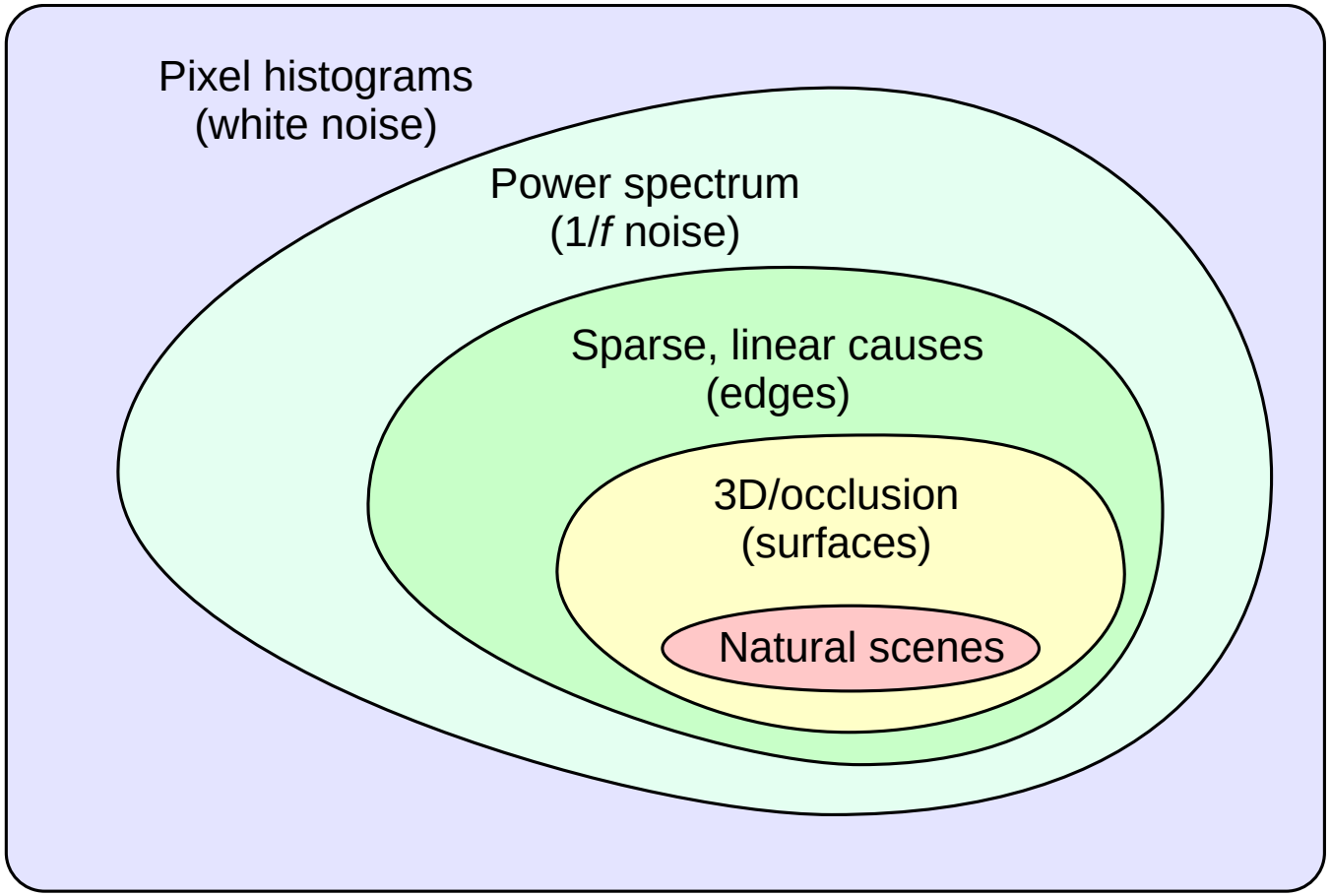


Learning invariance via slowness

Bilinear model of basis function coefficients:

$$a_i(t) = \underbrace{\sigma_i(t)}_{\text{slow}} \times \underbrace{z_i(t)}_{\text{fast}}$$

Image models



Conclusions

- The response properties of many neurons in the nervous system may be understood in terms of principles of efficient coding.

- The receptive fields of V1 neurons are well suited for producing sparse representations of natural images, and there is accumulating neurophysiological evidence that V1 employs such a strategy.

- One can also understand the local invariance properties of complex cells in terms of a bilinear model that factors apart amplitude (what) and phase (where) information.

- A full understanding of V1 function will require us to consider how it performs inference on natural scenes in the context of feedback from higher cortical areas.

Further information and details

baolshausen@ucdavis.edu
http://redwood.ucdavis.edu/bruno