Things we *think* we know, and things we *should* know, about visual cortex

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

REDWOOD CENTER
for Theoretical Neuroscience

Main points

- The efficient coding hypothesis
- Vision as inference
- Sparse coding in V1
- Towards hierarchical models

The efficient coding hypothesis
(Barlow 1961; Attneave 1954)

Nervous systems should exploit the statistical dependencies contained in sensory signals

Natural image statistics and efficient coding

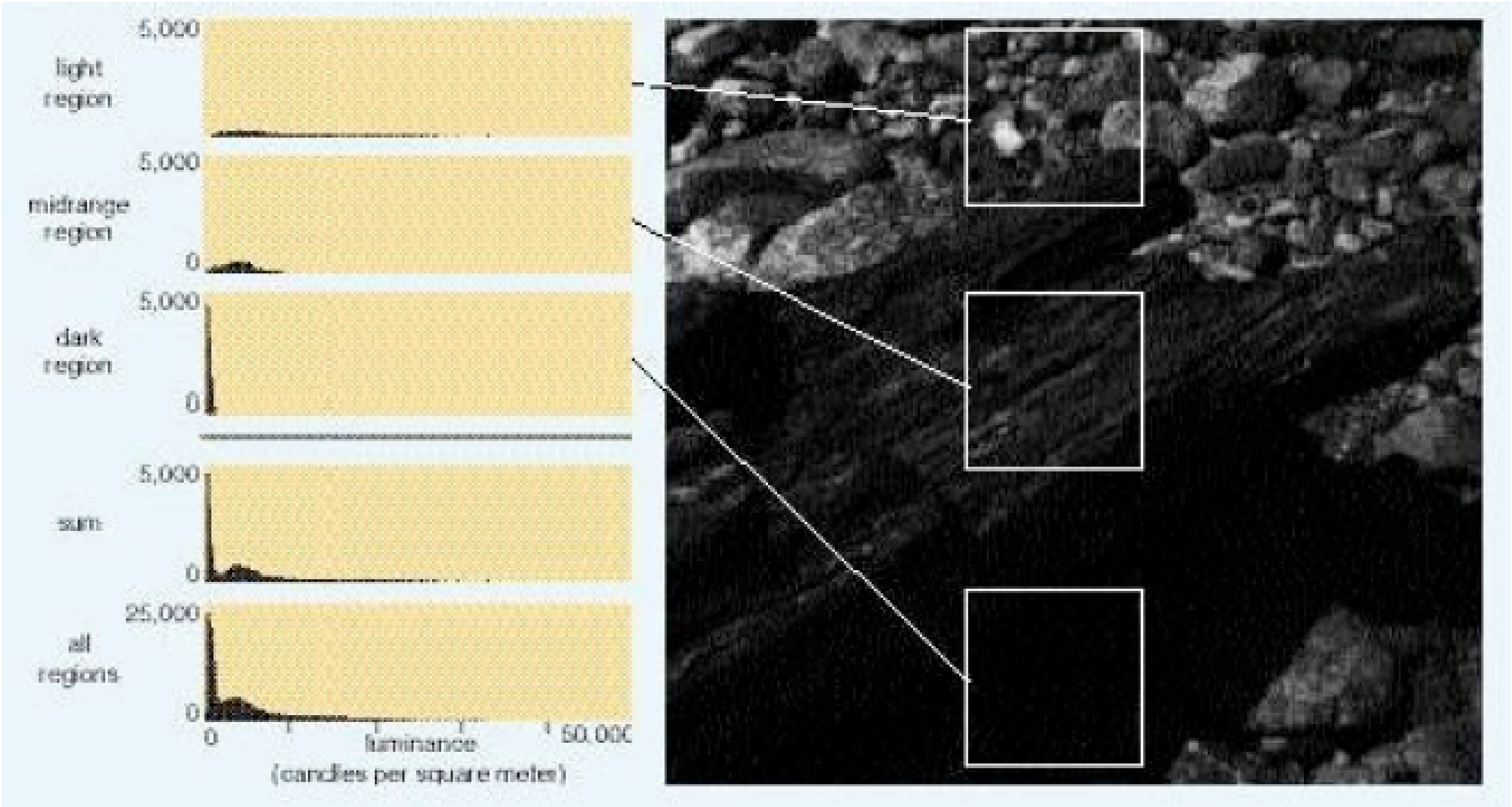
- First-order statistics
  - Intensity/contrast histograms
  - Histogram equalization

- Second-order statistics
  - Autocorrelation function $\rightarrow 1/f^2$ power spectrum
  - Decorrelation/whitening

- Higher-order statistics
  - Orientation, phase spectrum
  - Projection pursuit/sparse coding

First-order statistics (pixel histograms)

First-order statistics (pixel histograms)



Contrast: reduces dynamic range

$$C = \frac{I - \langle I \rangle}{\langle I \rangle}$$

Histogram equalization - fly LMC (Laughlin 1981)

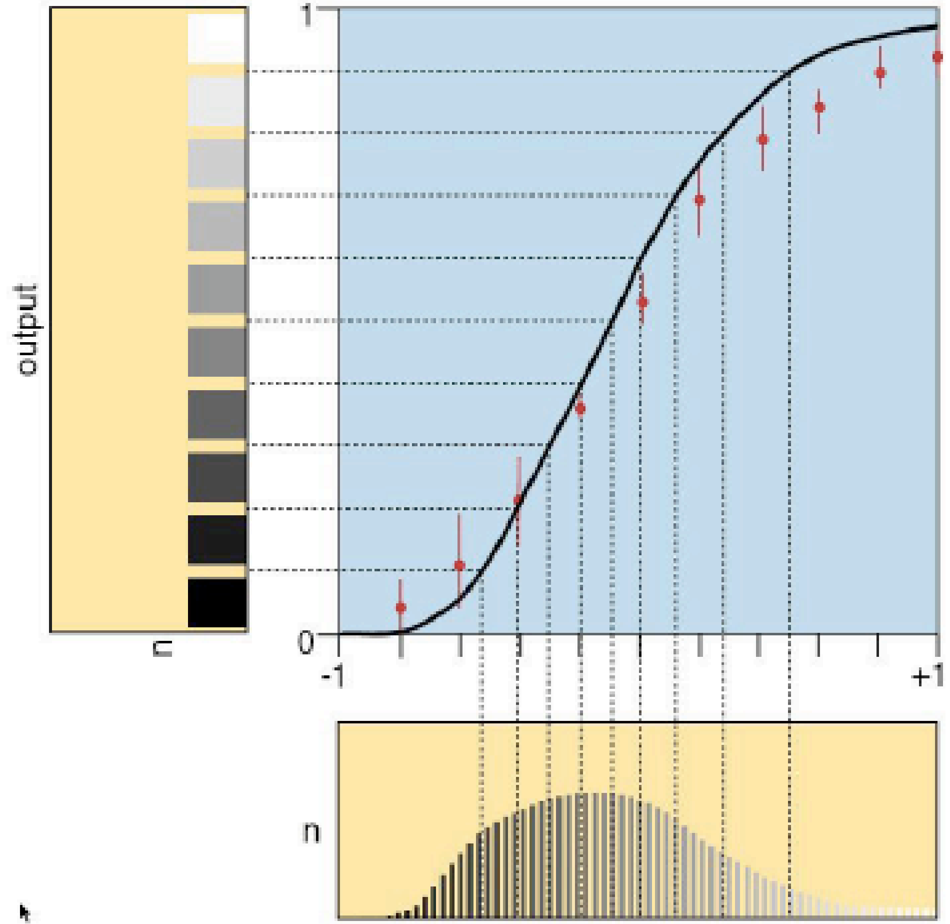
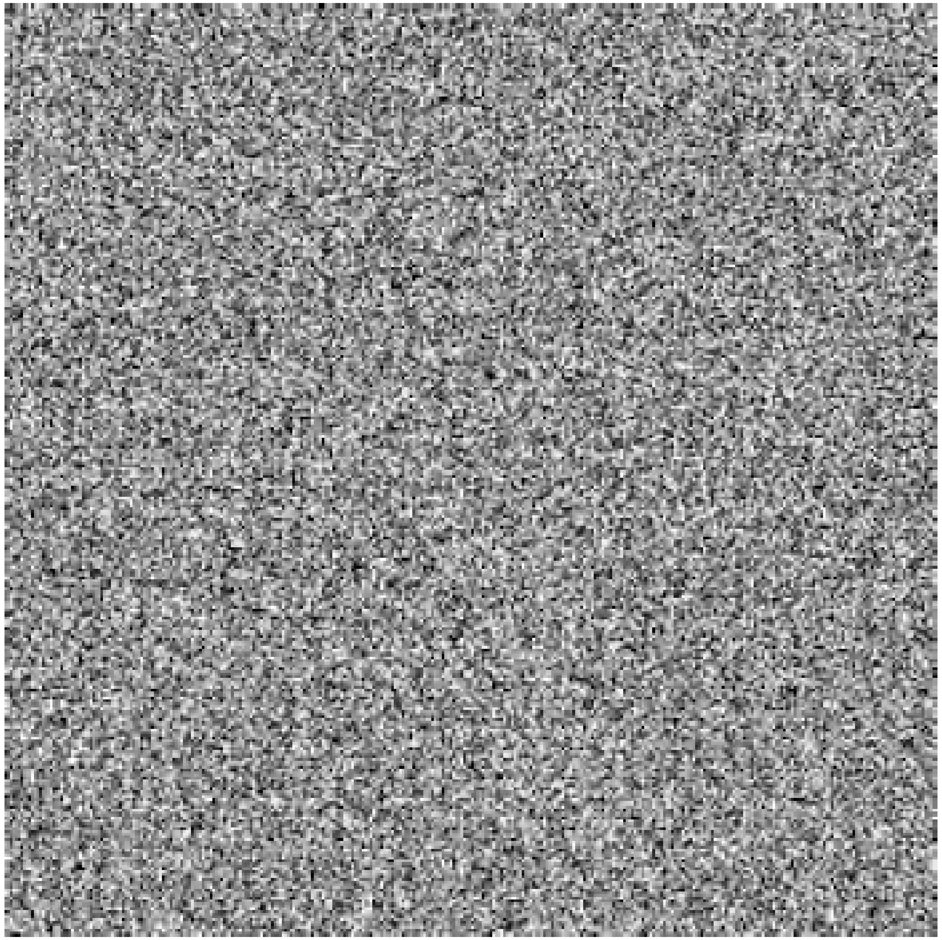
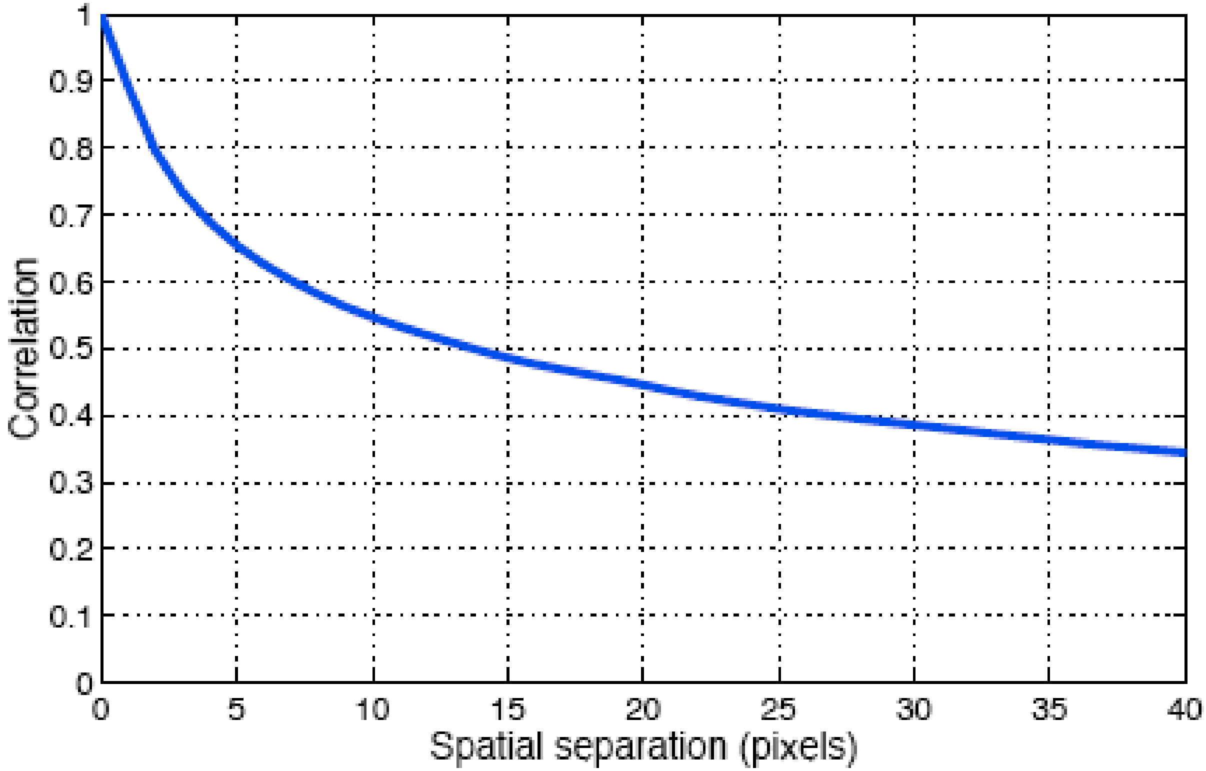


Image synthesis - first-order statistics

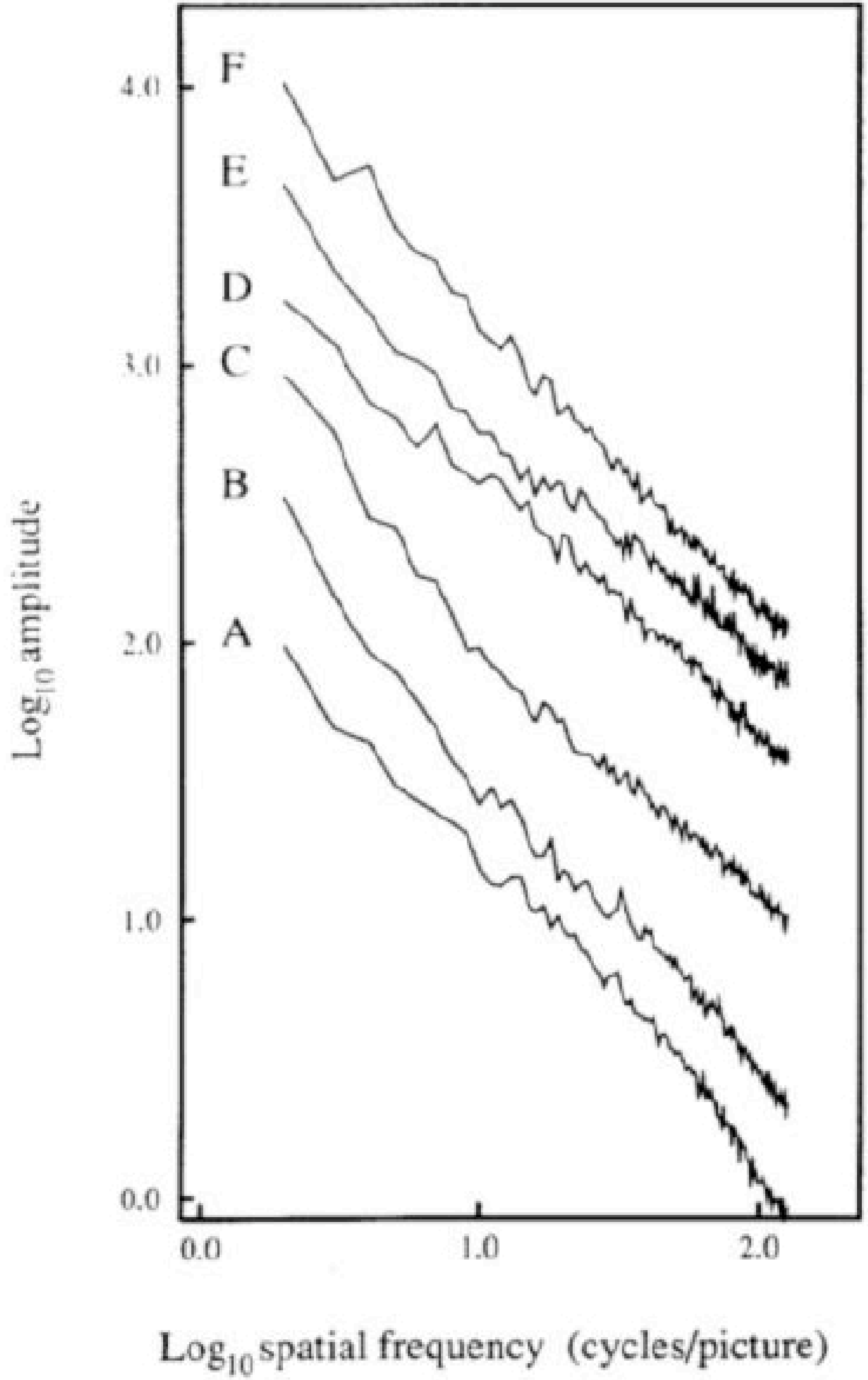


Second-order statistics (auto-correlation function)

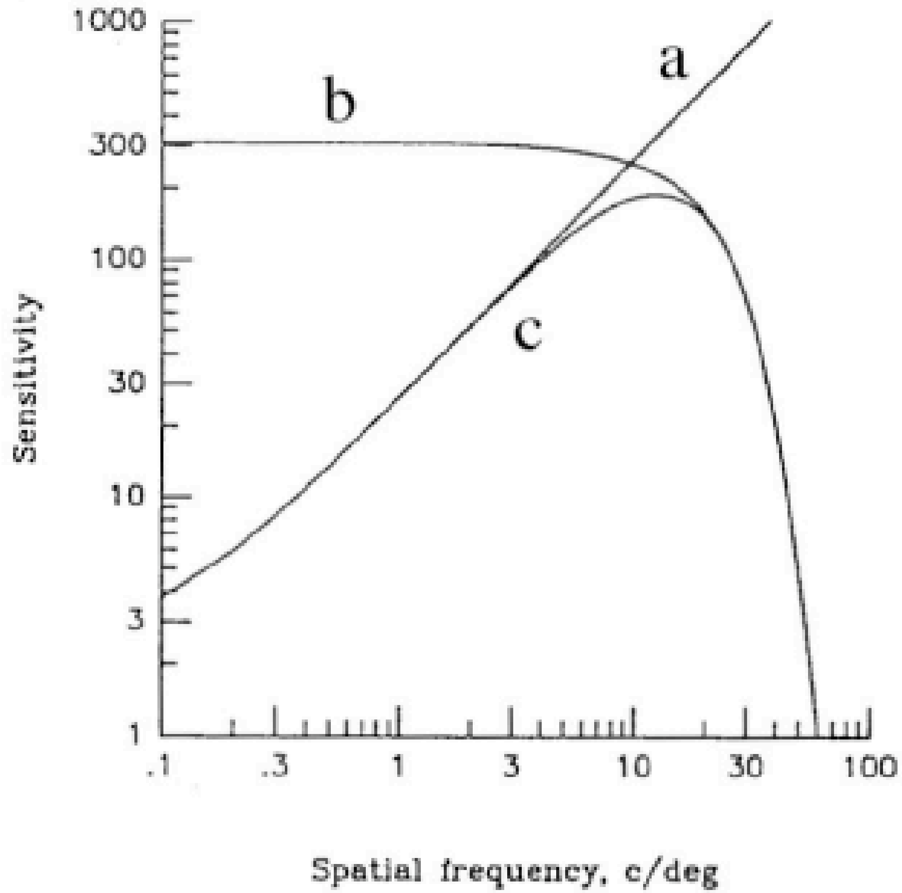
$$C(\Delta x) = \langle I(x)\, I(x + \Delta x) \rangle_x$$



Power spectrum of natural images (Field, 1987)



Whitening (Atick & Redlich, 1992)



Whitening removes second-order correlations

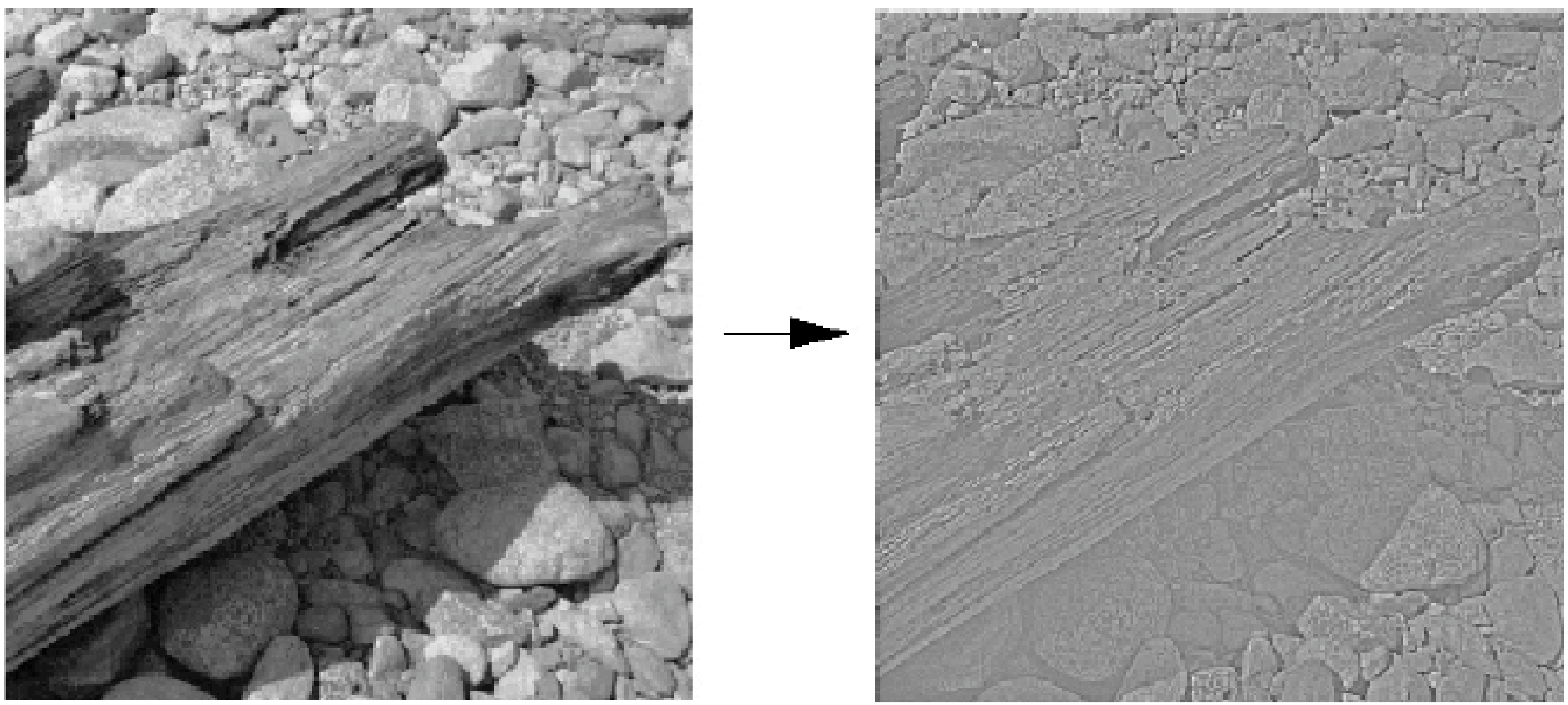
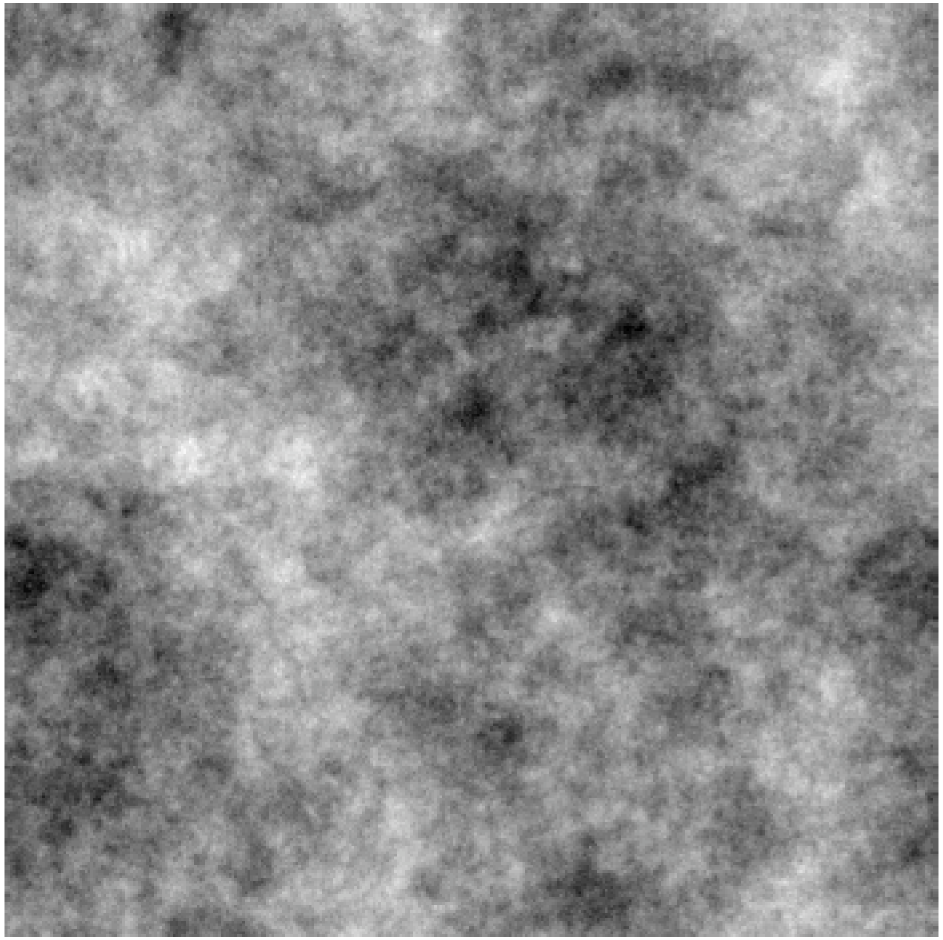
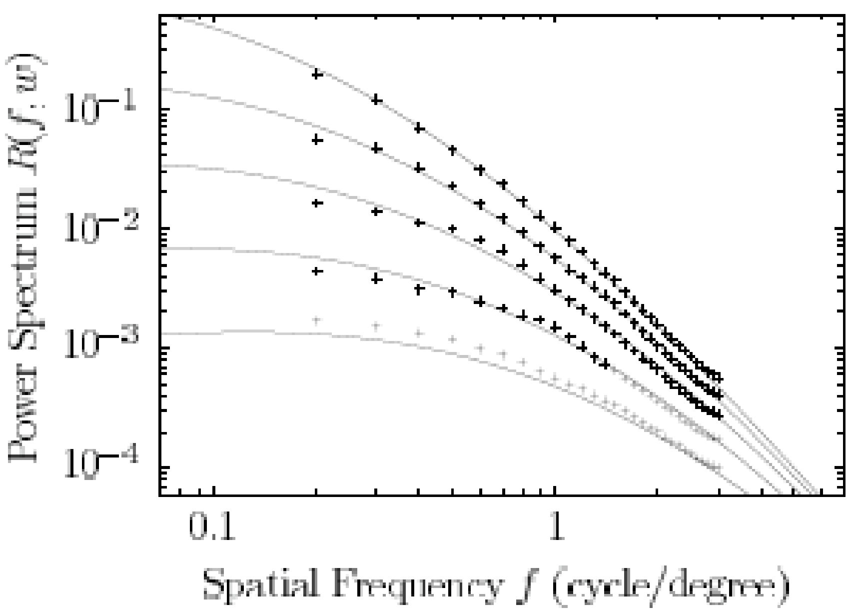


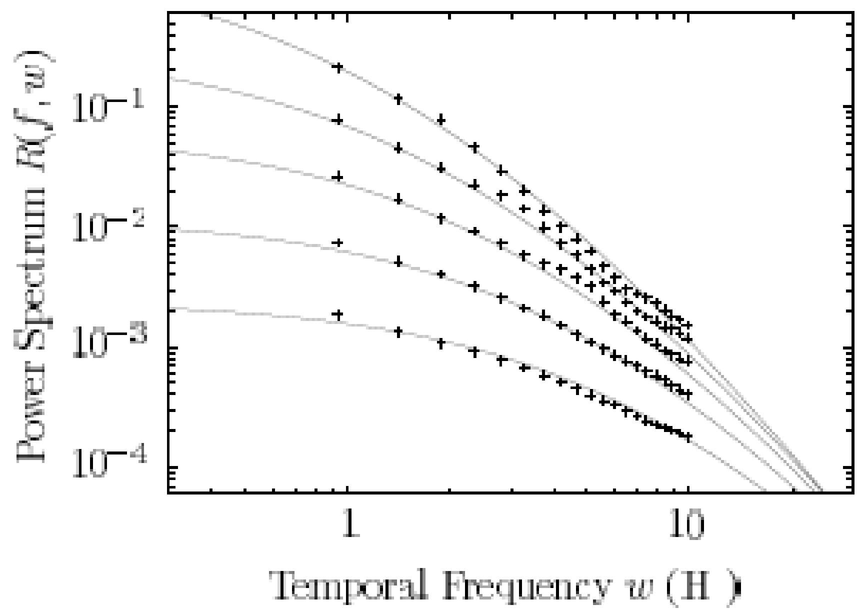
Image synthesis - second-order statistics



Spatiotemporal power spectrum of natural scenes

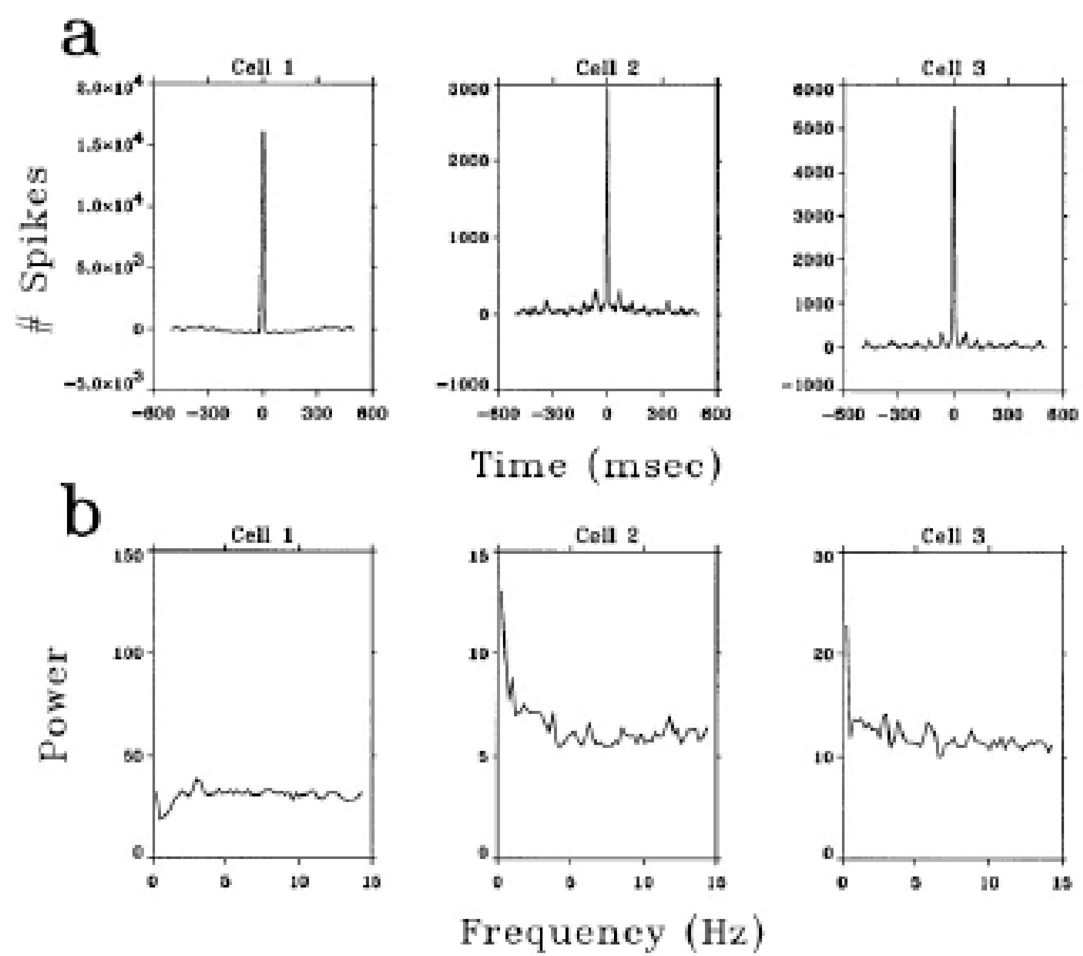
- Characterizes pairwise correlations across space and time.
- "$1/f^2$" but non-separable (Dong & Atick, 1995)



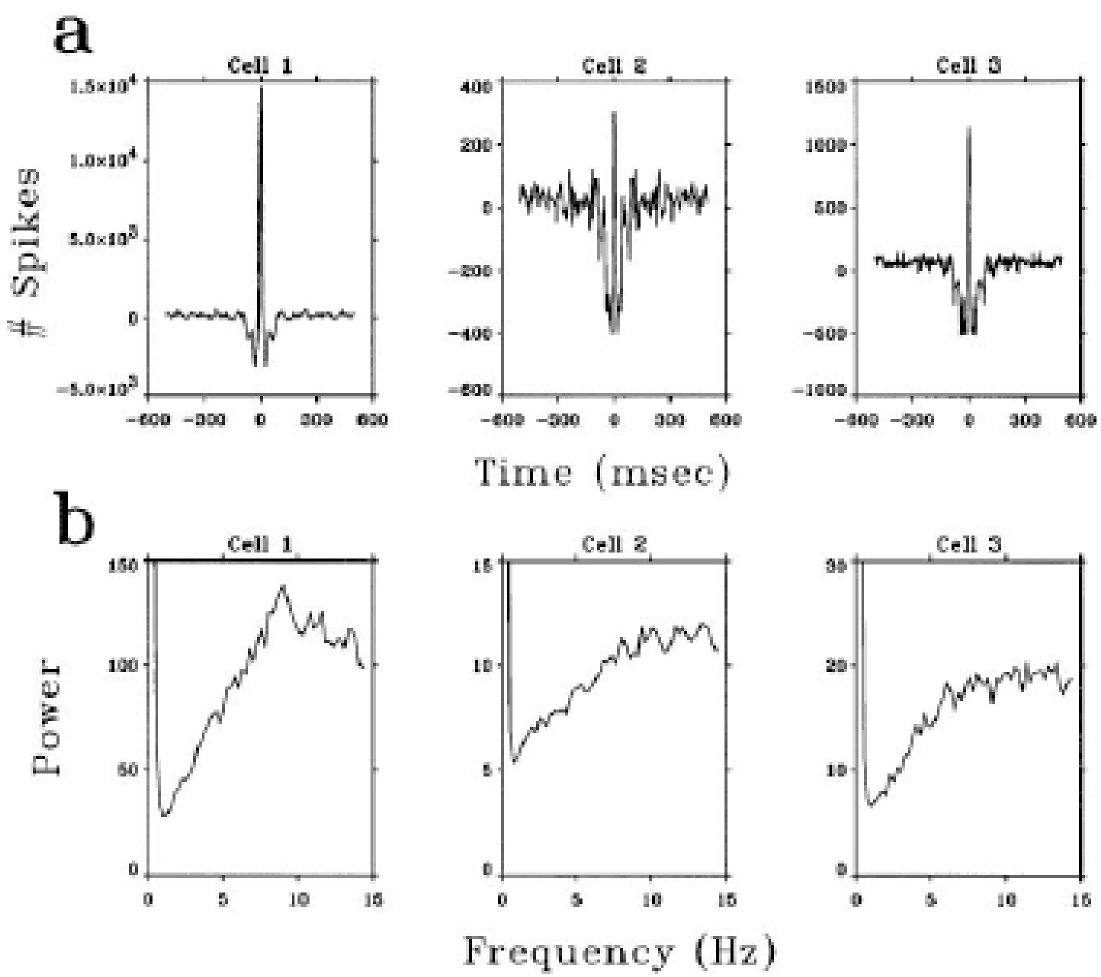


LGN neurons whiten time-varying natural images

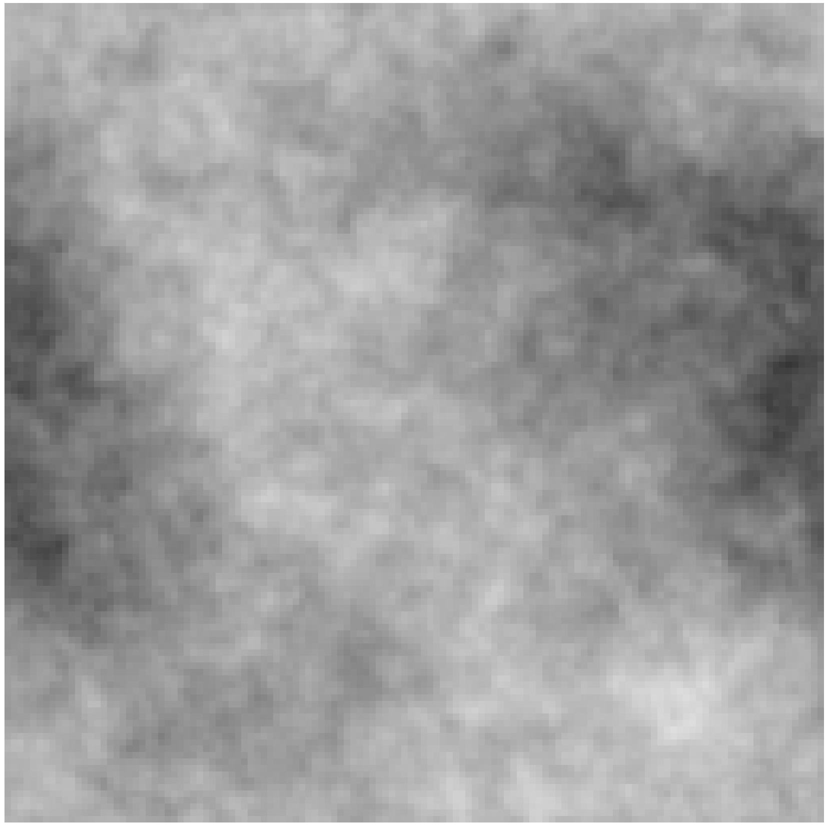
Dan et al, 1996



... but not white noise

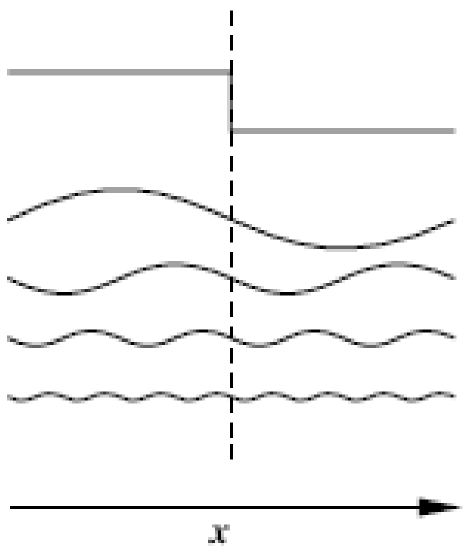


Movie synthesis - second-order, s-t statistics

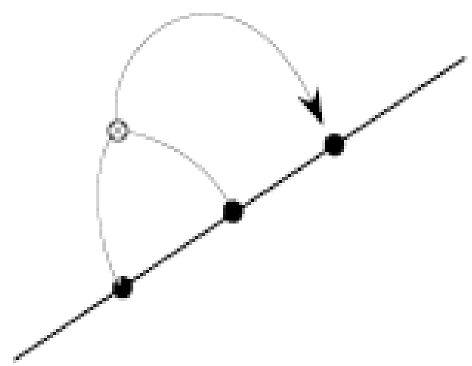


Higher-order statistics

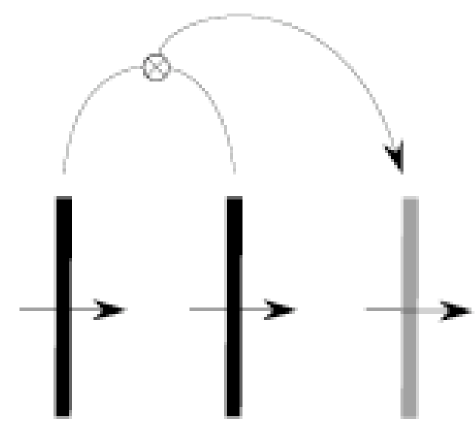
Phase alignment



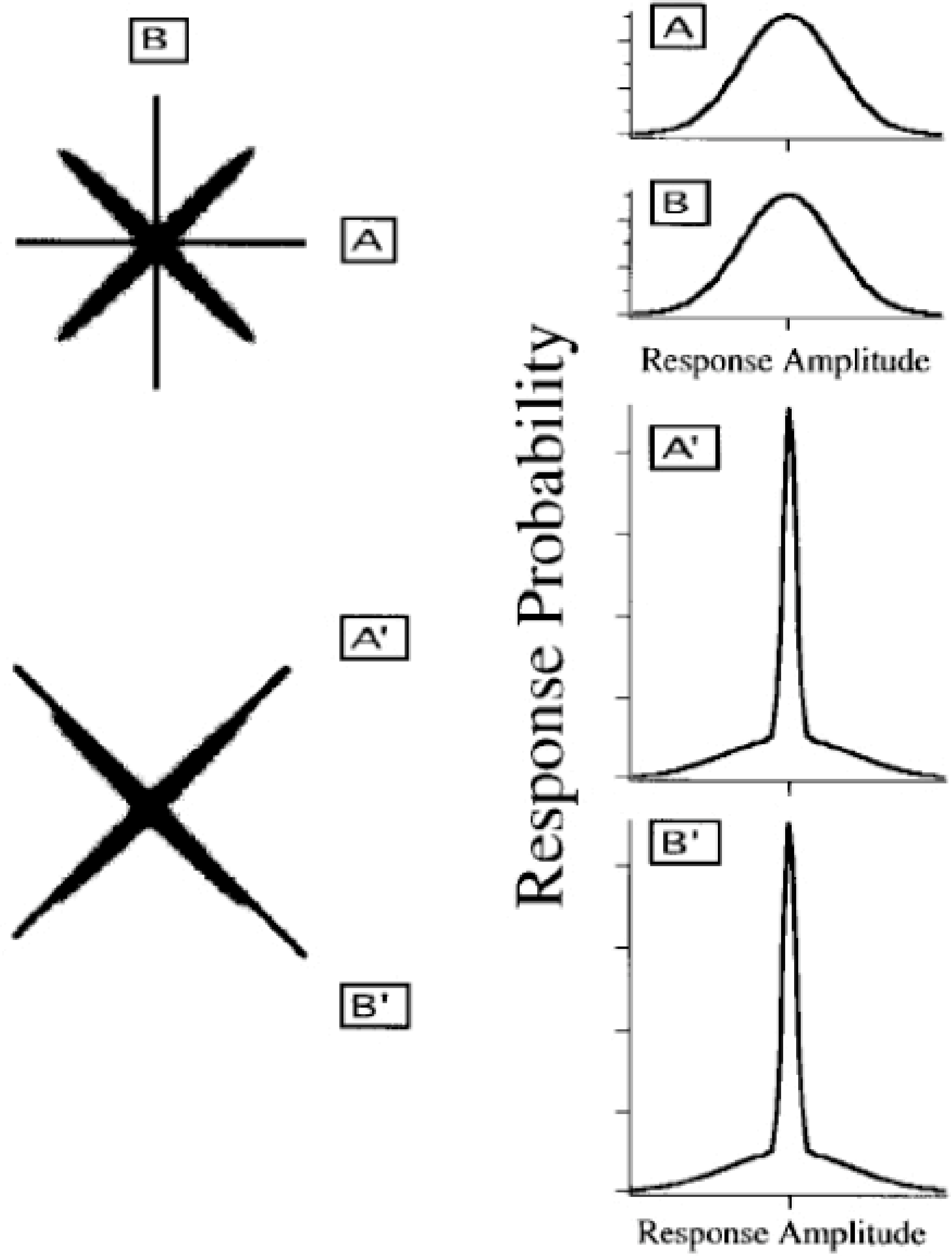
Oriented structure



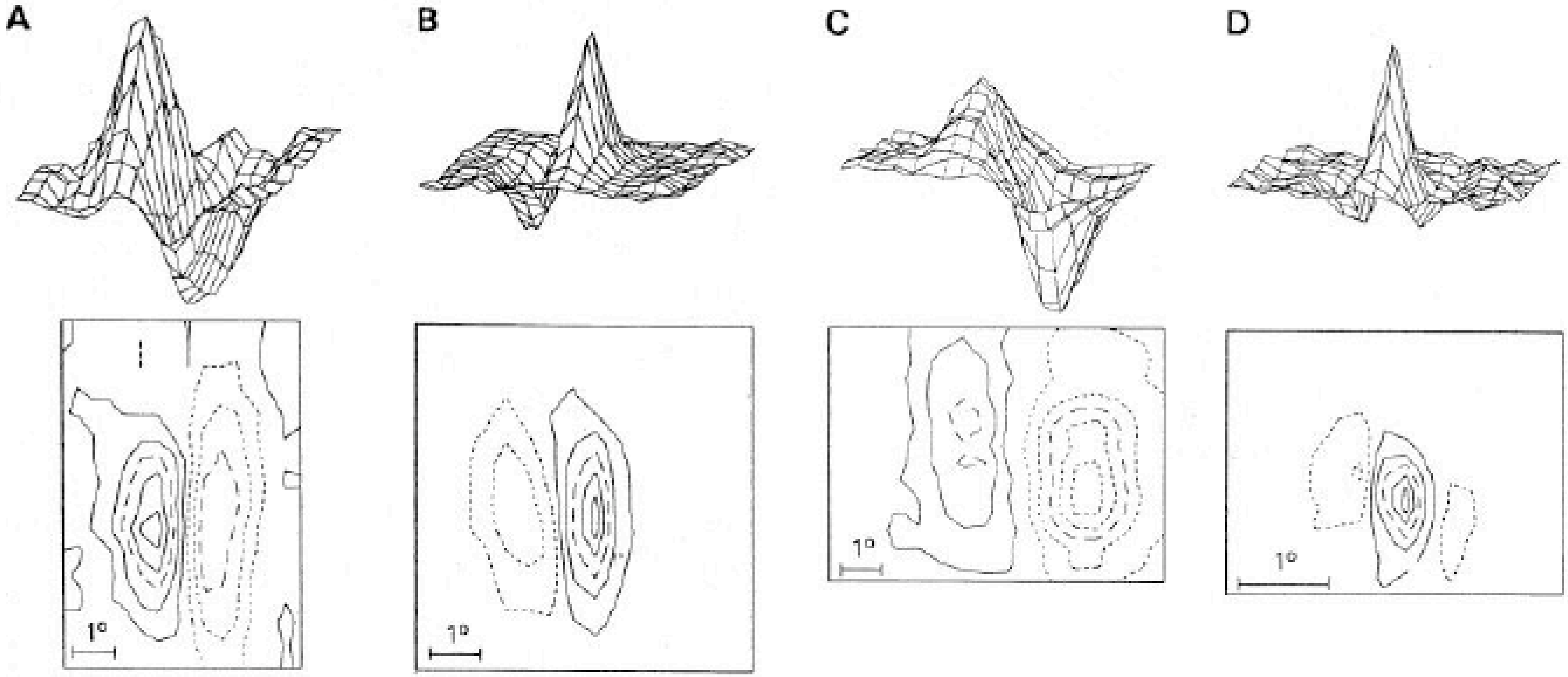
Motion



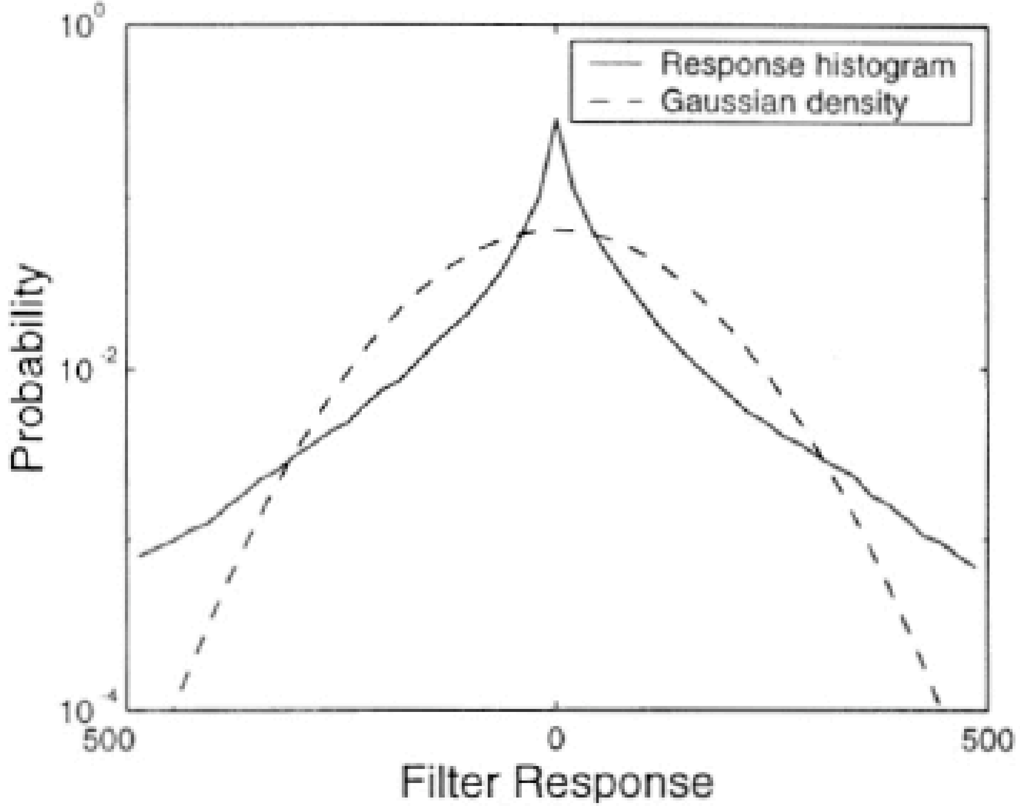
Projection Pursuit
(from Field 1994)



Simple cell receptive fields (Jones & Palmer, 1987)

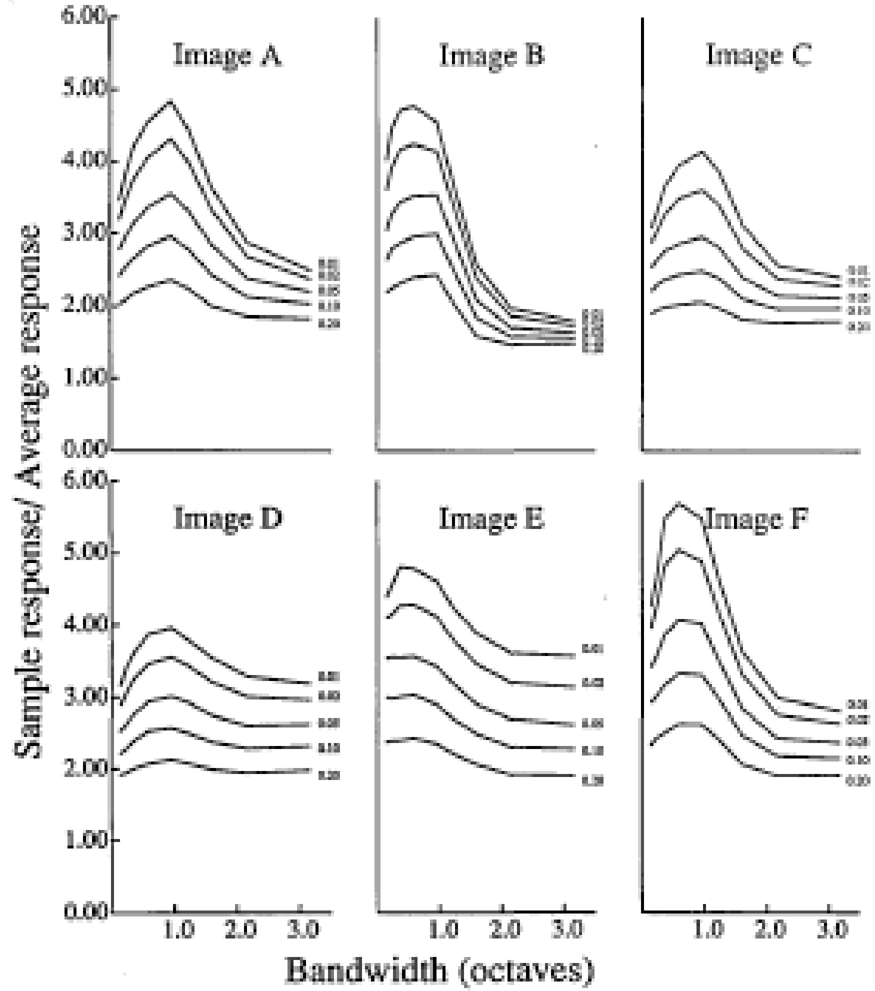


Gabor-filter histogram



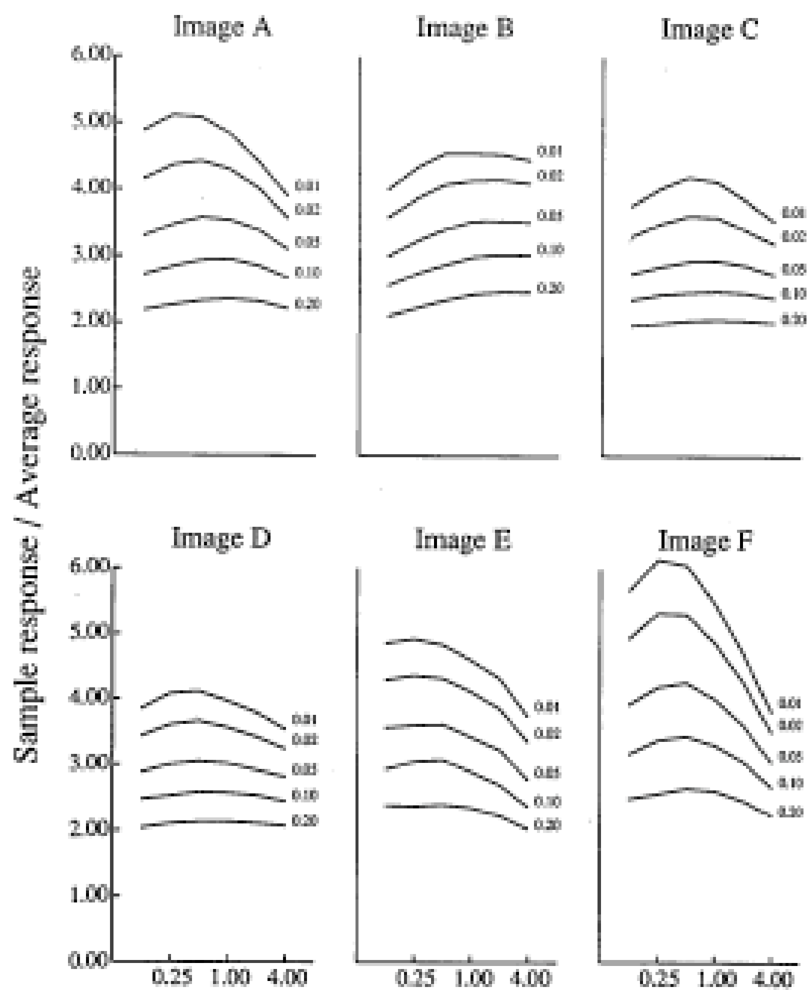
Optimal spatial-frequency bandwidth

Field (1987)



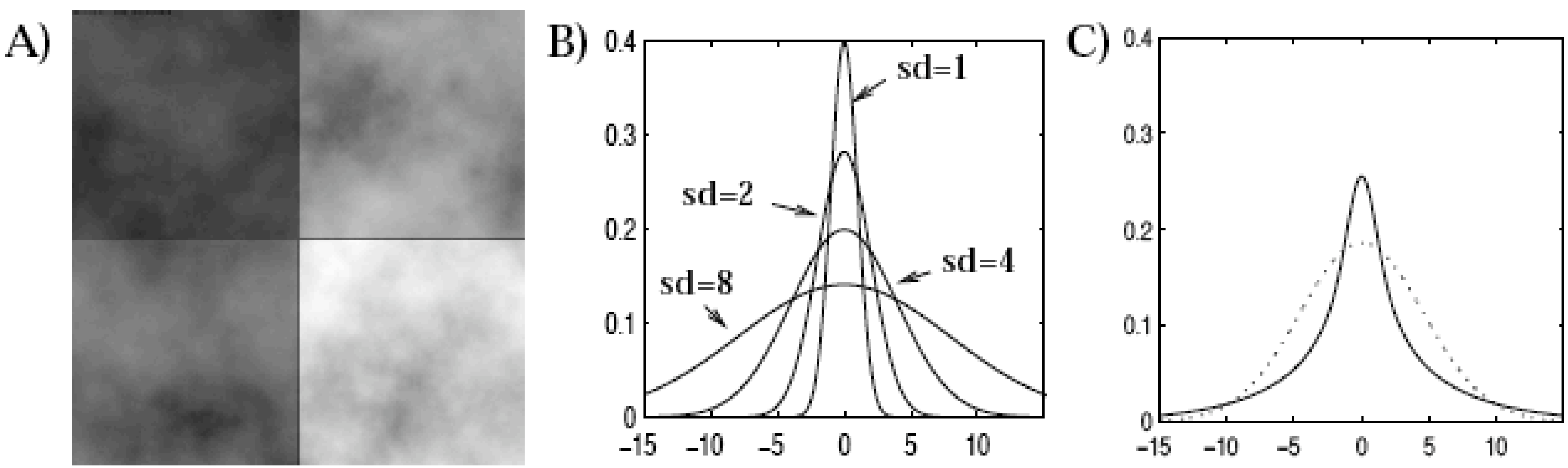
Optimal orientation bandwidth

Field (1987)

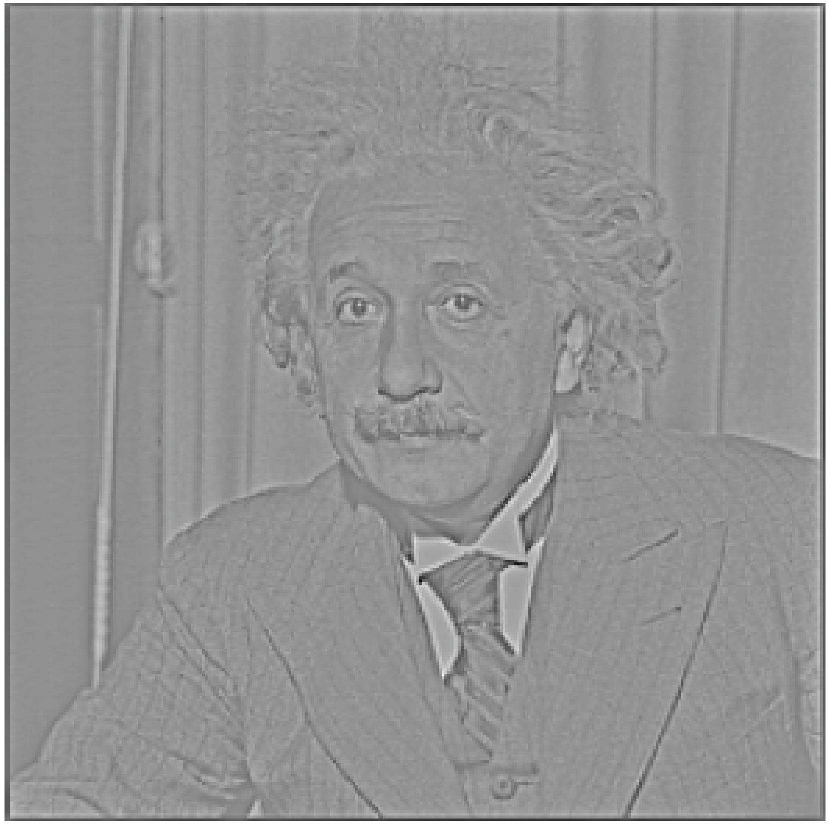


Searching for filters with ‘interesting’ output distributions: an uninteresting direction to explore?

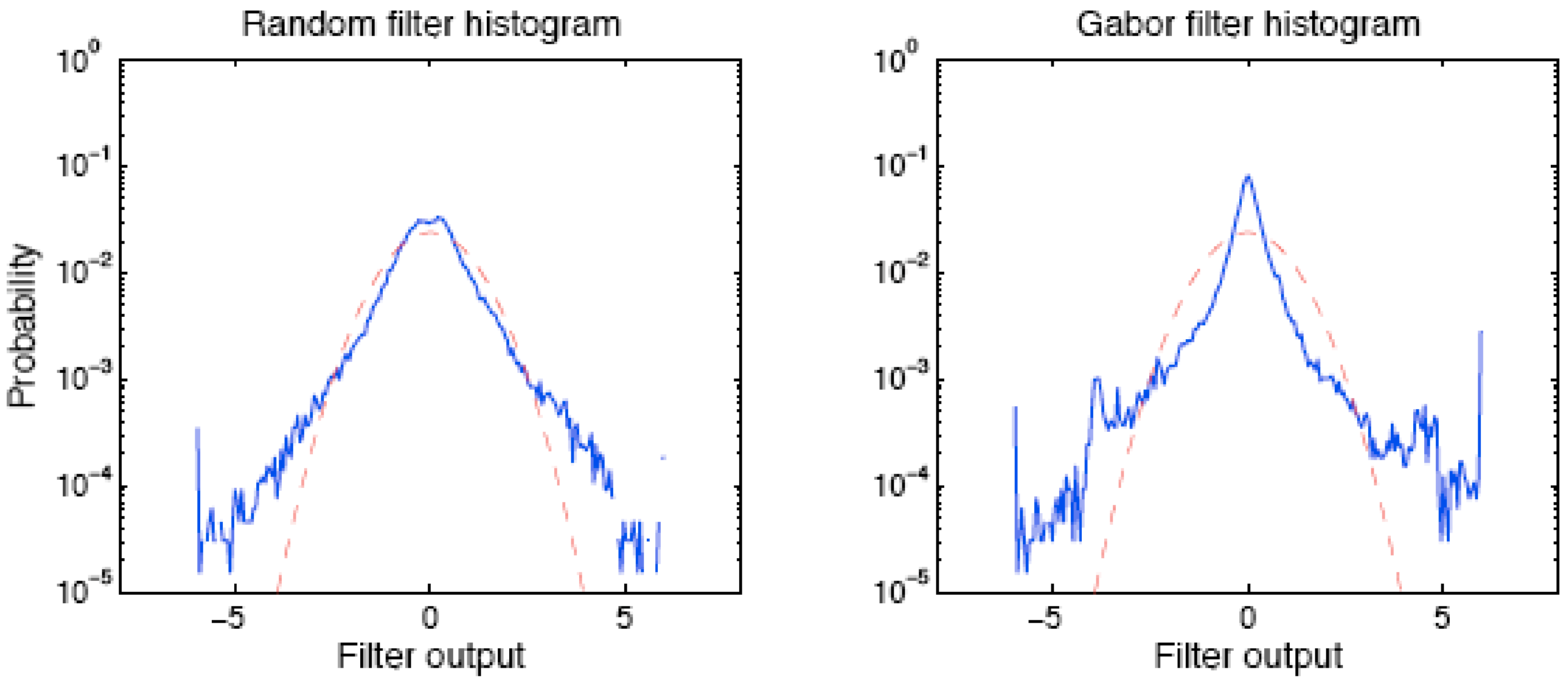
Baddeley (1996) *Network 7*, 409-421



Whitened image

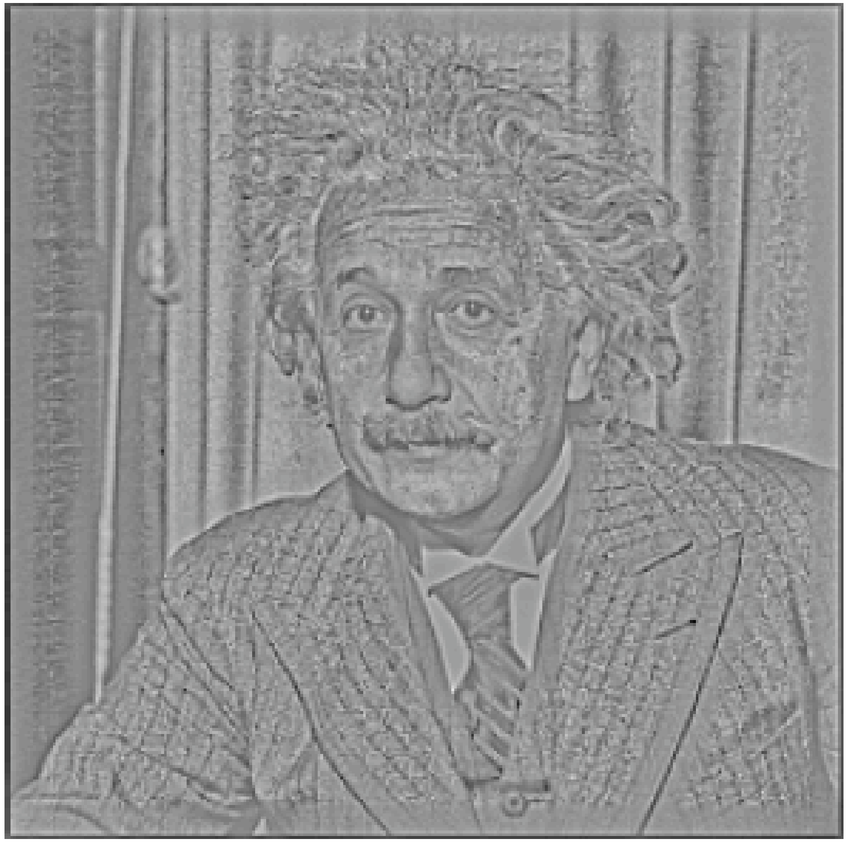


Filter histograms

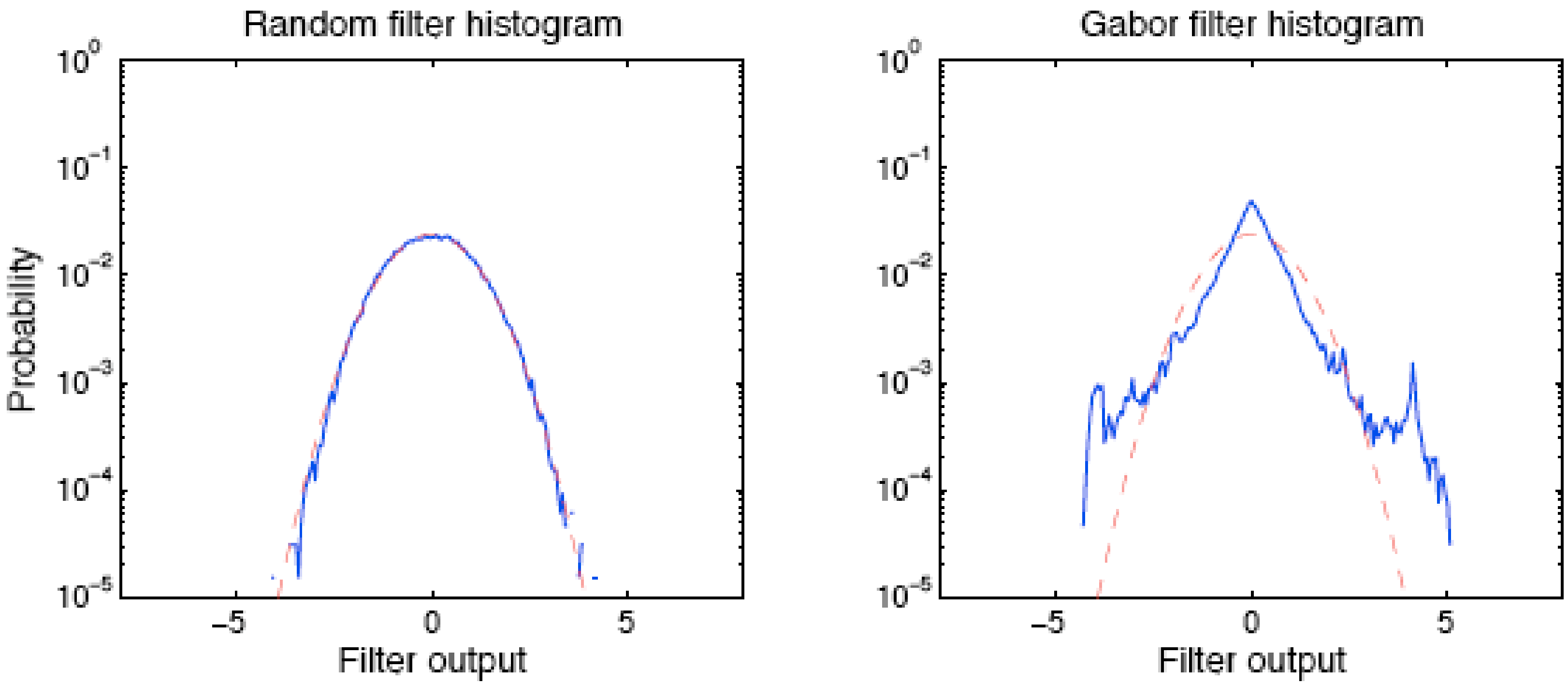


Contrast fluctuations

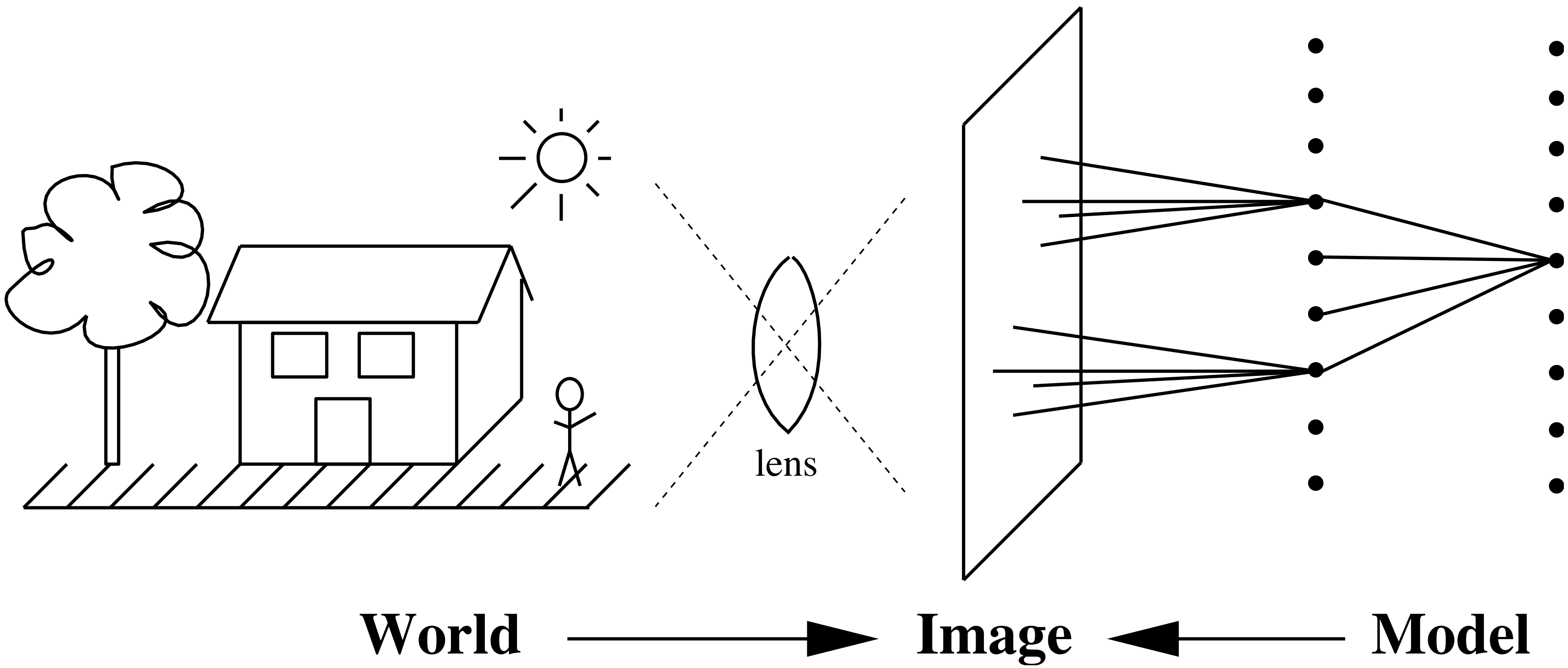
Whitened, contrast normalized image



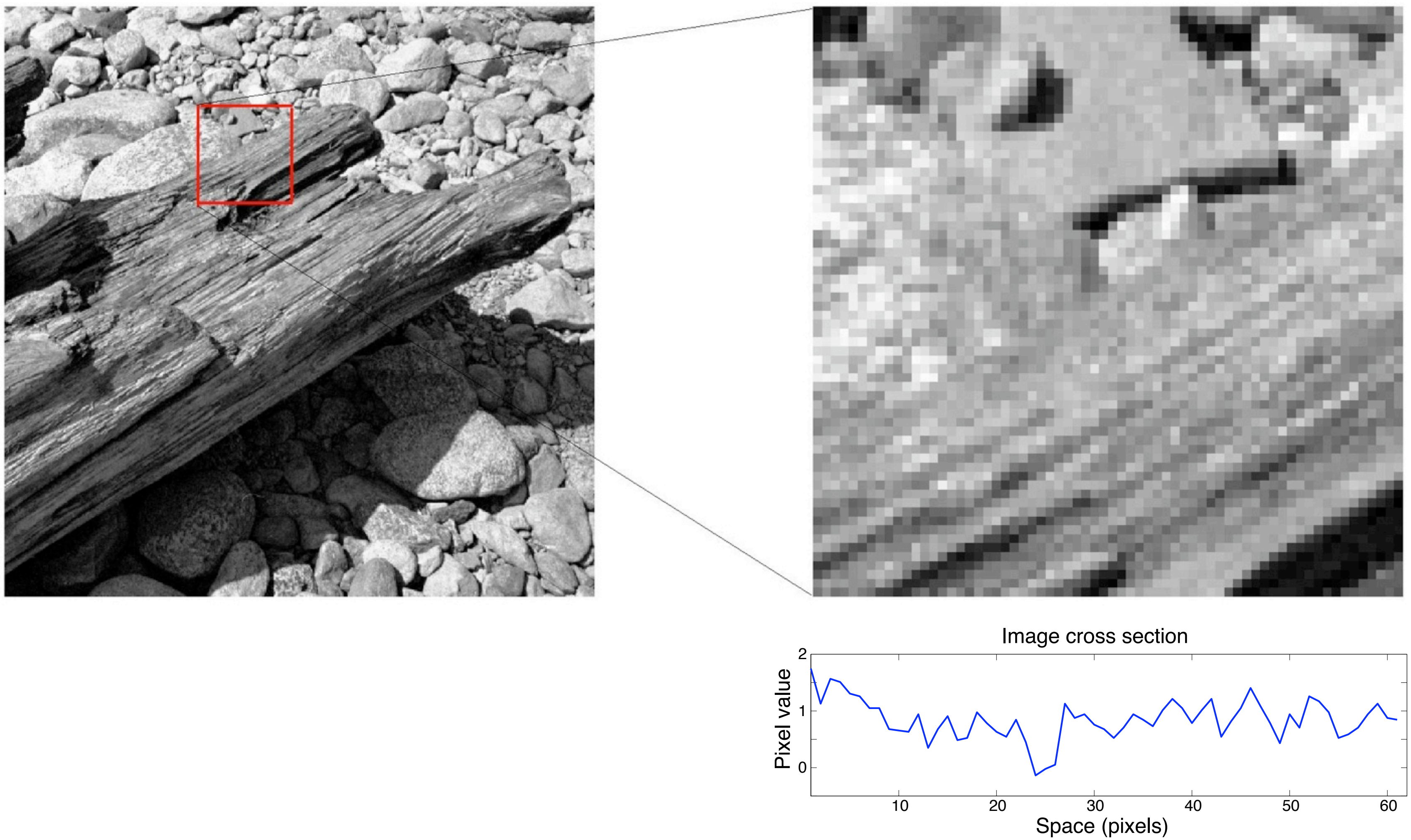
Filter histograms



Vision as inference



Natural scenes are filled with ambiguity

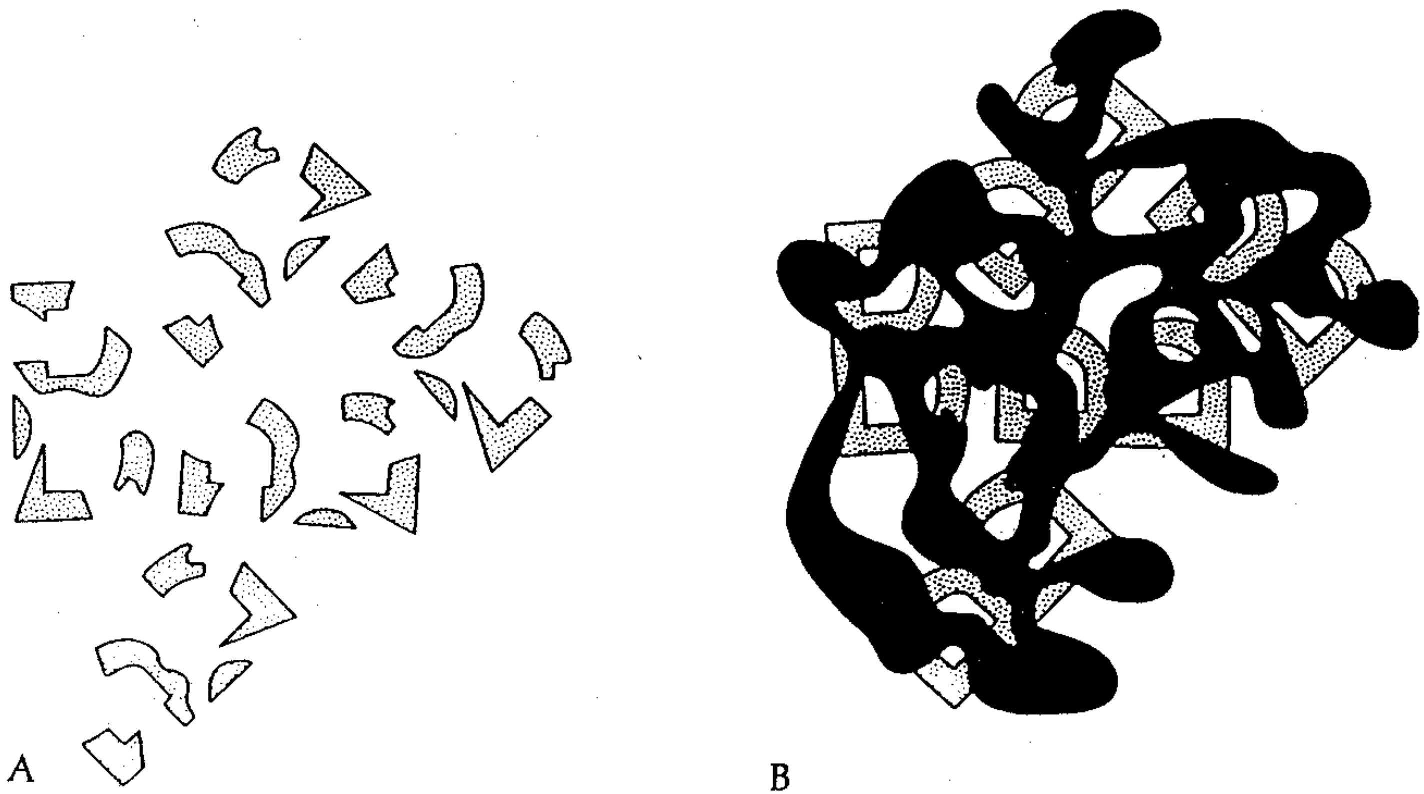


Mooney faces

Mooney faces

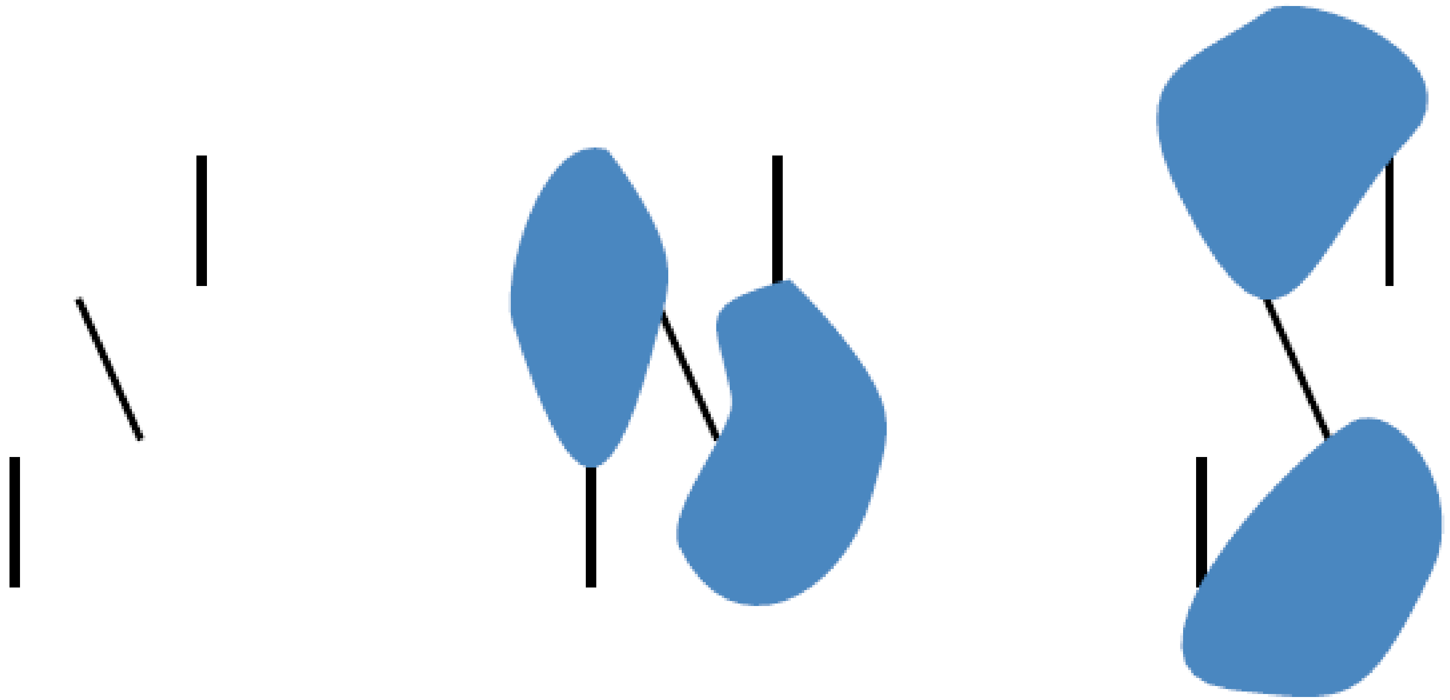
Mooney faces

Bregman B's



A B

Occluders determine object completion

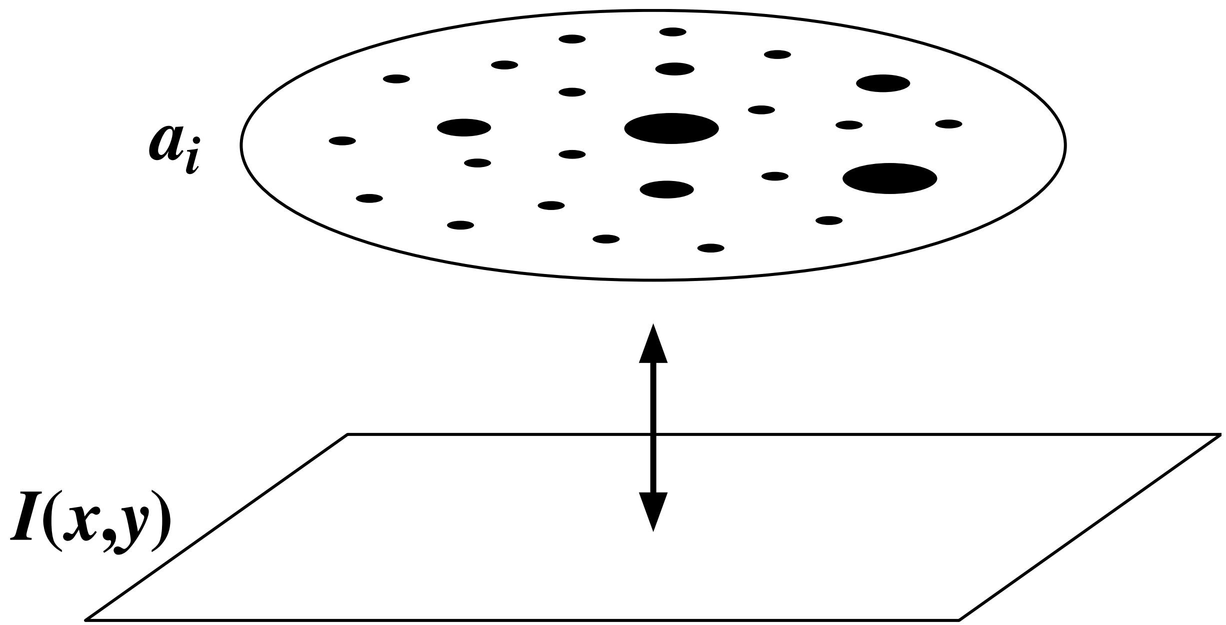


Object recognition depends on scene context

Object recognition depends on scene context

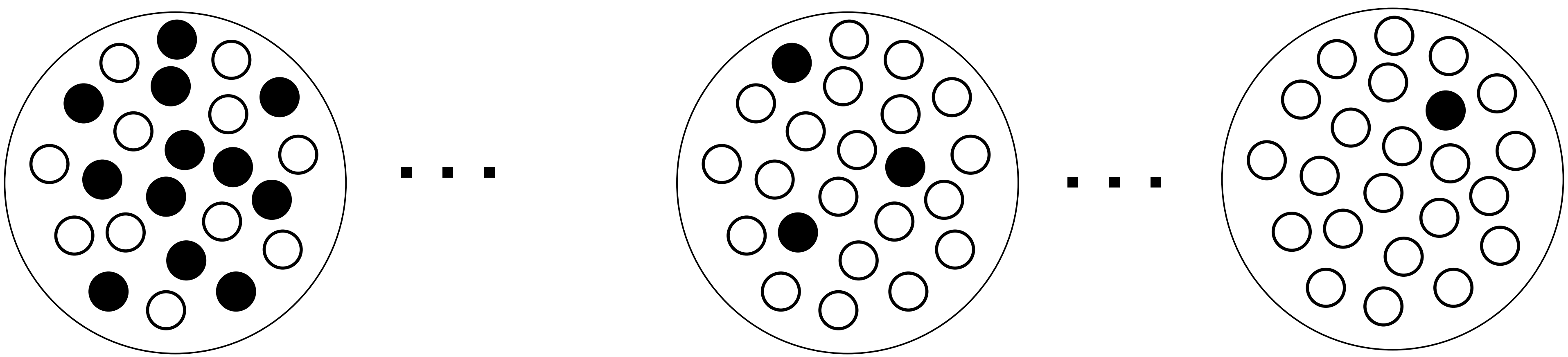
Object recognition depends on scene context

Sparse, distributed representations



Sparse vs. dense coding

Dense codes (ascii) | Sparse, distributed codes | Local codes (grandmother cells)



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

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

- 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.

Image model

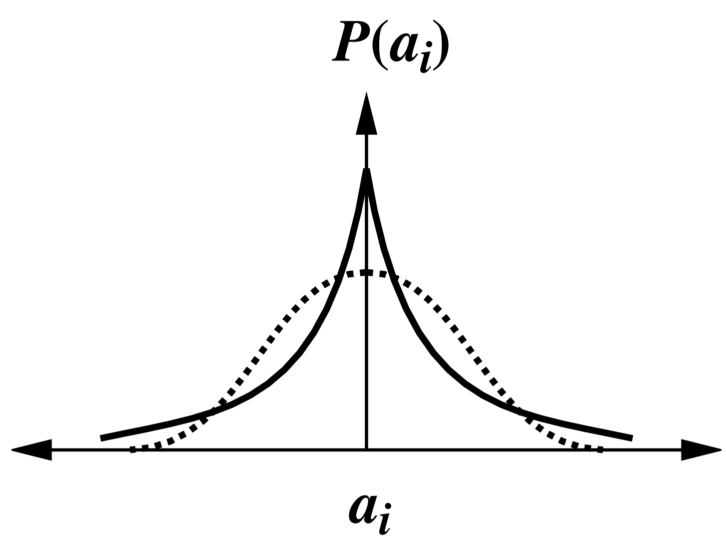
$$I(x,y) = \sum_i a_i \, \phi_i(x,y) + \epsilon(x,y)$$

Goal: Find a dictionary $\{\phi\}$ which enables a sparse representation of the image in terms of the coefficients a_i

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:

$$E(\mathbf{I},\mathbf{a}) = -\log P(\mathbf{a}|\mathbf{I},\theta)$$

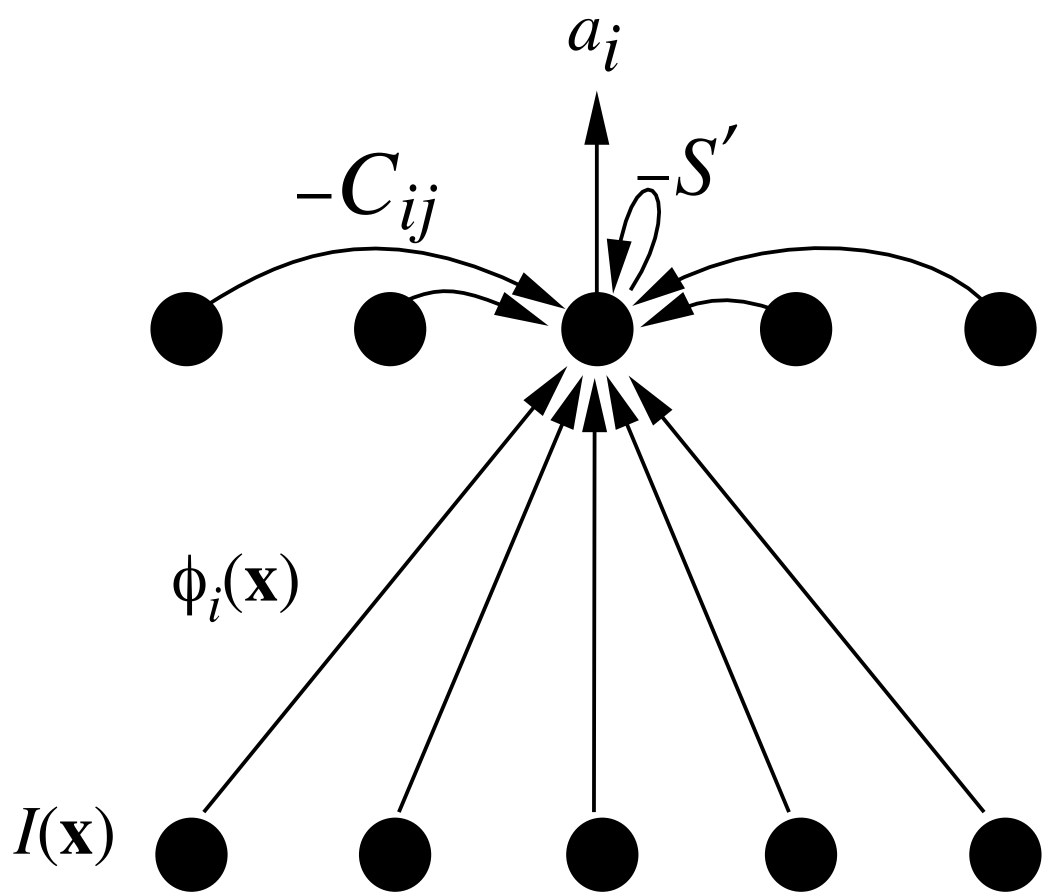
$$= \frac{\lambda_N}{2}|\mathbf{I} - \Phi\mathbf{a}|^2 + \sum_i S(a_i) + \text{const.}$$

Dynamics:

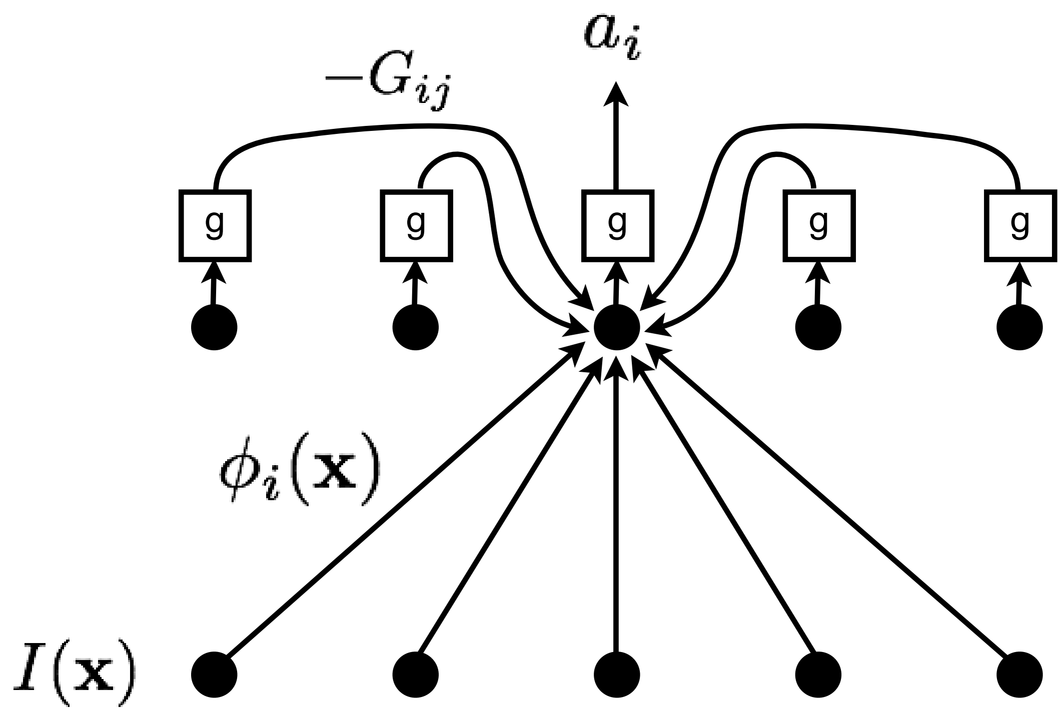
$$\dot{\mathbf{a}} \propto -\frac{\partial E}{\partial \mathbf{a}}$$

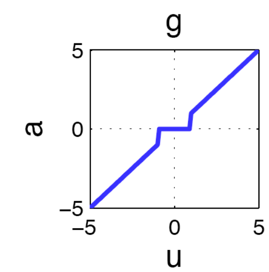
$$= \lambda_N \Phi^T \mathbf{I} - \lambda_N \Phi^T \Phi\, \mathbf{a} - S'(\mathbf{a})$$

Neural circuit implementation

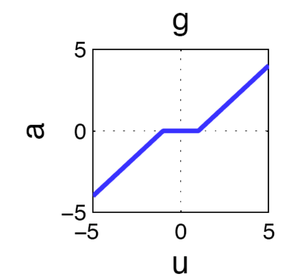


Neural circuit implementation
(much more efficient)





or



$$\tau \dot{u}_i + u_i = b_i - \sum_{j \neq i} G_{ij} a_i$$

← leaky integrator with feedforward excitation and local inhibition

$$a_i = g(u_i)$$

← thresholding

Adaptation (learning)

Objective function:

$$\mathcal{L} = \langle \log P(\mathbf{I}|\theta) \rangle$$

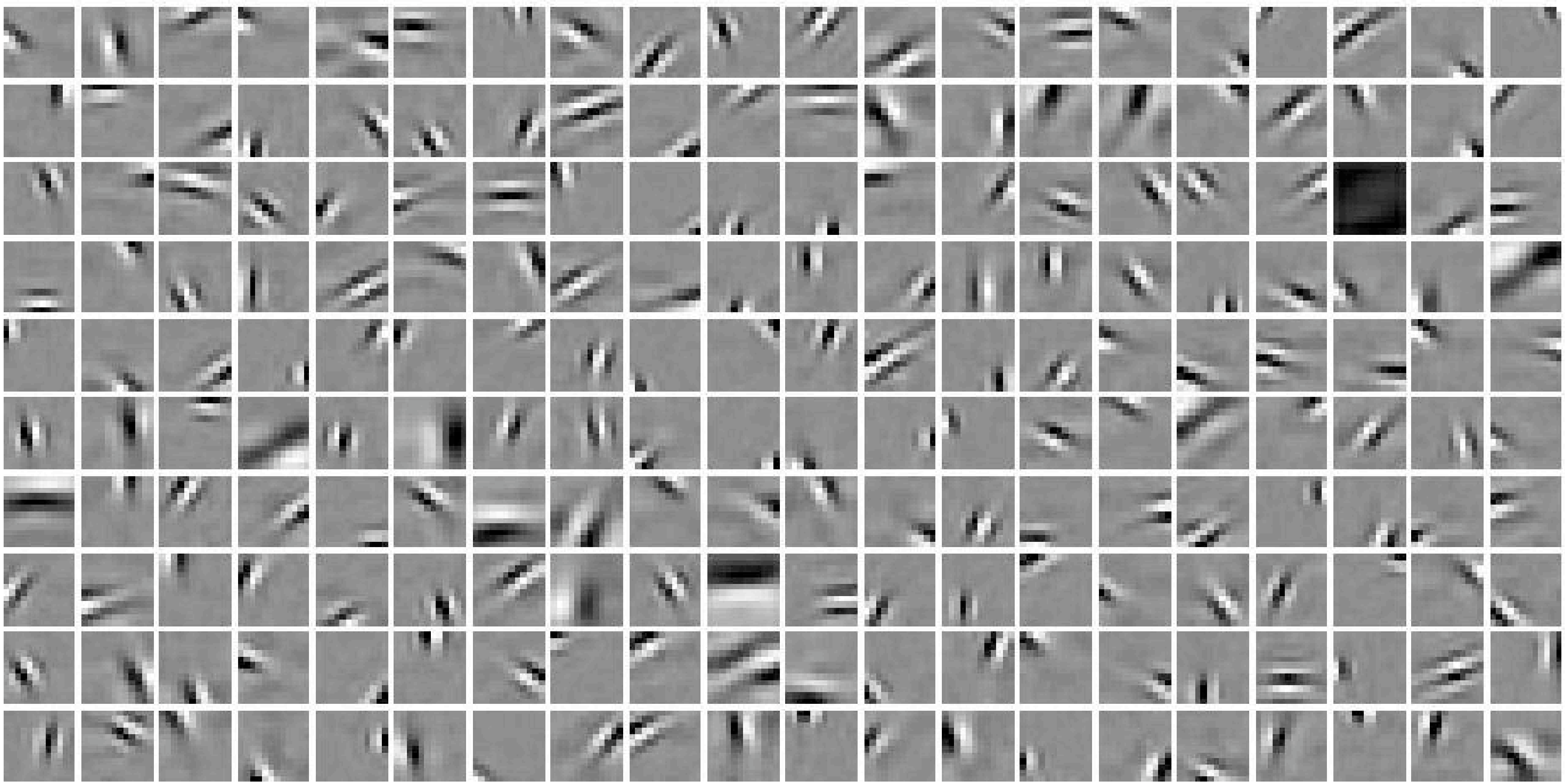
$$P(\mathbf{I}|\theta) = \int P(\mathbf{I}|\mathbf{a}, \theta)\, P(\mathbf{a}|\theta)\, d\mathbf{a}$$

Learning rule:

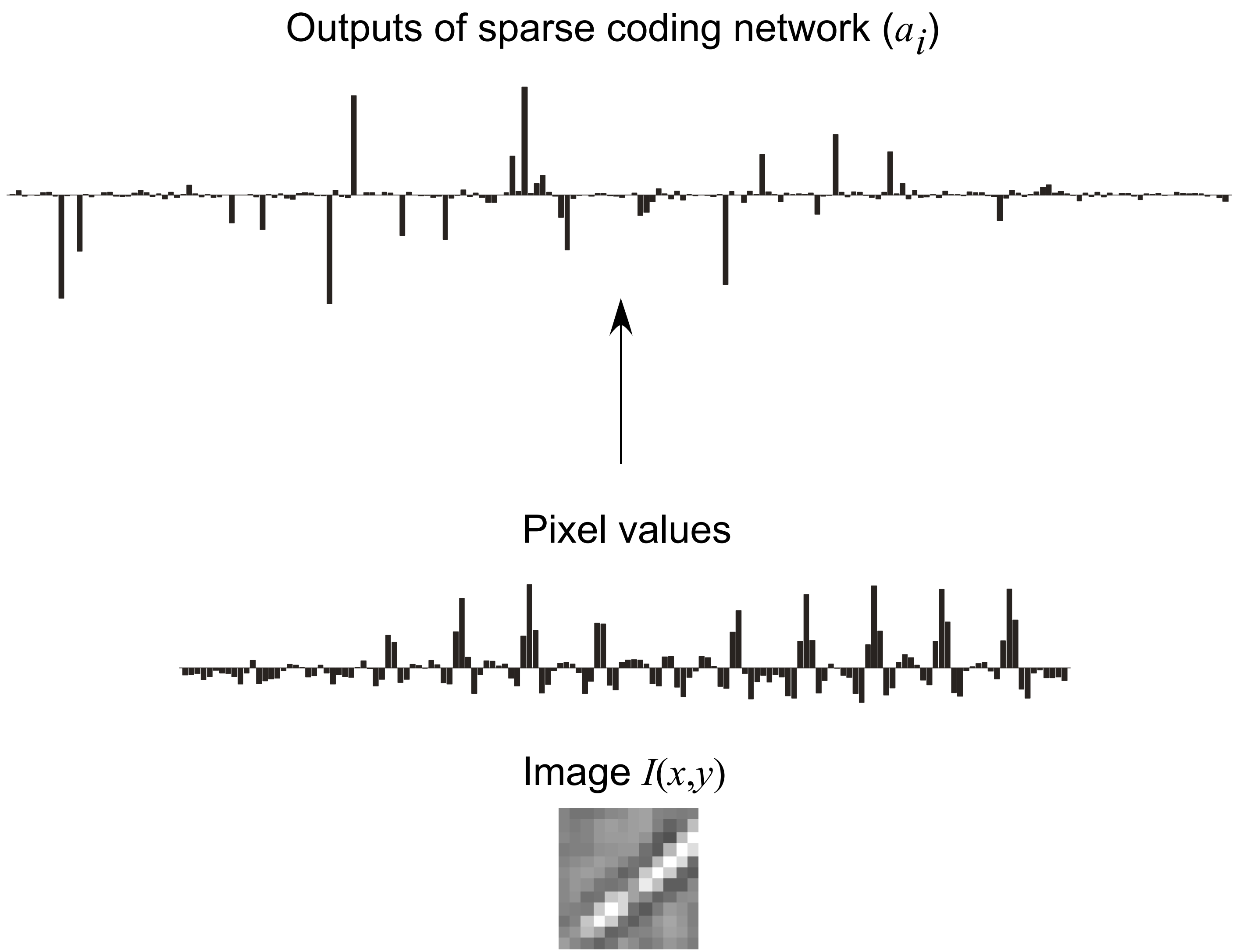
$$\Delta\Phi \propto \frac{\partial \mathcal{L}}{\partial \Phi}$$

$$= \lambda_N \left\langle \int [I - \Phi\, \mathbf{a}]\; \mathbf{a}^T\, P(\mathbf{a}|\mathbf{I}, \theta)\, d\mathbf{a} \right\rangle$$

Learned basis functions
(200, 12x12 pixels)

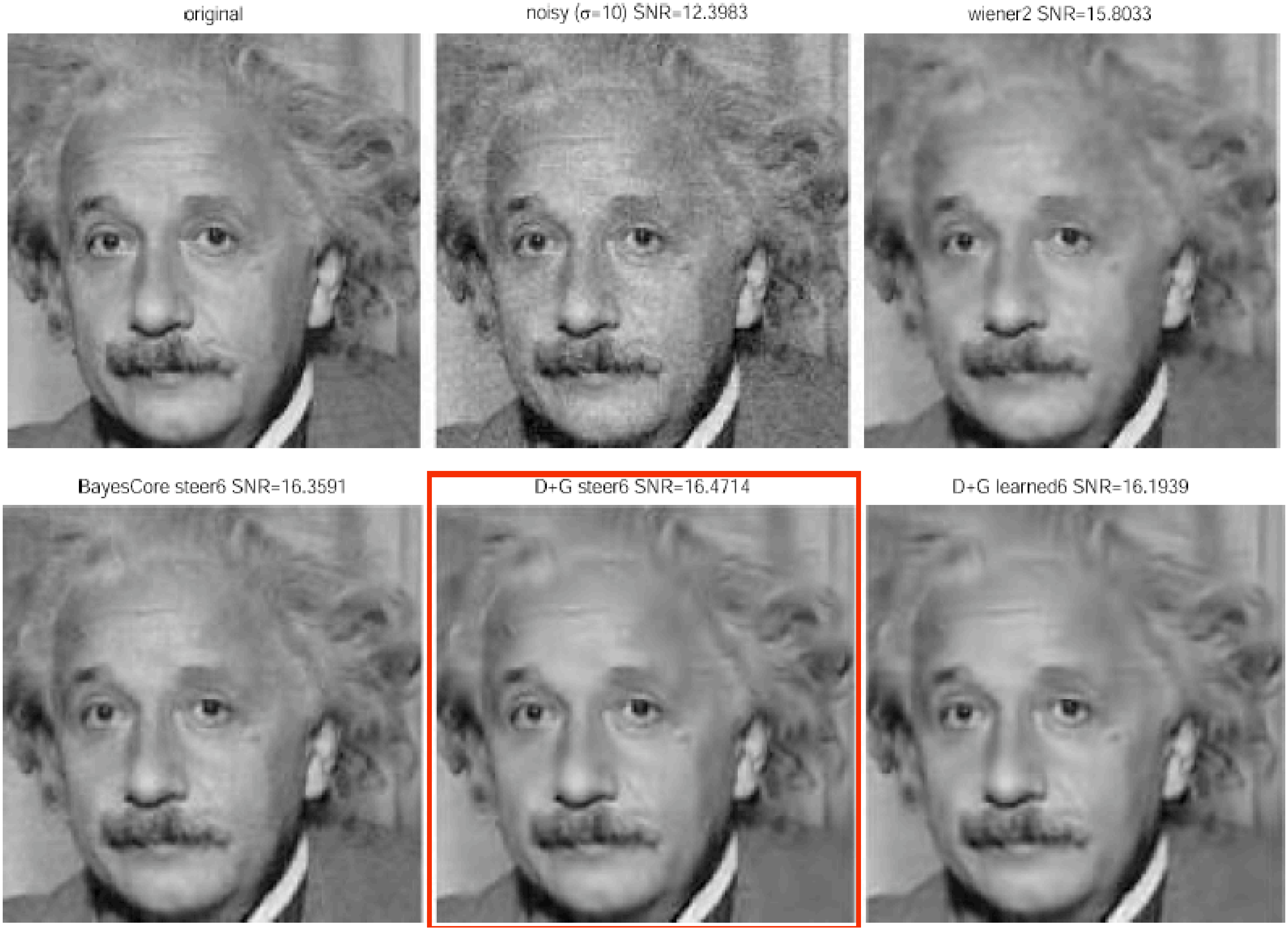


Sparsification



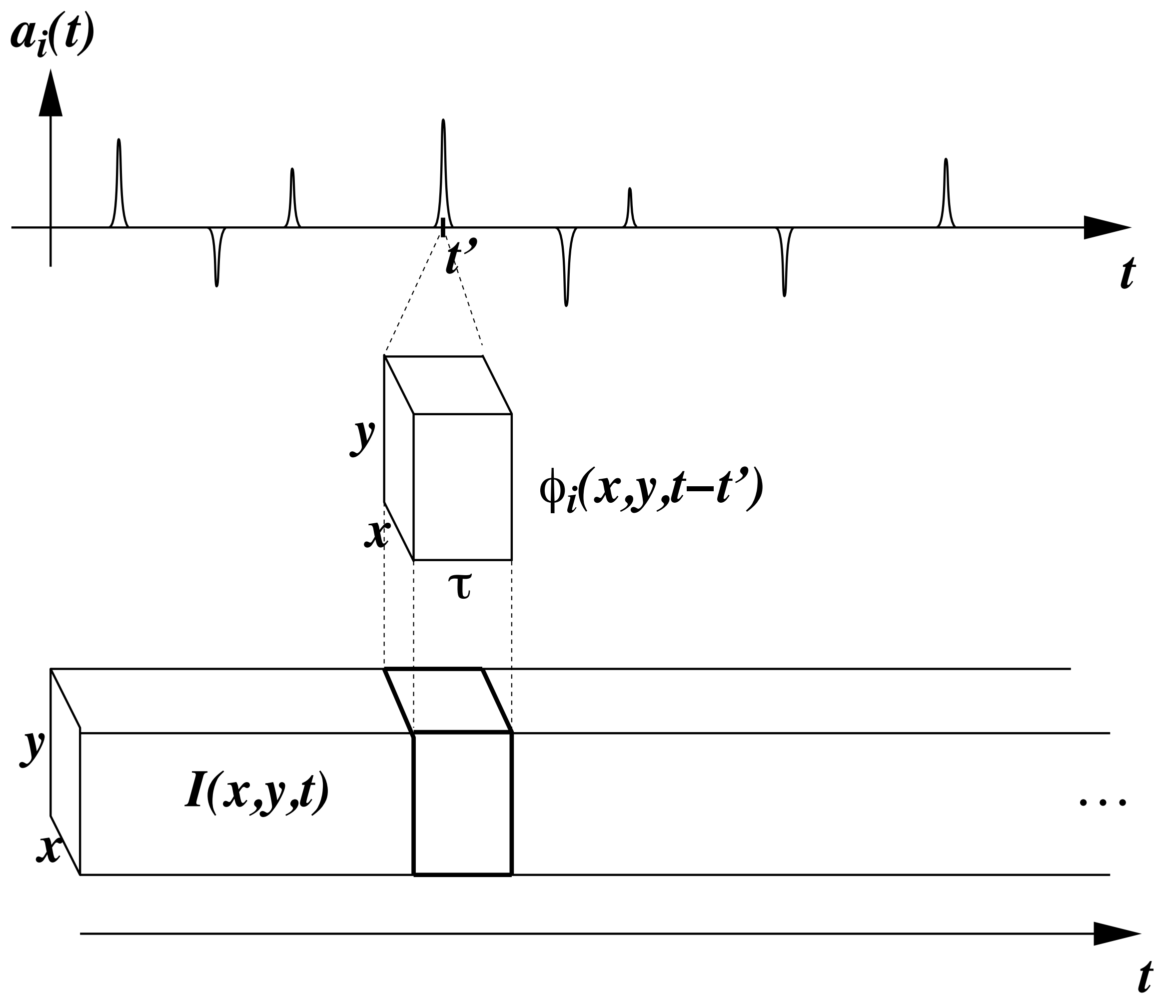
Denoising

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.

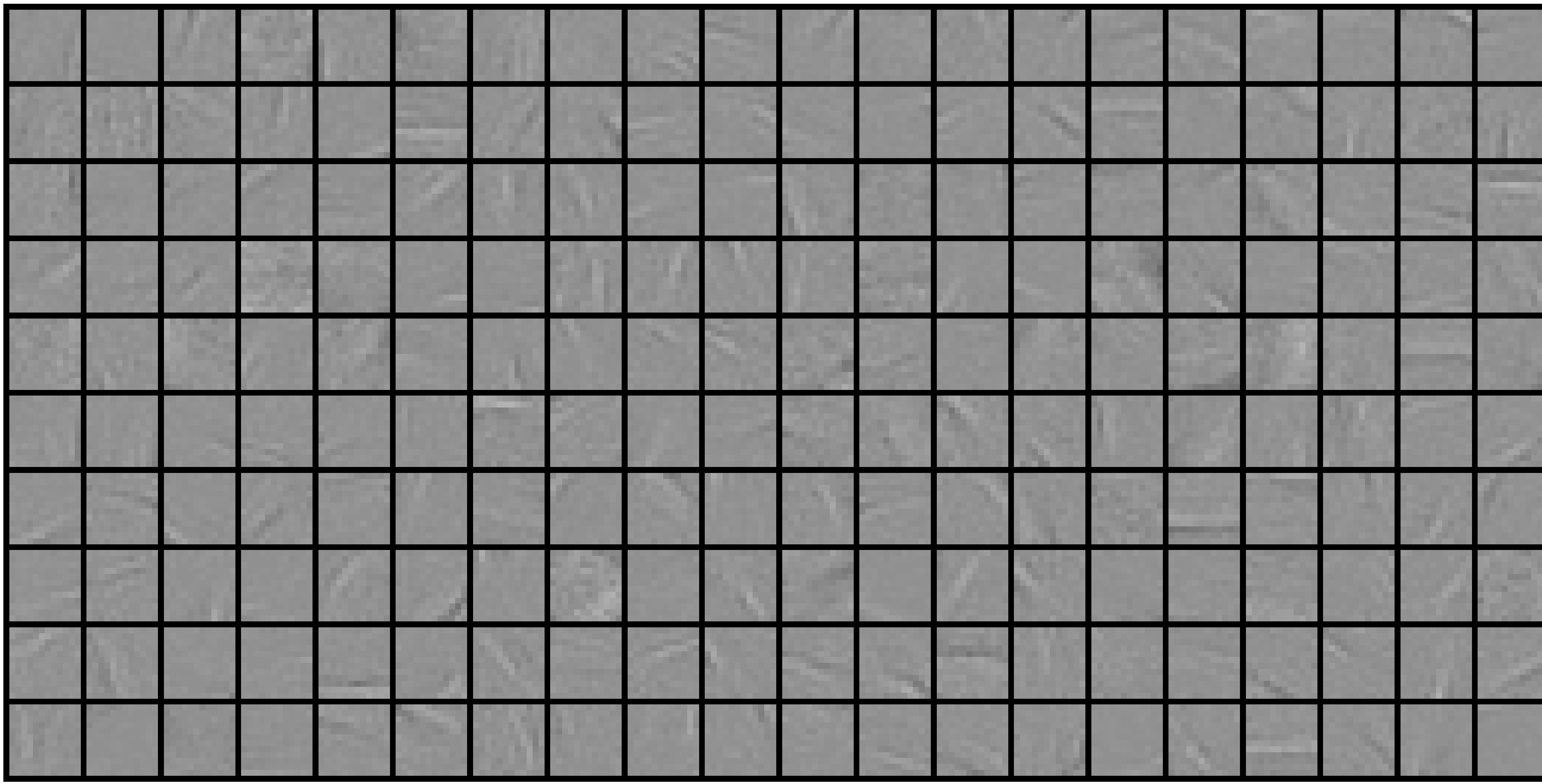

original
noisy (σ=10) SNR=12.3983
wiener2 SNR=15.8033
BayesCore steer6 SNR=16.3591
D+G steer6 SNR=16.4714
D+G learned6 SNR=16.1939

Space-time image model

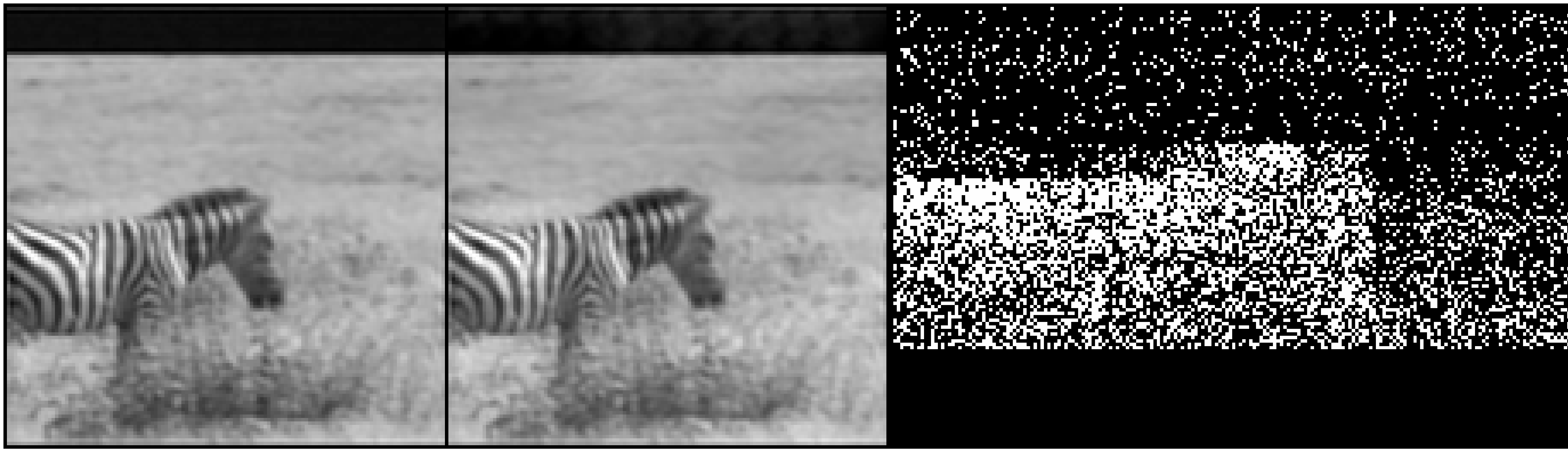
$$I(x,y,t) = \sum_i a_i(t) * \phi_i(x,y,t) + \nu(x,y,t)$$

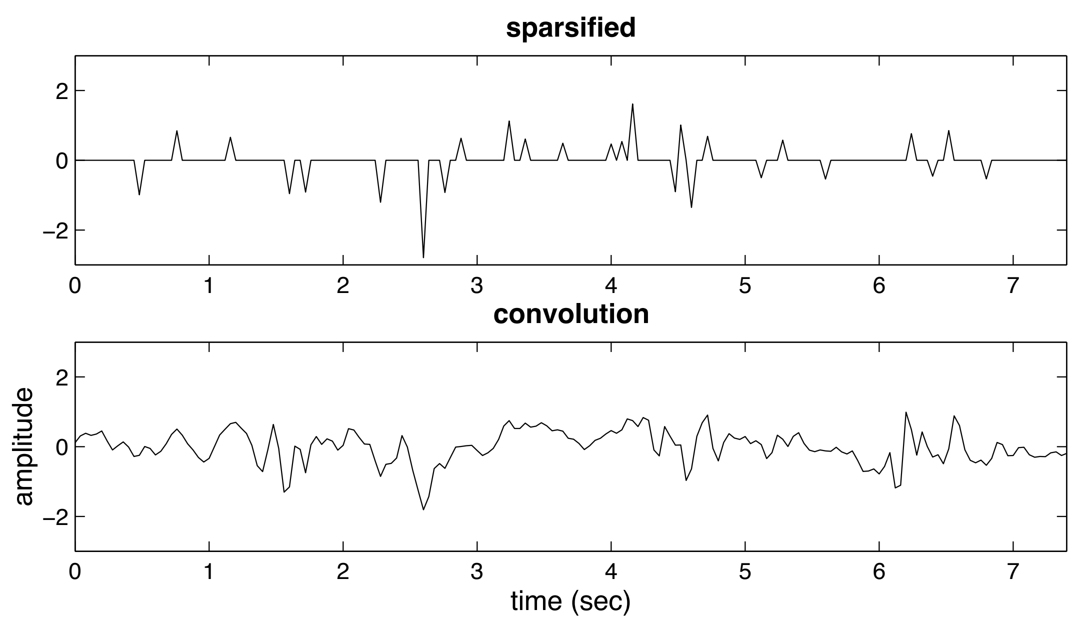


Learned basis space-time basis functions
(200 bfs, 12 x12 x 7)



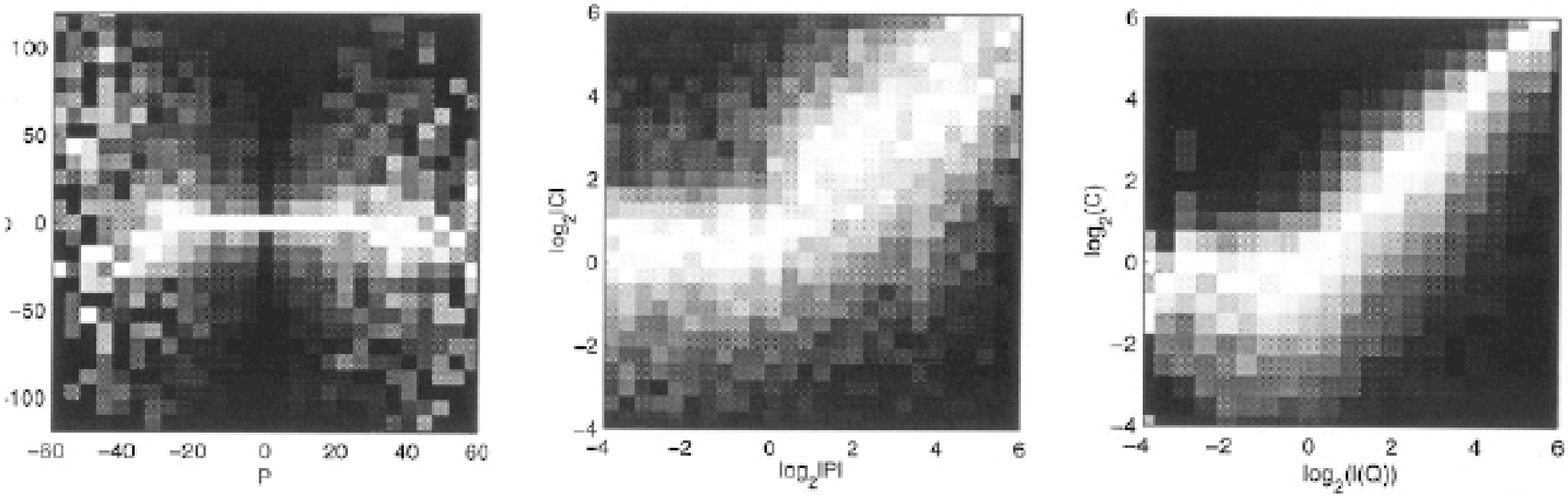
Sparse coding and reconstruction



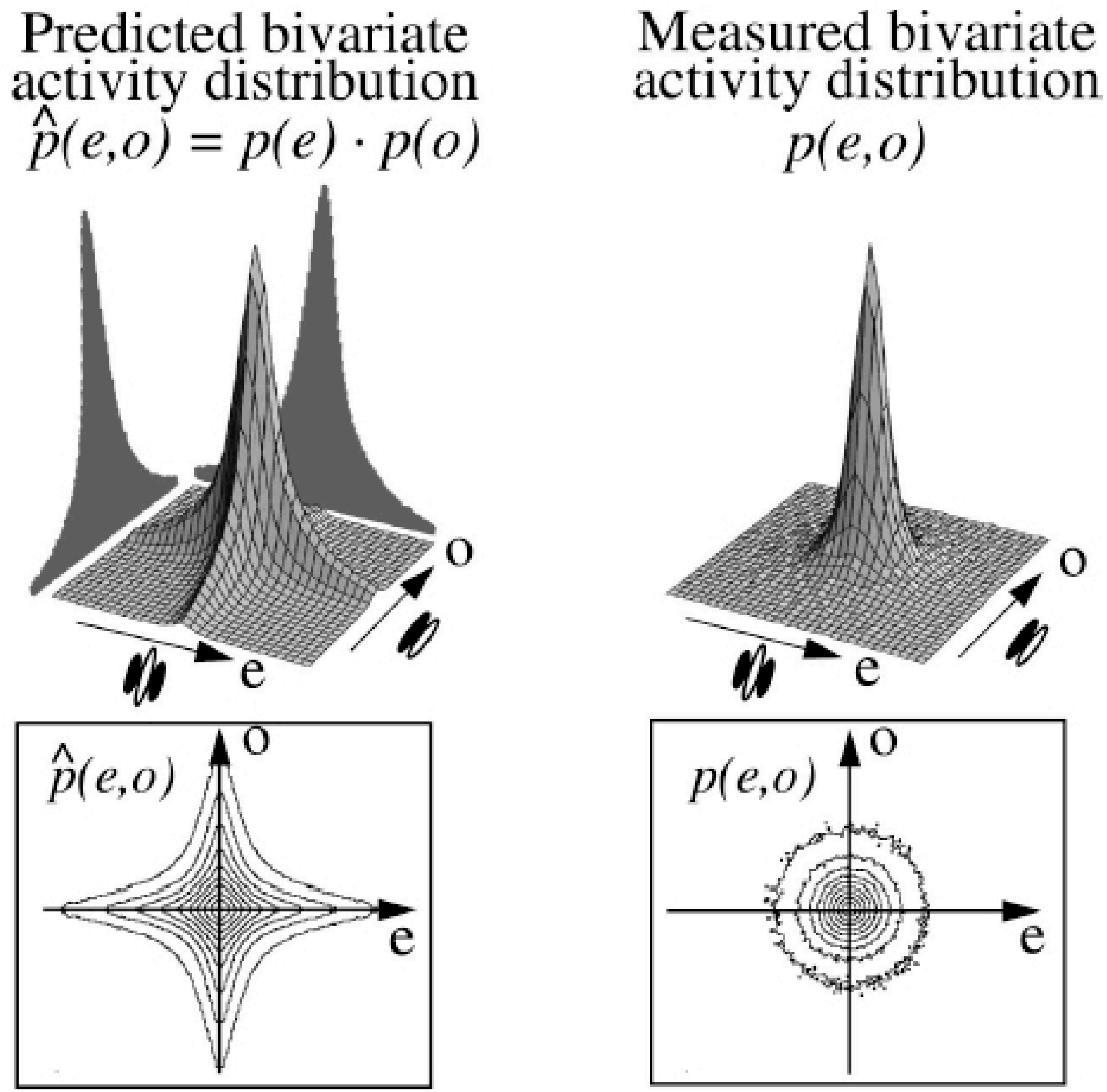


Statistical dependencies among coefficients

Buccigrossi & Simoncelli (1997)



Statistical dependencies among coefficients
(Zetzsche et al., 1999)

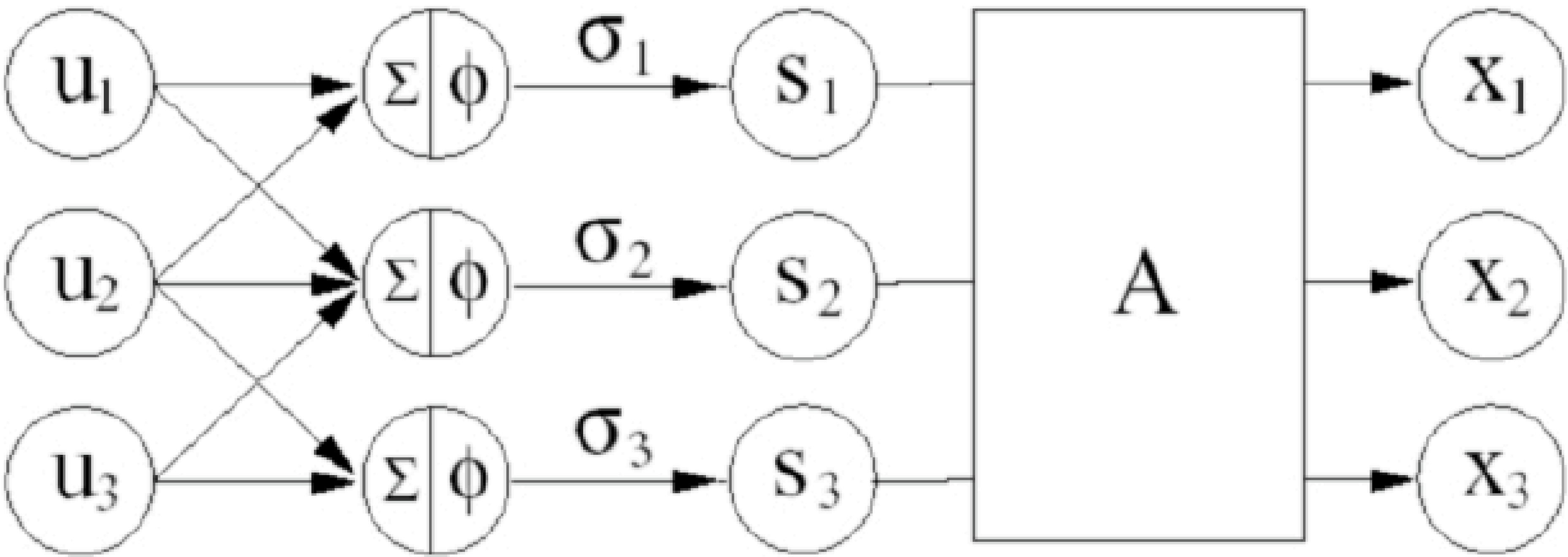


Hierarchical models for capturing dependencies among sparse components

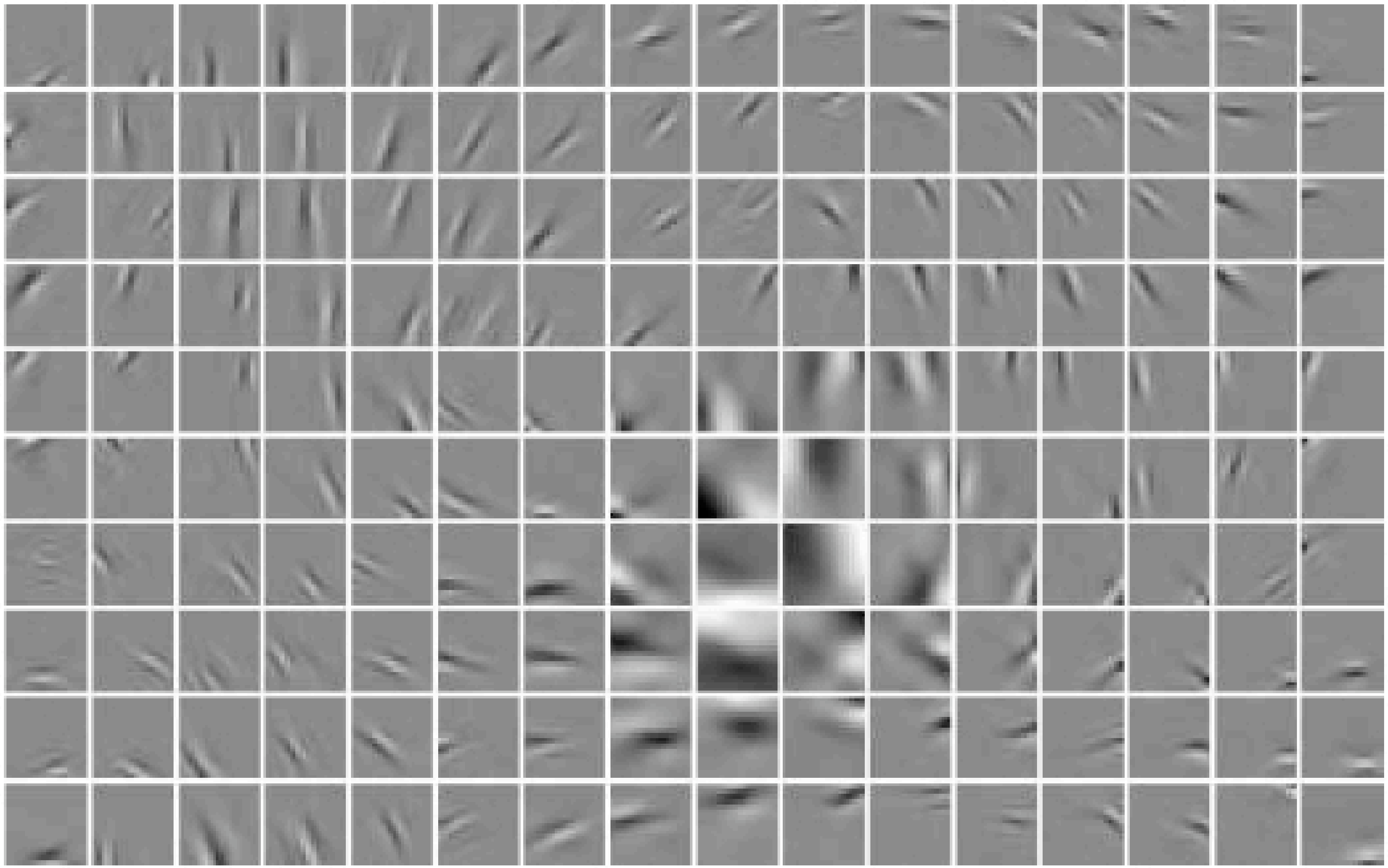
$$
\begin{aligned}
a_i &= \overbrace{\sigma_i}^{\text{power-law}} \times \overbrace{z_i}^{\text{Gaussian}} \\
\sigma_i &= f\Big(\sum_j \Psi_{ij}\, b_j\Big)
\end{aligned}
$$

Wainwright & Simoncelli (2002)
Hyvarinen & Hoyer (2002)
Karklin & Lewicki (2003)
Schwartz & Sejnowski (2004)
Osindero & Hinton (2005)

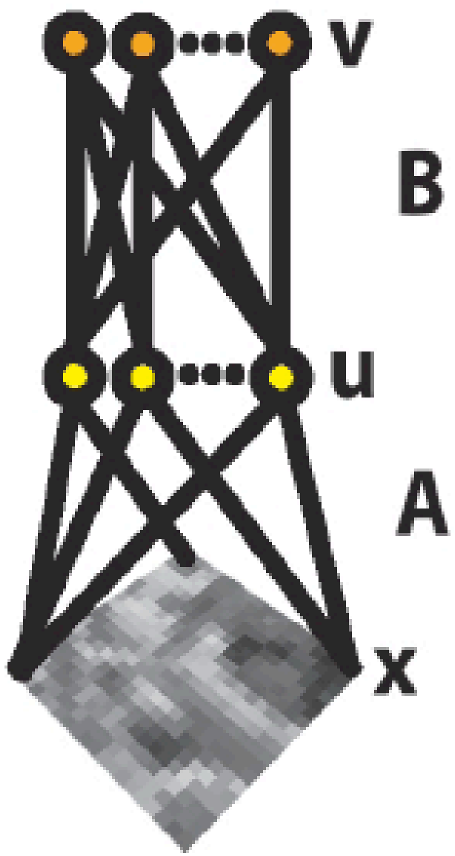
'Topographic ICA' - Hyvarinen & Hoyer (2002)



'Topographic ICA' - Hyvarinen & Hoyer (2002)



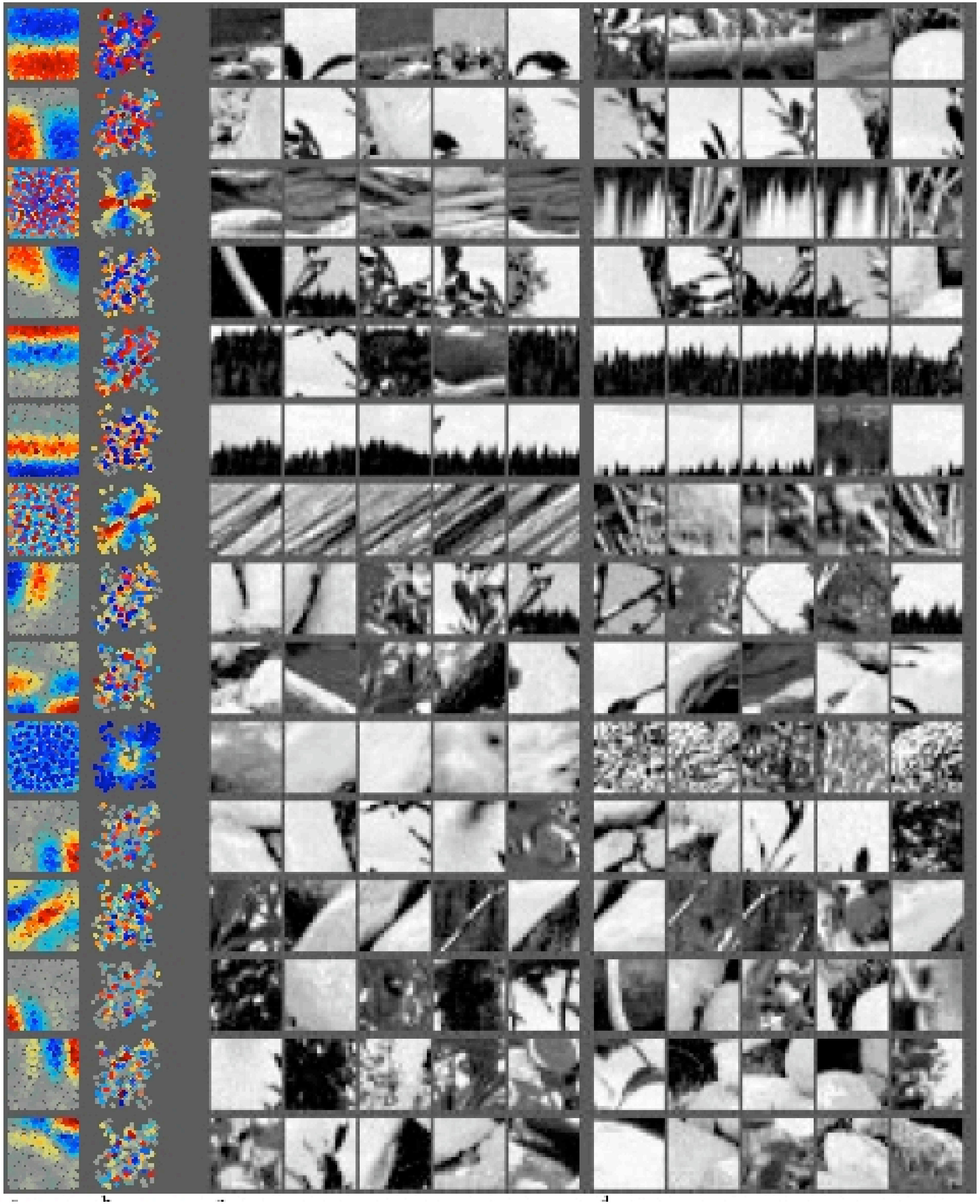
Learning the neighborhoods - Karklin & Lewicki (2003)



$$\mathbf{x} = \mathbf{A}\,\mathbf{u}$$

$$u_i = \sigma_i\, z_i$$

$$\sigma_i = e^{\sum_j B_{ij} v_j}$$

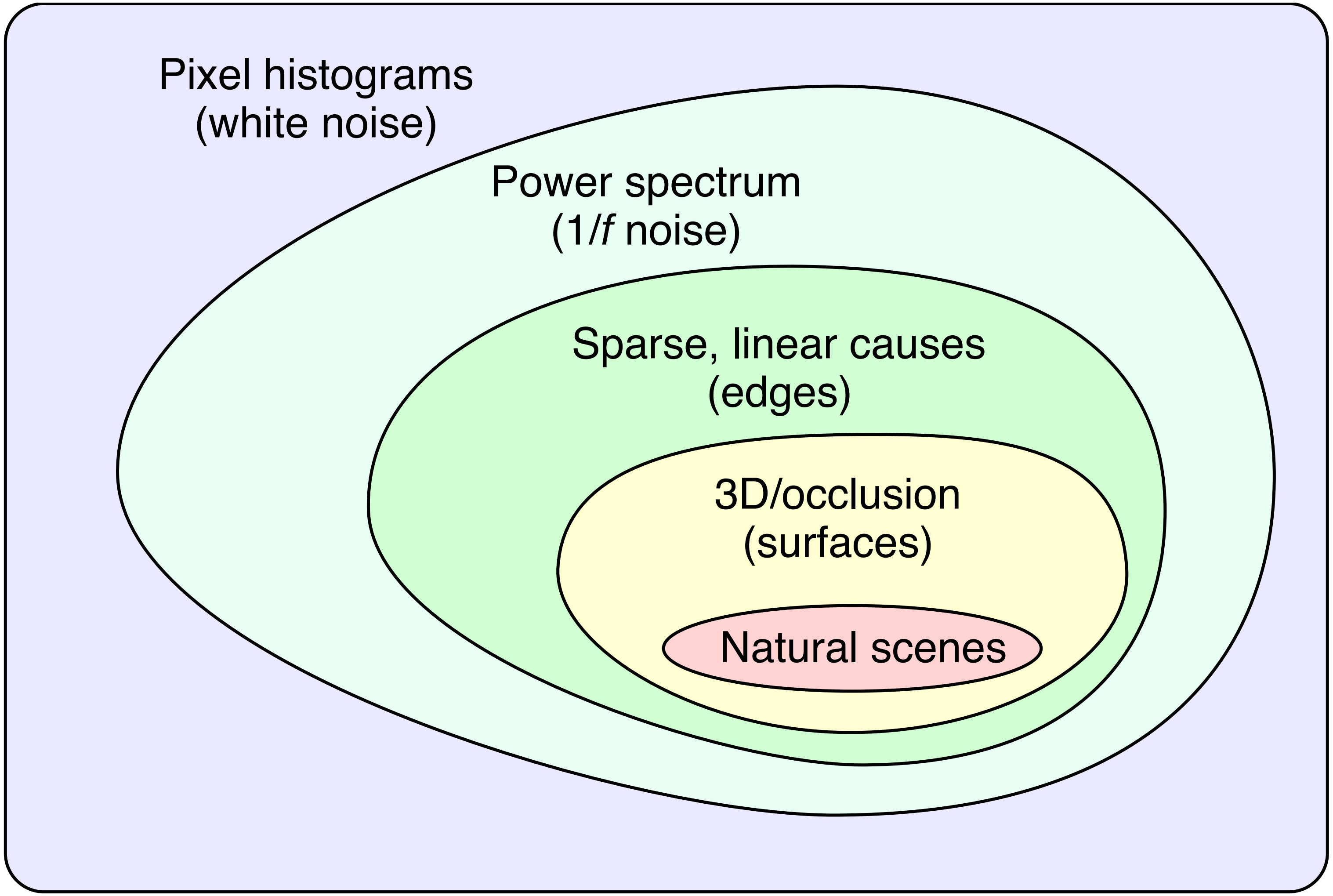


Bilinear models for learning invariant representations

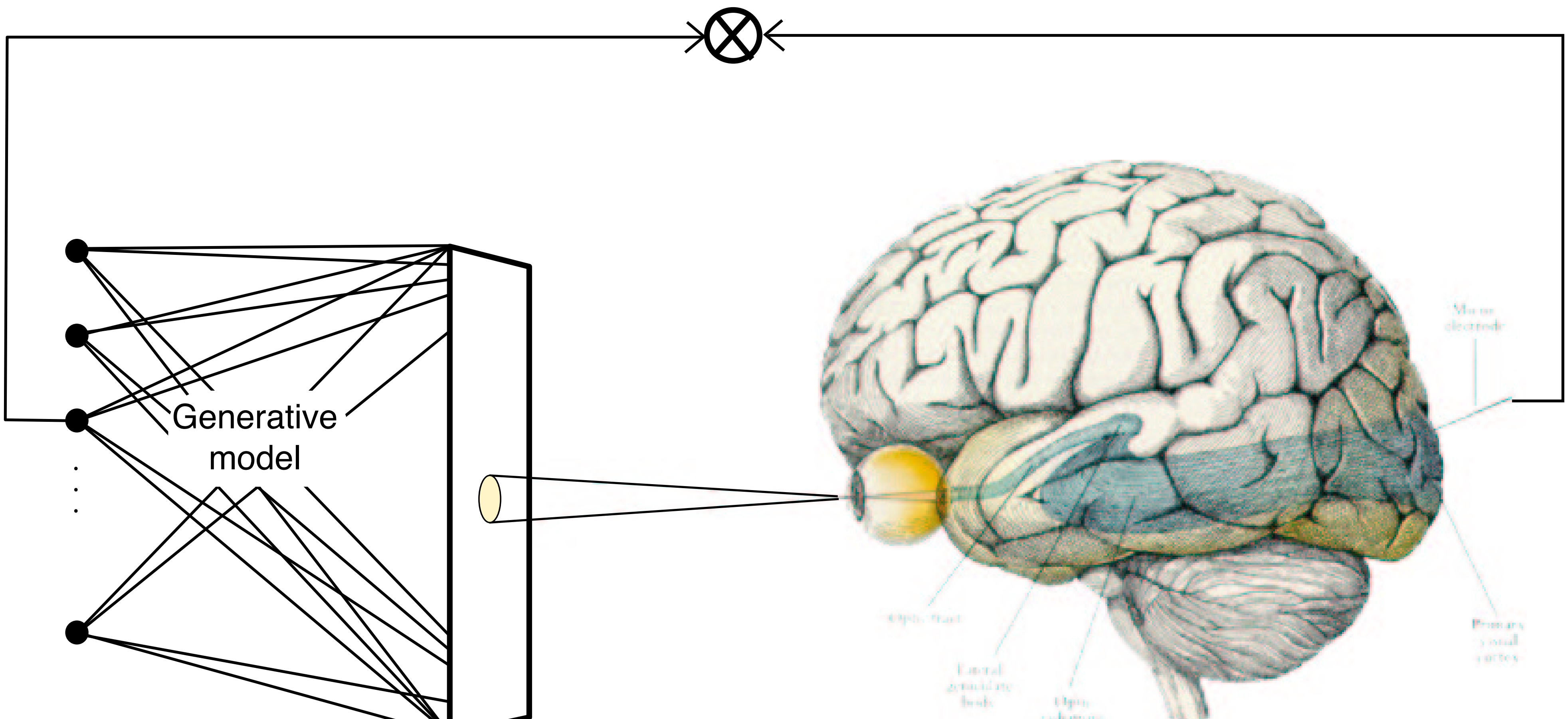
$$\mathbf{z} = \sum_{ij} \mathbf{w}_{ij} \underbrace{x_i}_{\text{`what'}} \underbrace{y_j}_{\text{`where'}}$$

- Tenenbaum & Freeman (2000) - SVD
- Grimes & Rao (2005) - sparse coding

Image models



Generative models as experimental tools



Further information and papers

http://redwood.berkeley.edu/bruno

baolshausen@berkeley.edu