What I think the other 85% is doing

Bruno A. Olshausen

REDWOOD
Neuroscience Institute

&

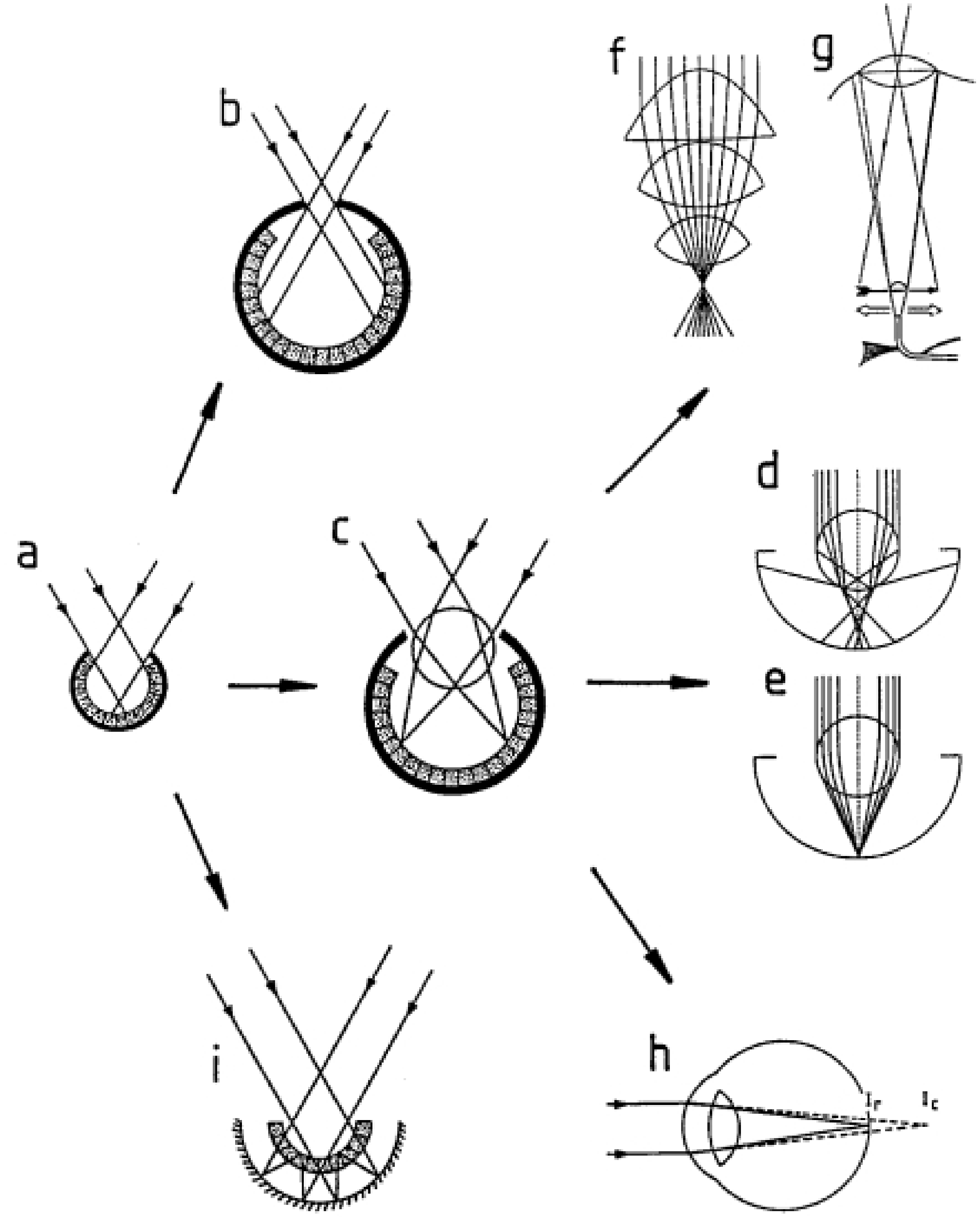
Center for Neuroscience and
Dept. of Neurobiology, Physiology & Behavior, UC Davis

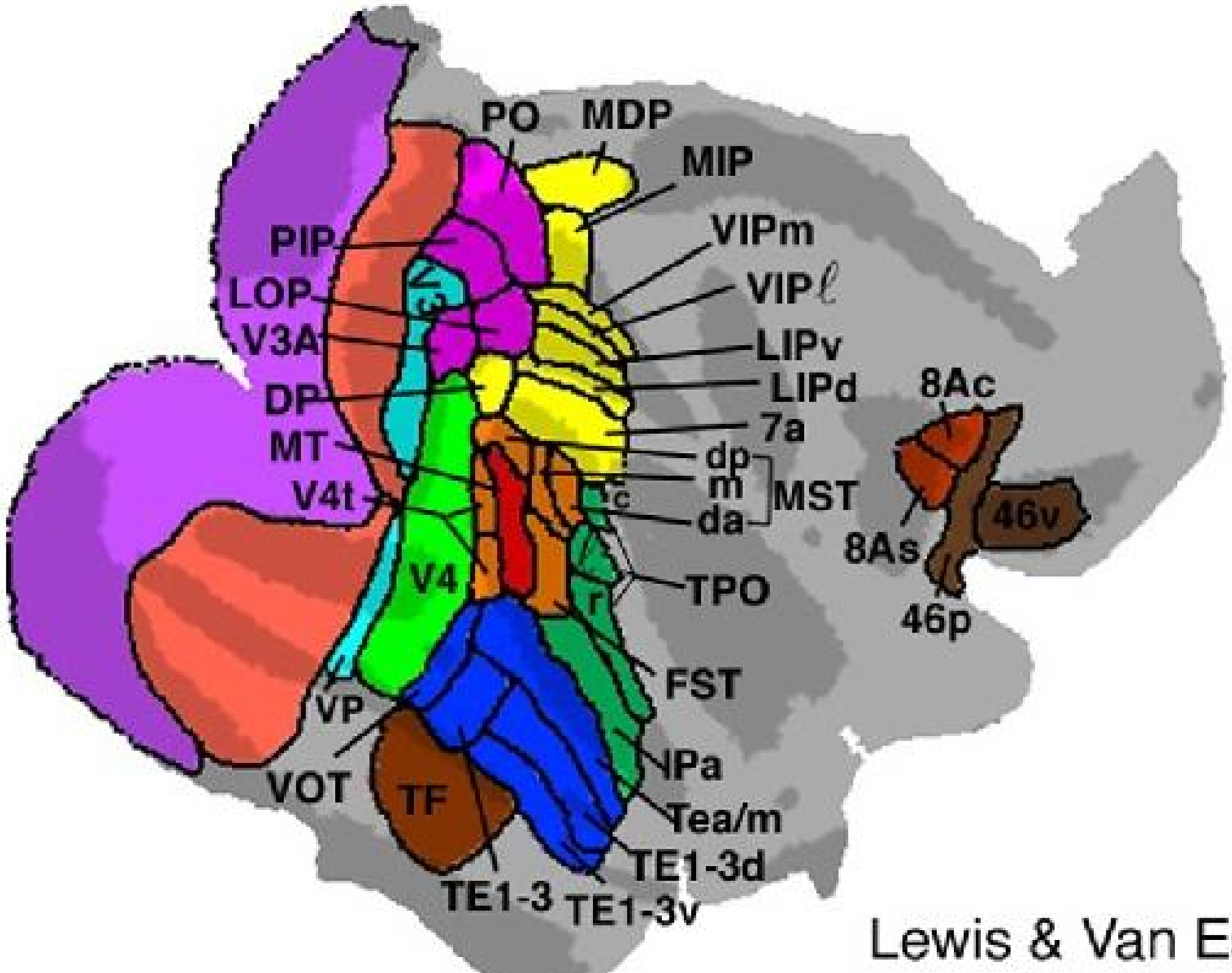
Main Points

- Vision as inference
- Sparse coding
- Learning what and where in images

THE EVOLUTION OF EYES

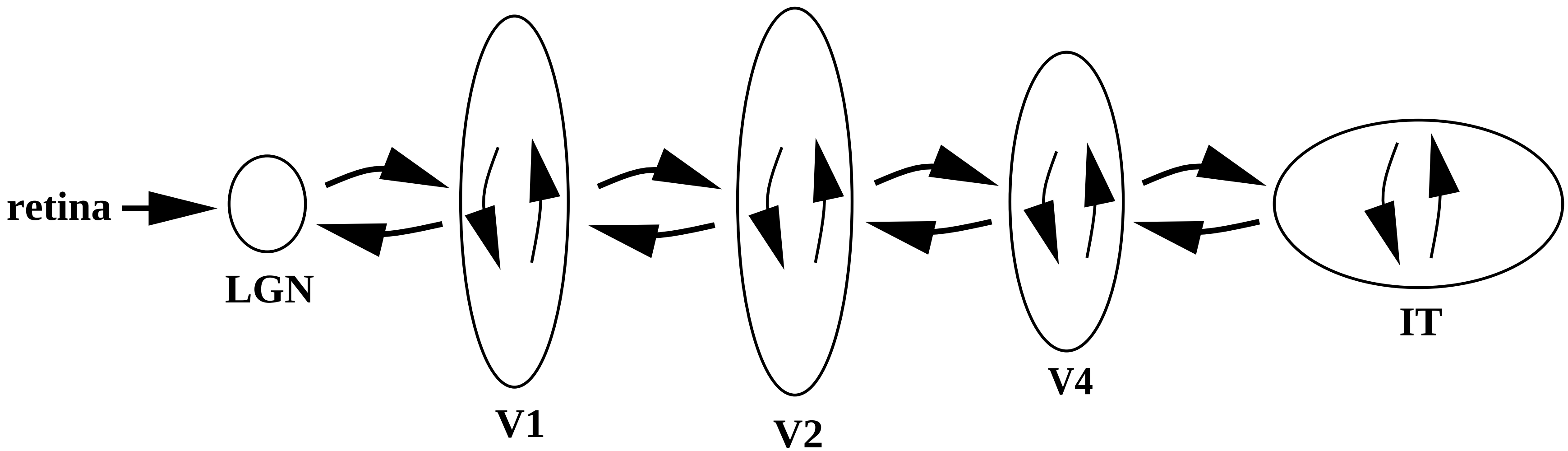
Michael F. Land *Russell D. Fernald*



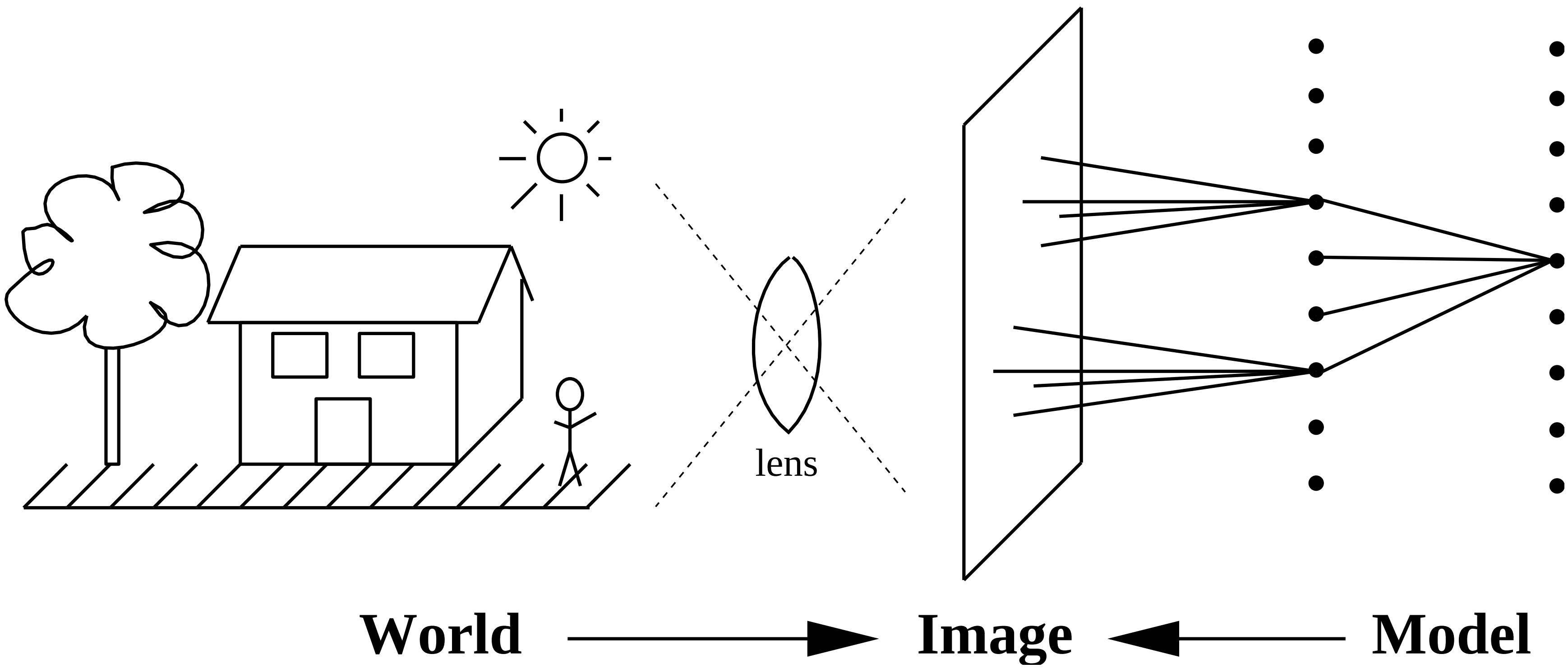


Lewis & Van Essen (2000)
visual areas

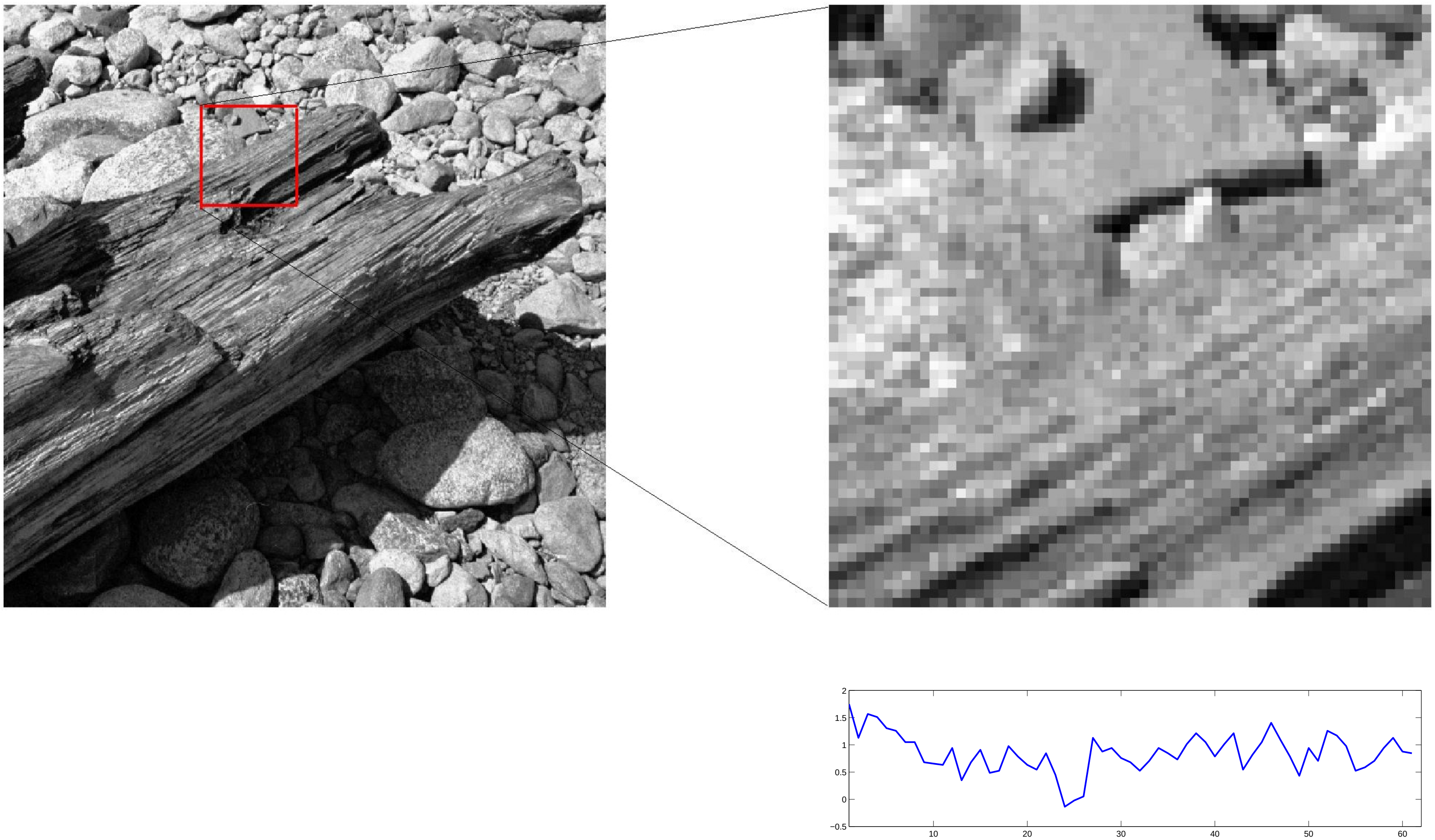
Recurrent computation is pervasive throughout cortex



Vision as inference



Natural scenes are filled with ambiguity

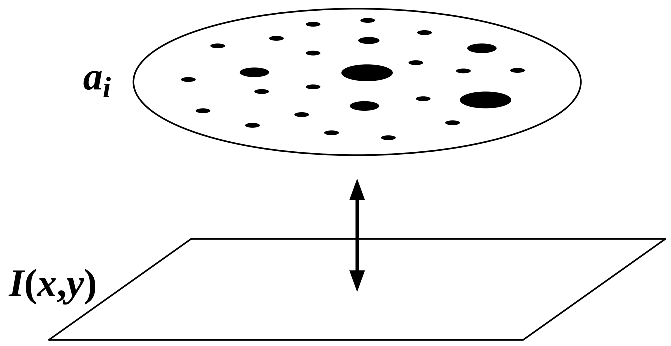


Bayes' rule

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

$$\begin{aligned} E &= \text{the actual state of the environment} \\ D &= \text{data about the environment} \end{aligned}$$

Sparse coding



- Provides a simple description of images
- Makes image structure explicit $\longrightarrow$ Grouping
- Makes it easier to learn associations
- Field's (1987) analysis of simple-cell receptive fields suggests they have been optimized for sparseness.

Overcomplete representations

- In oriented, multiscale pyramids, overcompleteness is necessary for shiftability (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

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

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

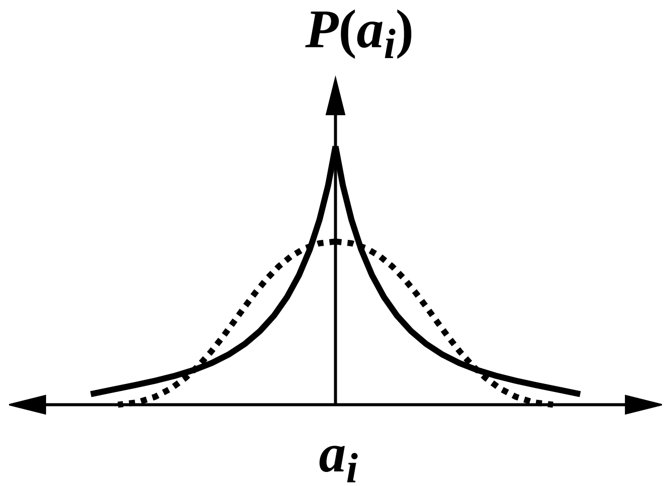
$$P(\mathbf{I}|\theta) = \int P(\mathbf{I}|\mathbf{a},\theta) \, P(\mathbf{a}|\theta) \, d\mathbf{a}$$

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



Objective functions for inference and learning

Inference (perception):

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

Learning:

$$\langle \log P(\mathbf{I}|\theta) \rangle = \left\langle \log \int P(\mathbf{I}|\mathbf{a},\theta)\, P(\mathbf{a}|\theta)\, d\mathbf{a} \right\rangle$$

Energy function

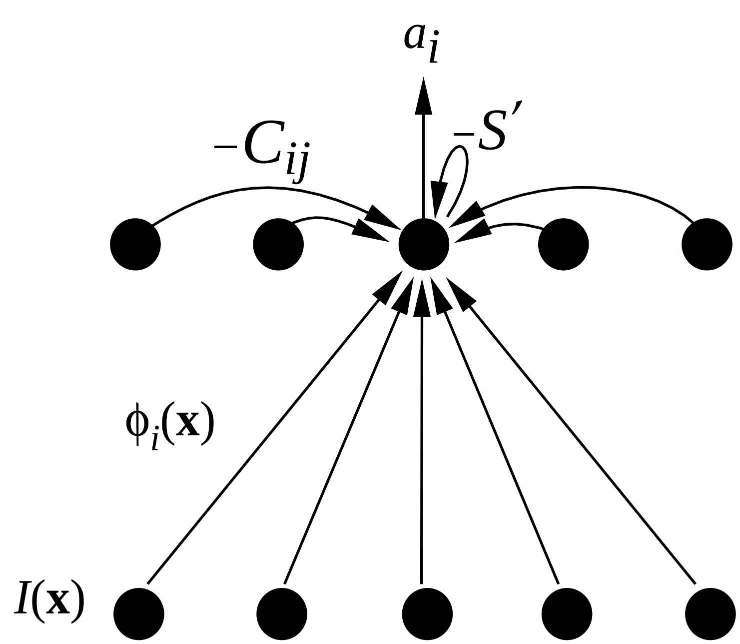
$$
\begin{aligned}
E &= \log P(\mathbf{a}|\mathbf{I},\theta) \\
&= \frac{\lambda_N}{2}\sum_{x,y}\left[I(x,y) - \sum_i a_i\phi_i(x,y)\right]^2 + \sum_i S(a_i) + \text{const.}
\end{aligned}
$$

Dynamics

$\dot{a}_i \propto -\frac{\partial E}{\partial a_i}$:

$$
\begin{aligned}
\tau \dot{a}_i &= b_i - \sum_j C_{ij} a_j - S'(a_i) \\
b_i &= \lambda_N \sum_{x,y} \phi_i(x,y)\, I(x,y) \\
C_{ij} &= \lambda_N \sum_{x,y} \phi_i(x,y) \phi_j(x,y)
\end{aligned}
$$

Network implementation

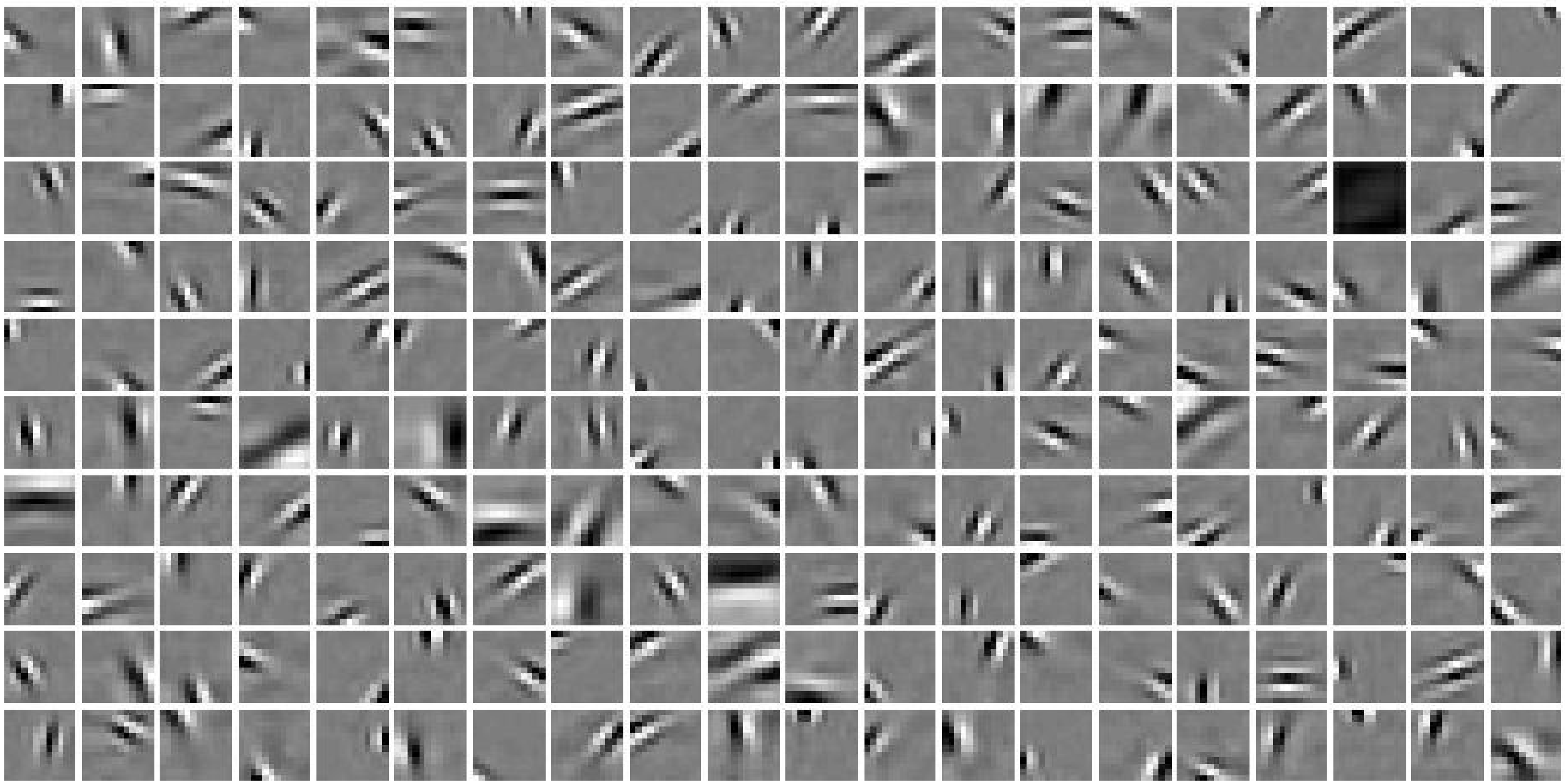


Learning

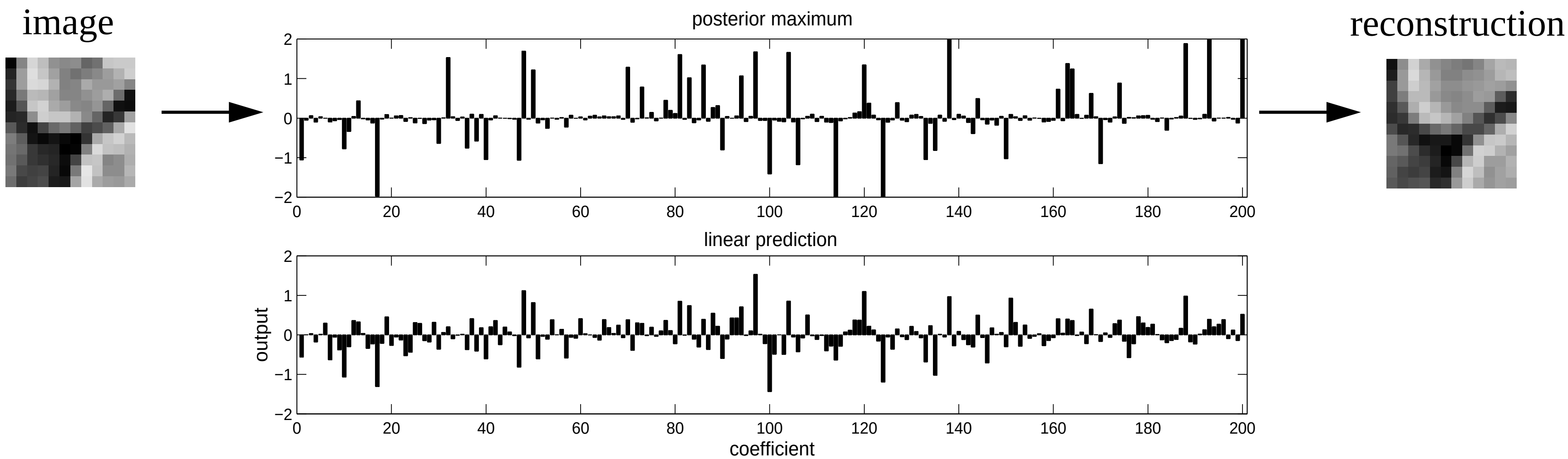
$\Delta\phi_i \propto -\left\langle \frac{\partial E}{\partial \phi_i} \right\rangle$:

$$\begin{aligned} \Delta\phi_i(x,y) &= \eta\,\langle a_i\,r(x,y)\rangle \\ r(x,y) &= I(x,y) - \sum_i a_i\phi_i(x,y) \end{aligned}$$

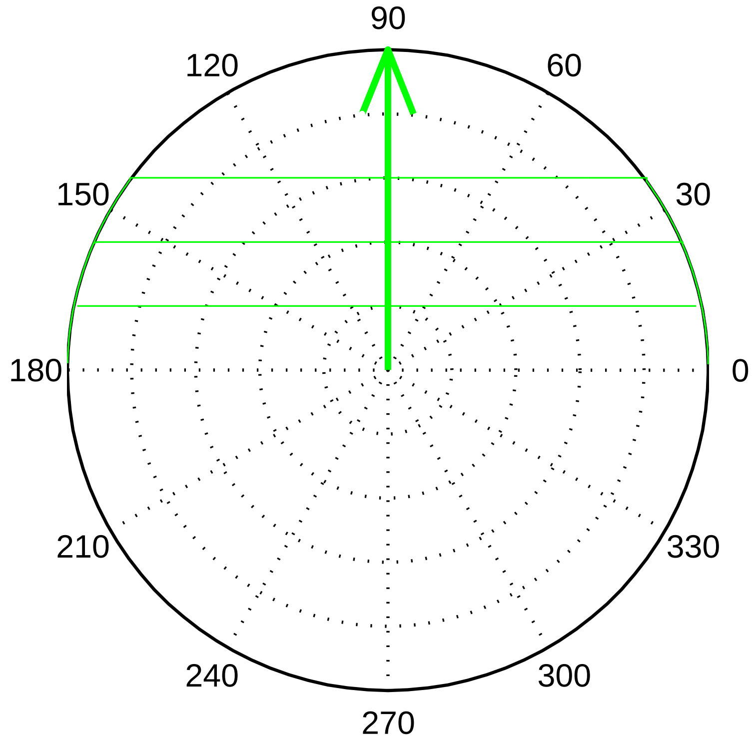
Learned basis functions (200, 12x12)



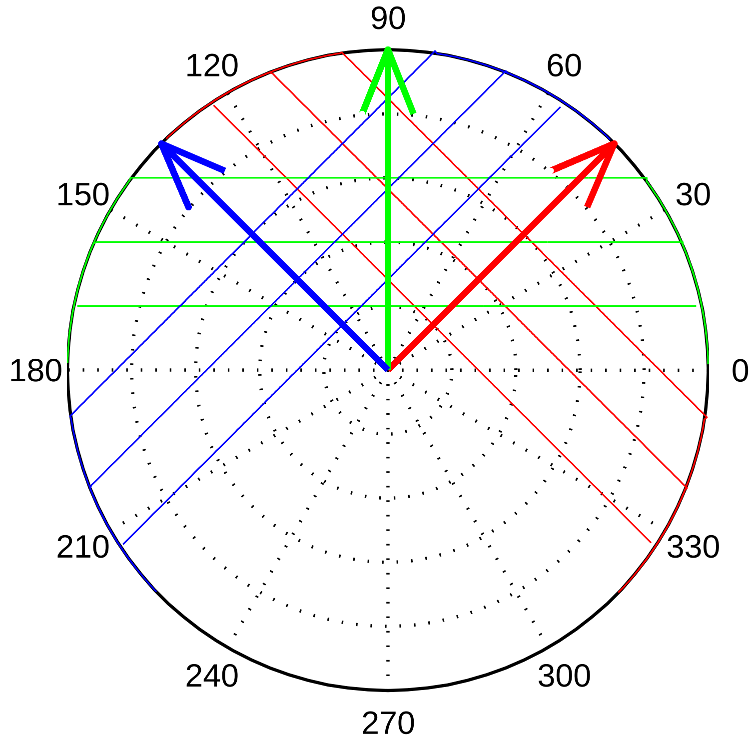
Sparsification



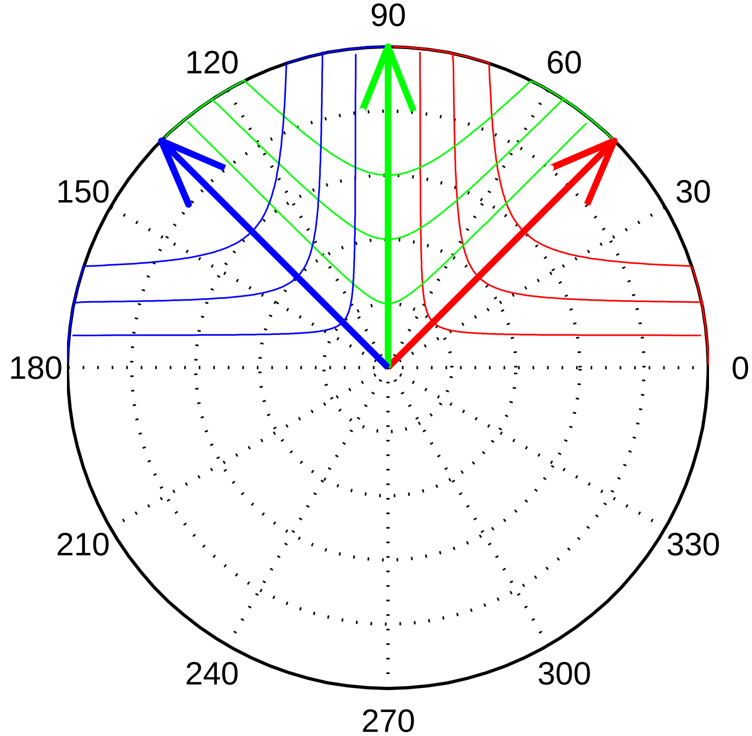
Iso-response contours



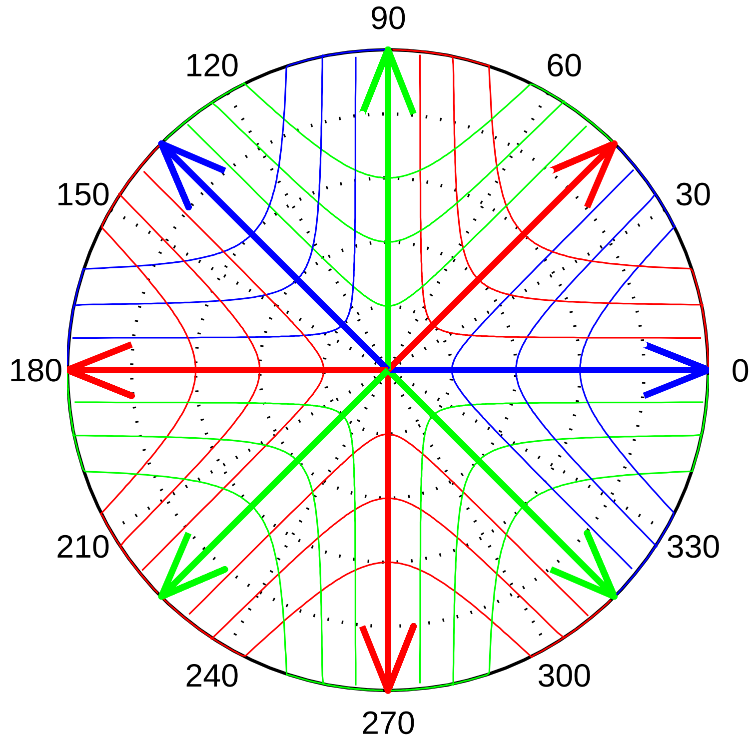
Iso-response contours



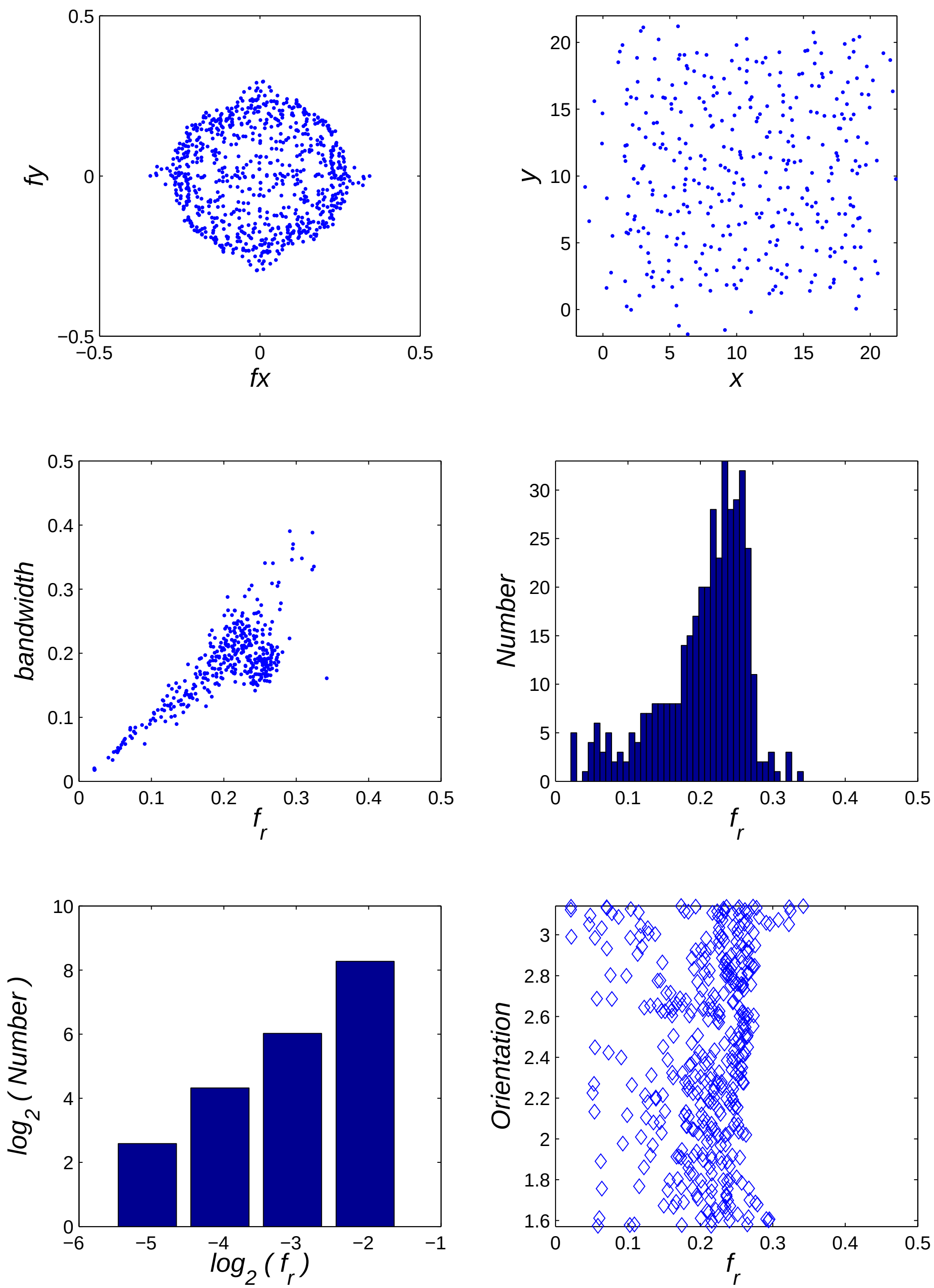
Iso-response contours



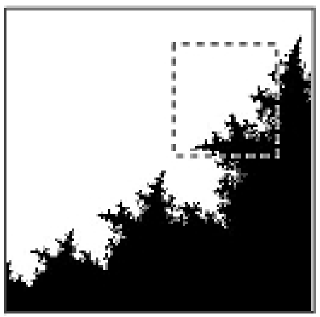
Iso-response contours

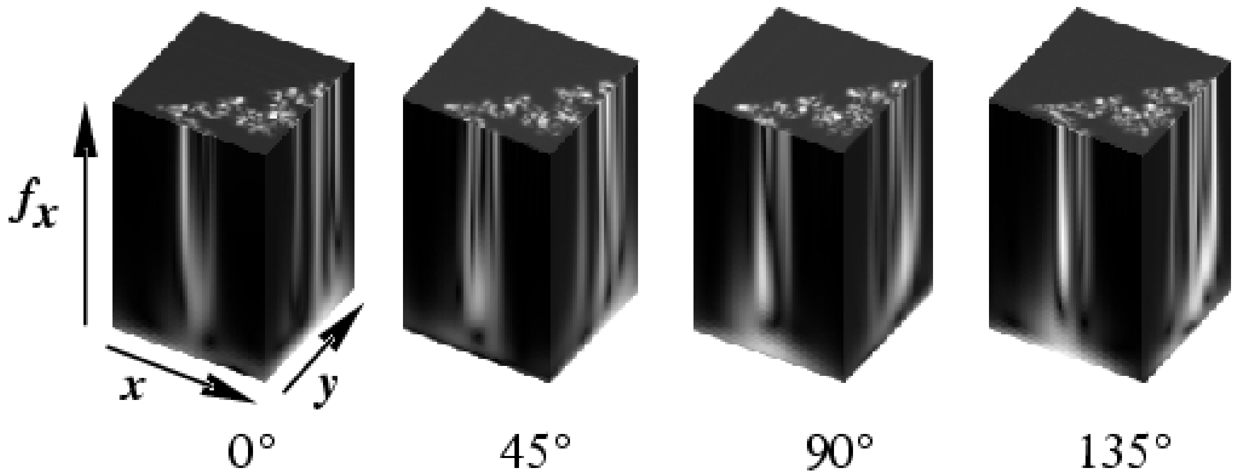


Tiling properties



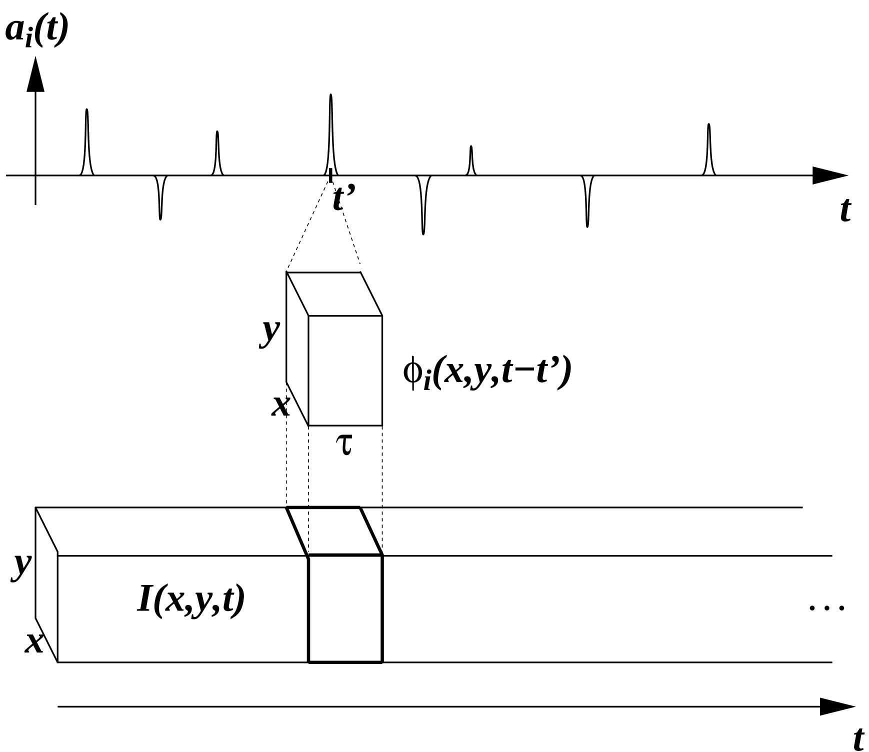
Scale space cross-section of a fractal contour





Space-time image model

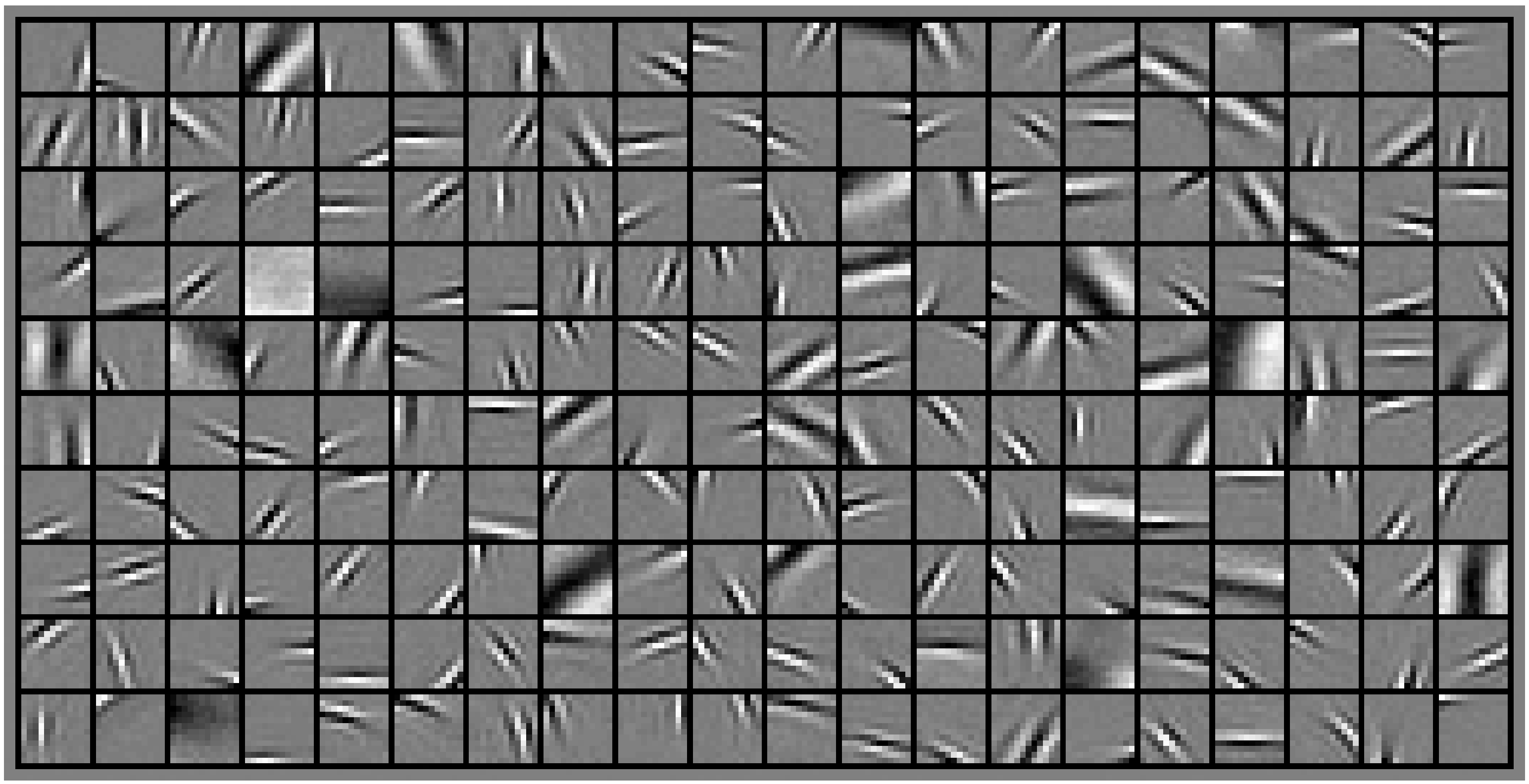
$$I(x, y, t) = \sum_i a_i(t) * \phi_i(x, y, t) + \nu(x, y, t)$$



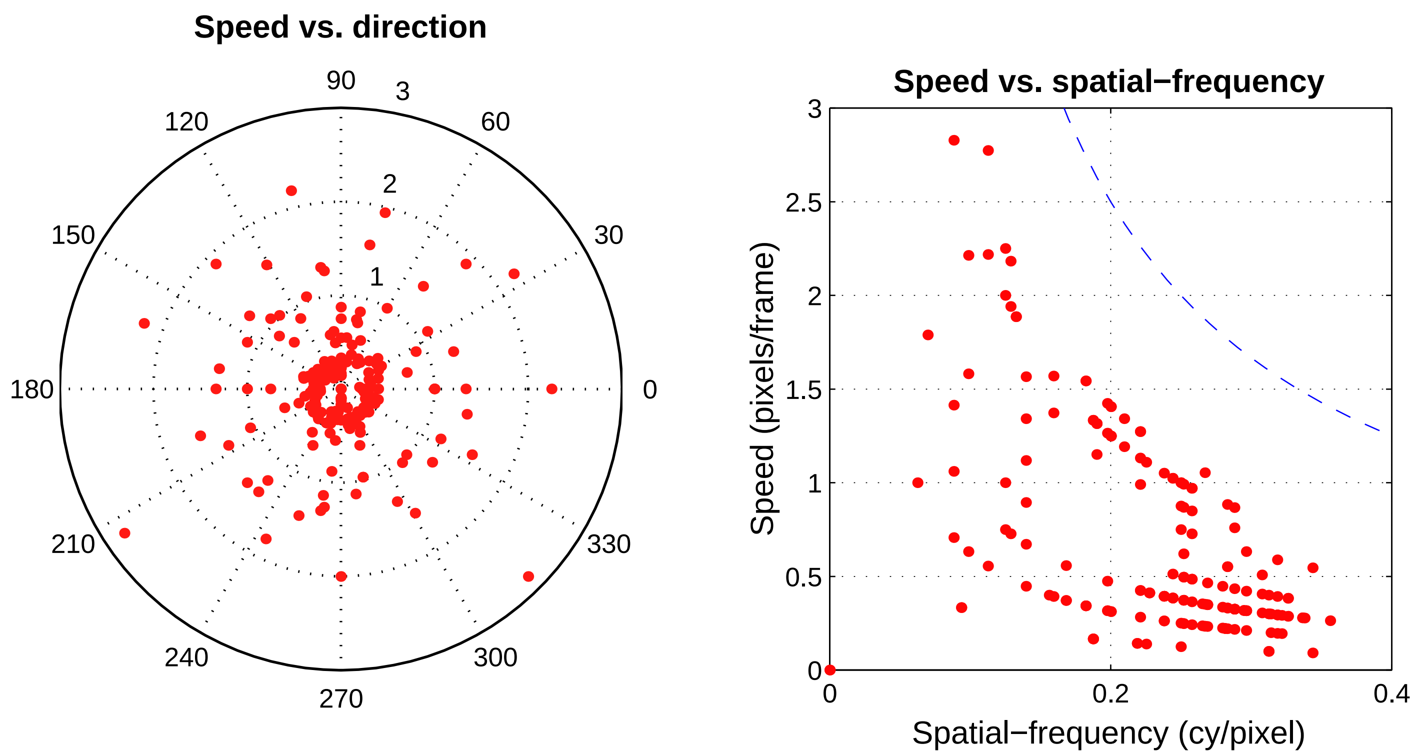
Goal: Find a set of space-time basis functions $\{\phi_i\}$ for representing natural images such that the *time-varying* coefficients $a_i(t)$ are as sparse and statistically independent as possible *over both space and time*.

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

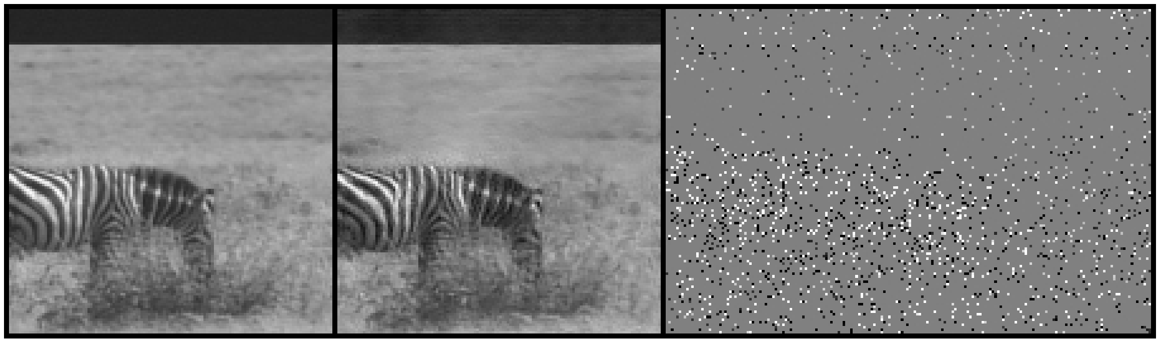
Training set: nature documentary

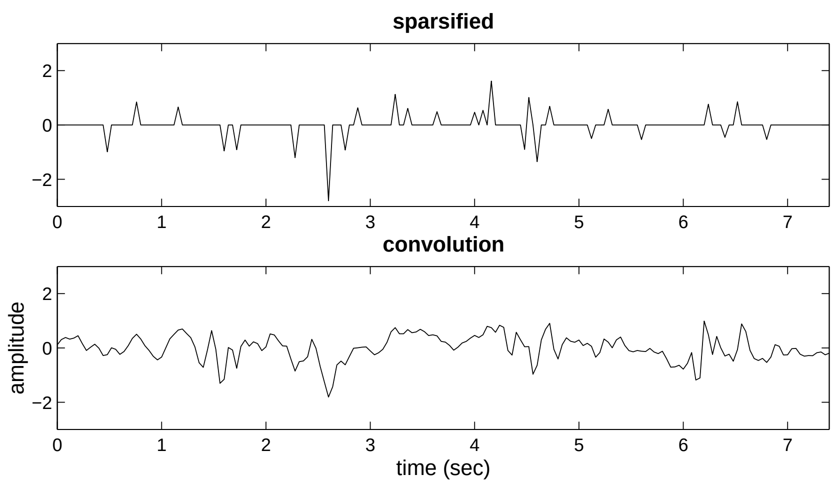


Basis function properties



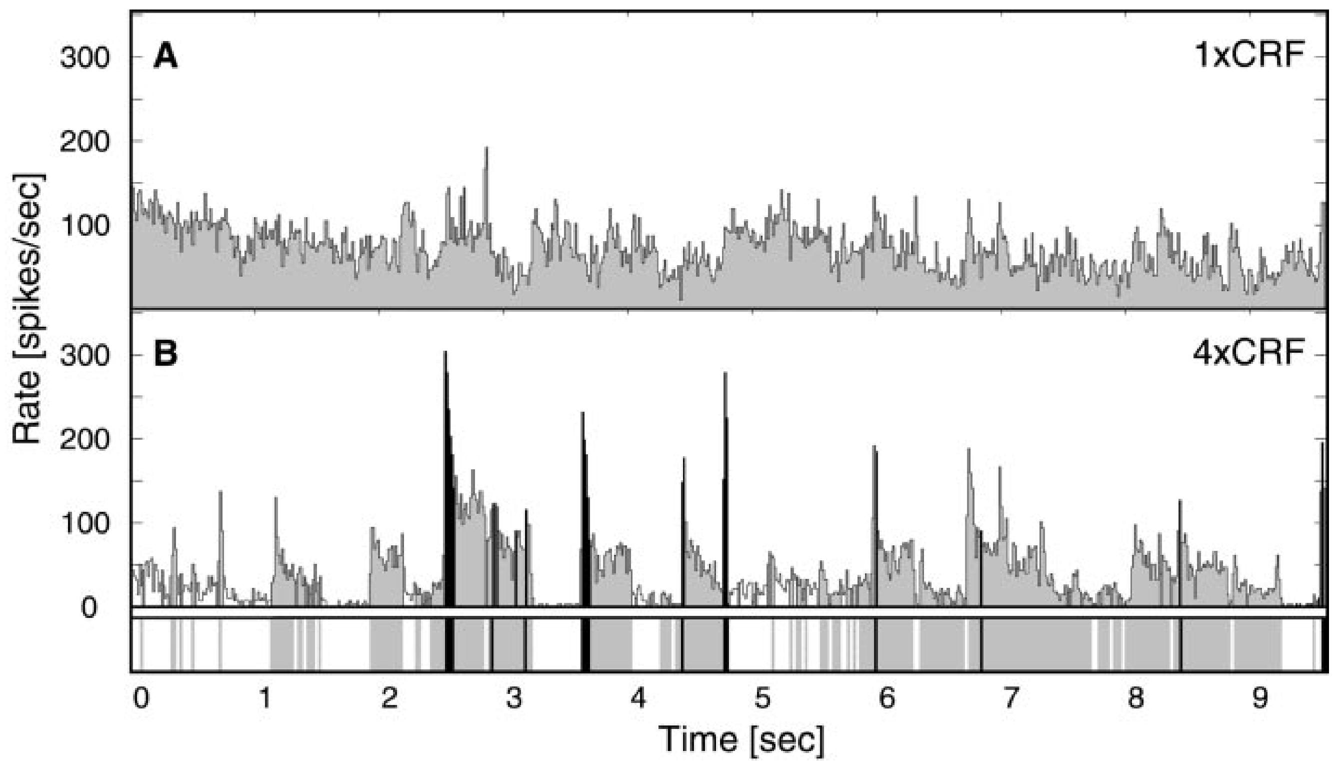
Spike encoding and reconstruction





Context in natural scenes sparsifies responses

Vinje & Gallant (2000, 2002)



Review article

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

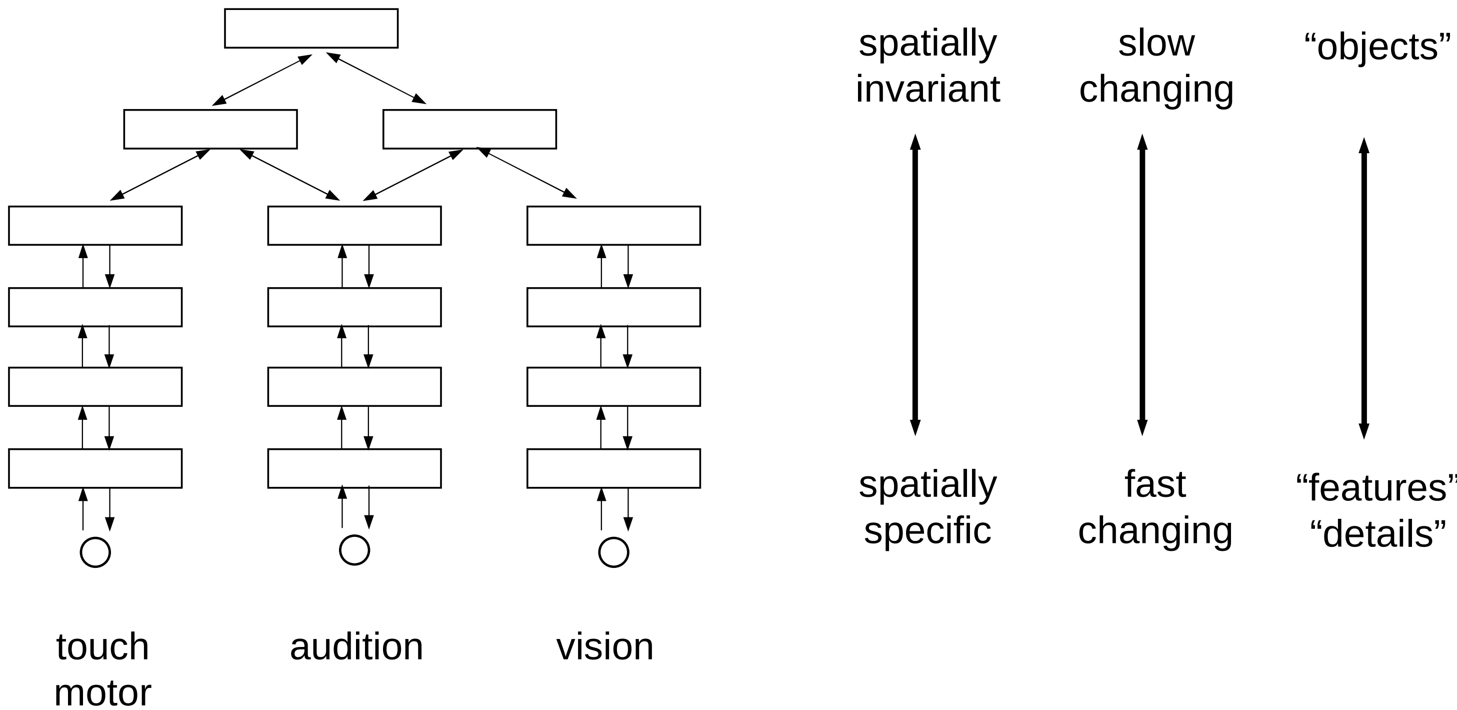
http://redwood.ucdavis.edu/bruno

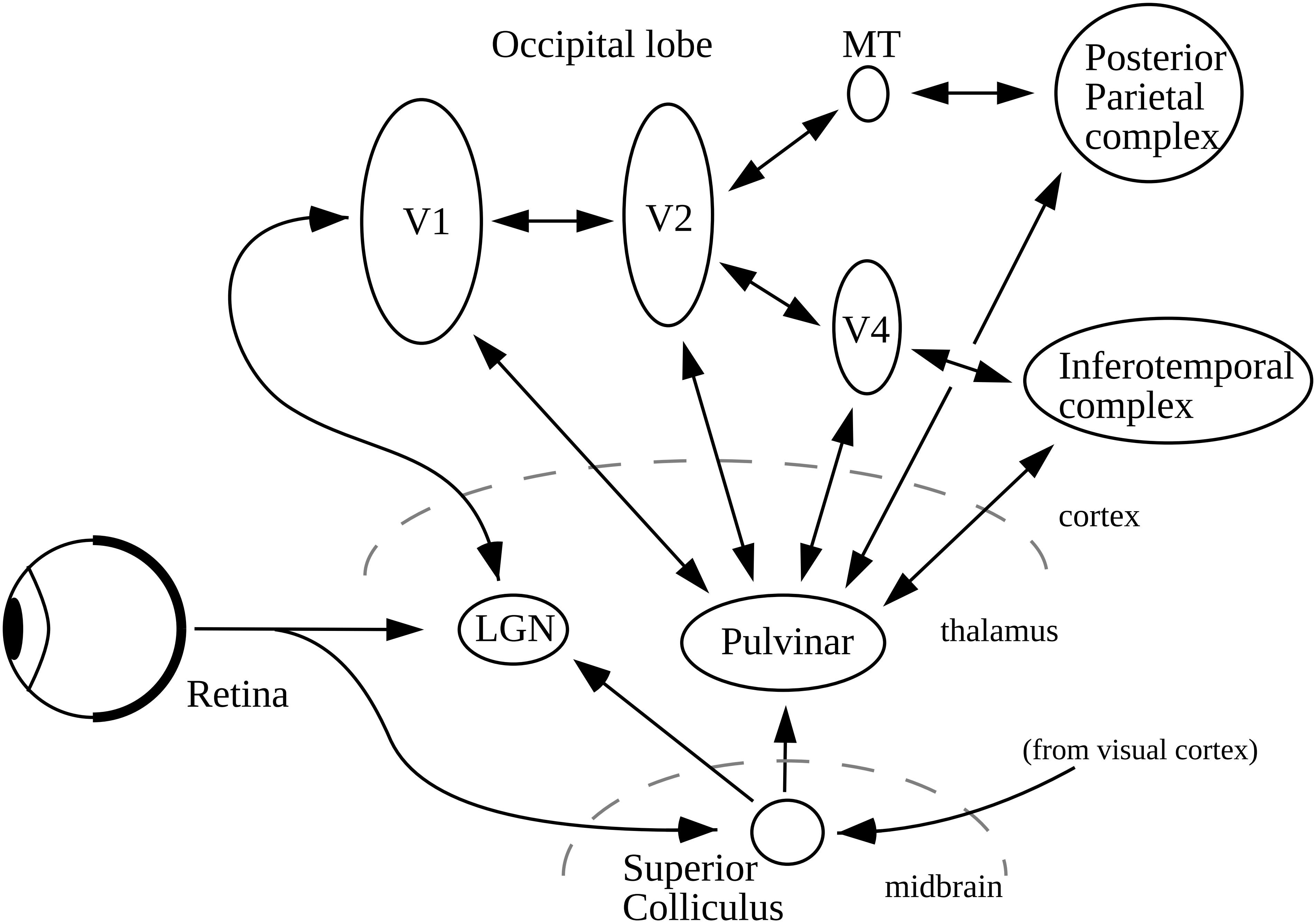
Problems with the current model

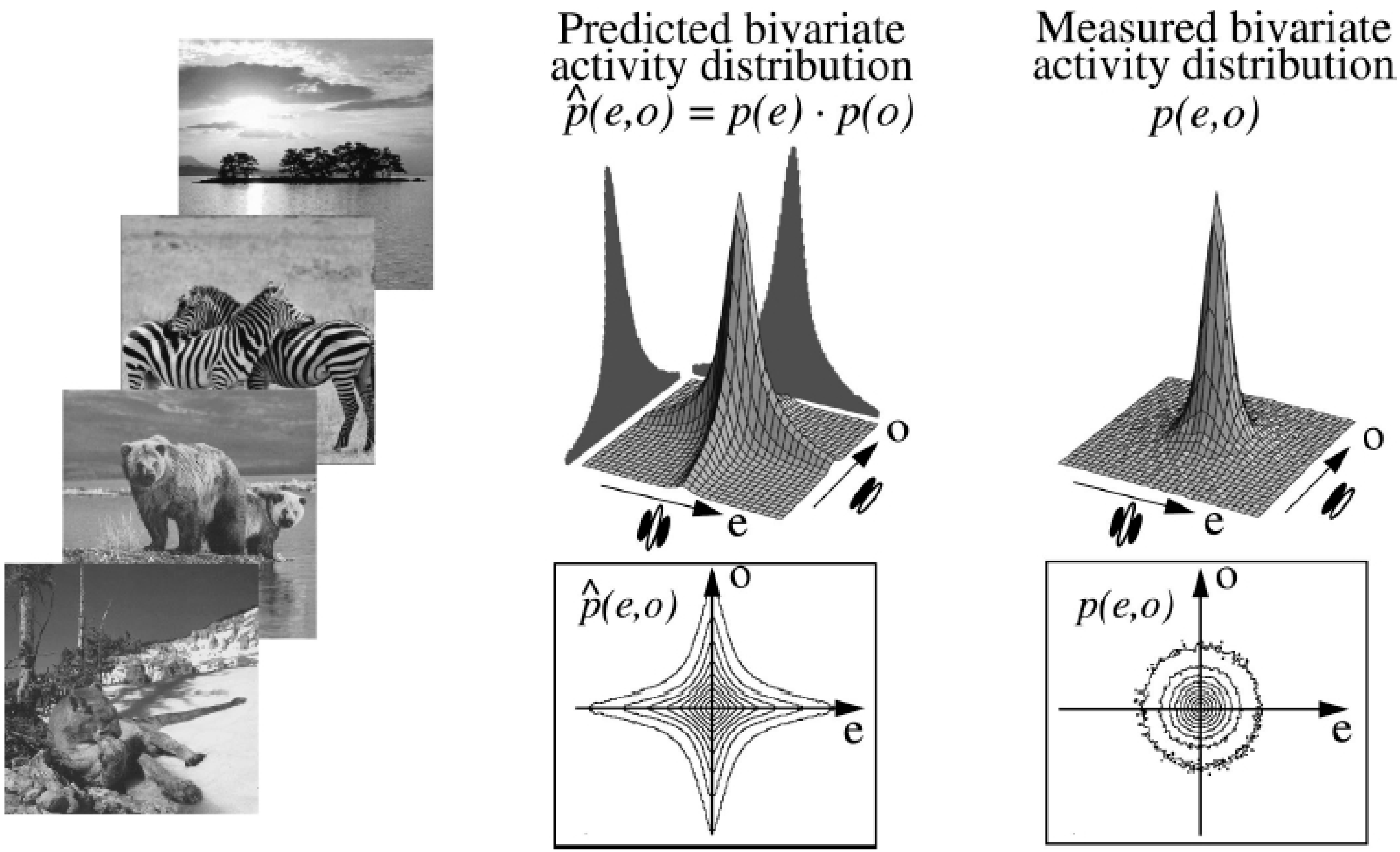
- Sparsification: small changes in the image could lead to drastic changes in the output representation.
- Factorial prior: coefficients exhibit strong dependencies, so the factorial prior is wrong.
- Linear model: how to extend to a hierarchical model?

Hierarchical representation

Hawkins (2004) - "On Intelligence"




Occipital lobe
MT
Posterior
Parietal
complex
V1
V2
V4
Inferotemporal
complex
cortex
LGN
Pulvinar
thalamus
Retina
(from visual cortex)
Superior
Colliculus
midbrain

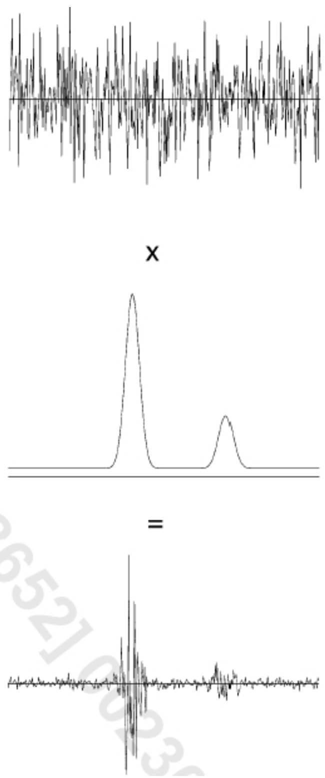

Predicted bivariate activity distribution
$\hat{p}(e,o) = p(e) \cdot p(o)$
Measured bivariate activity distribution
$p(e,o)$
o
e
o
e
$\hat{p}(e,o)$
o
e
$p(e,o)$
o
e

Bilinear model

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

Sparse bubbles

Hyvarinen et al. (2003) JOSA *20*



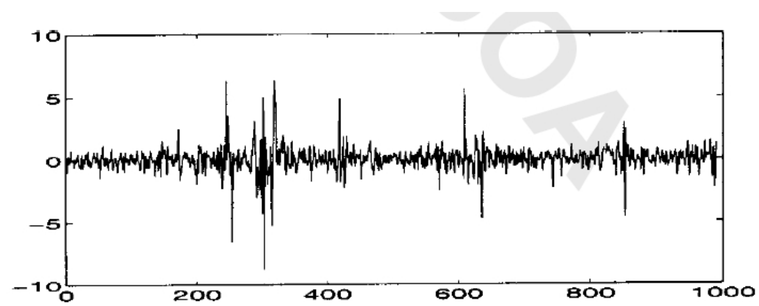
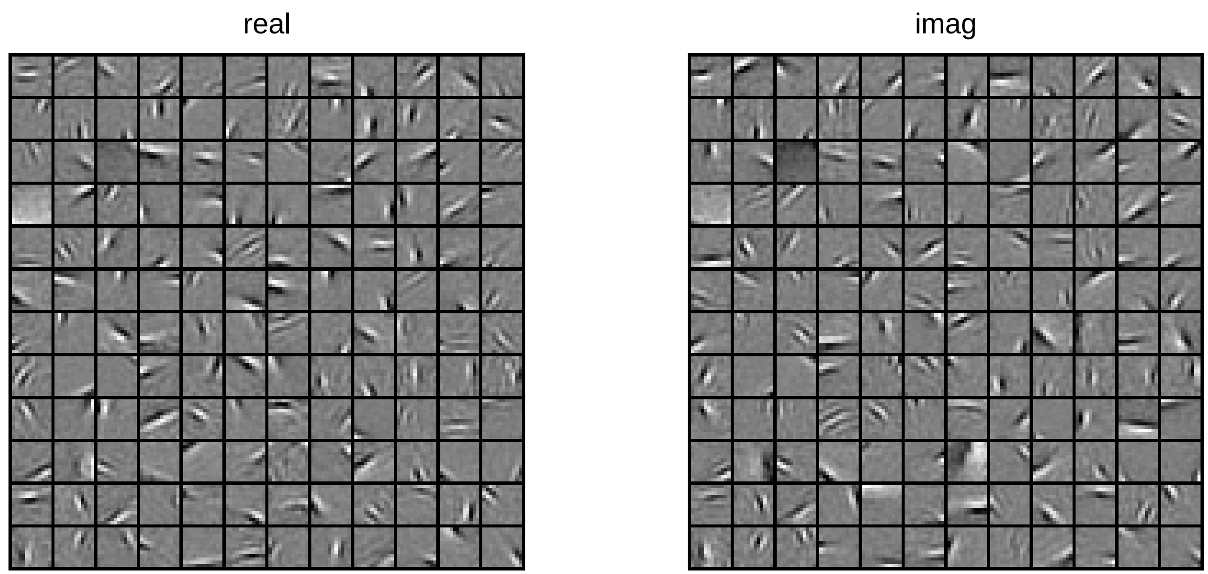


Image model with 'shiftable' basis functions

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

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

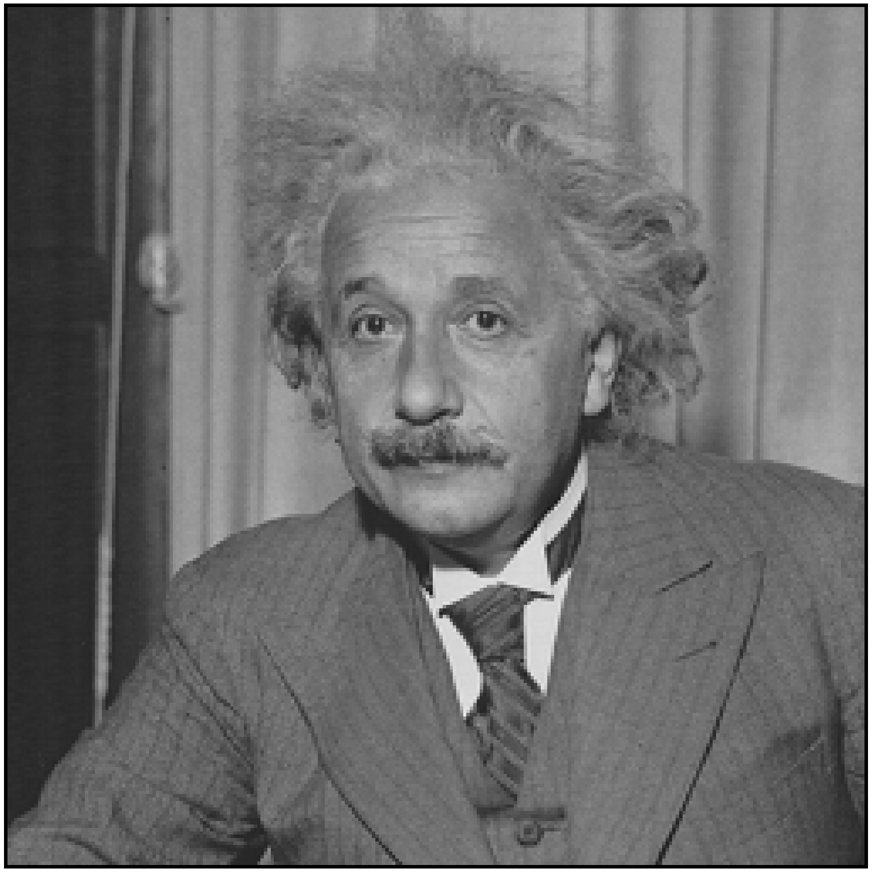
Learned complex basis functions (144, 12 $\times$ 12 patches)



animate!

Local phase is important

Original image

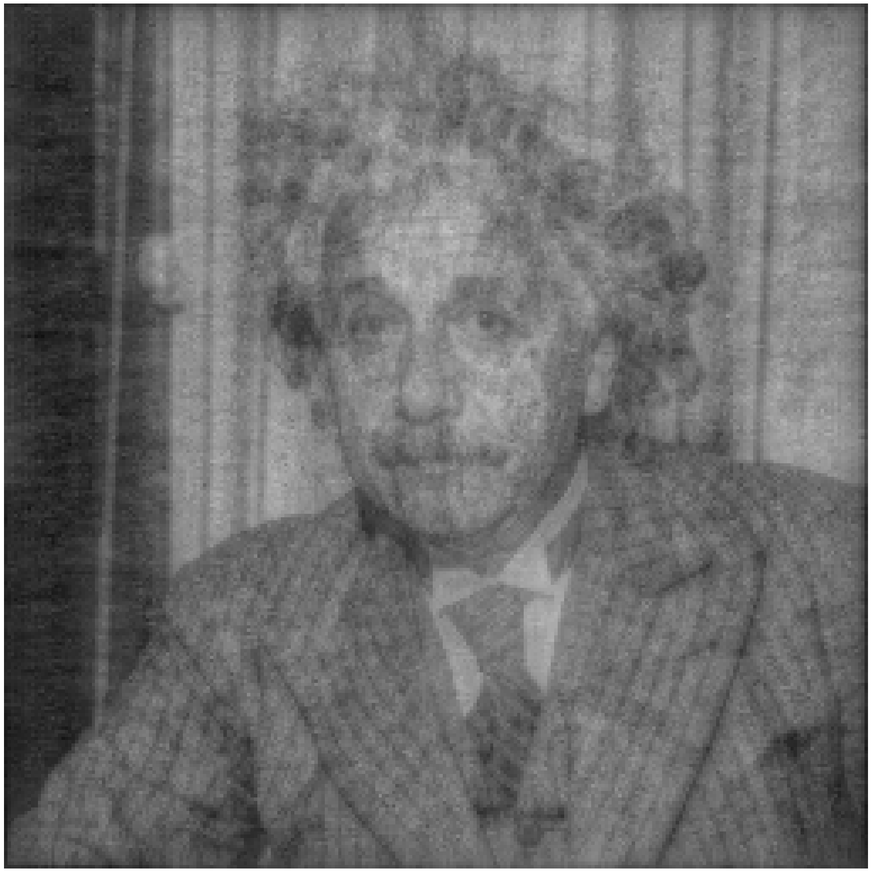


Local phase is important

Magnitudes only

Local phase is important

Phases only



Conclusions

- V1 neurons represent time-varying natural images in terms of sparse events.

- Joint dependencies among coefficients may be modeled with shiftable basis functions $\longrightarrow$ neurons carry both amplitude and phase?

Further information and details

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