Theory of efficient coding and inferential computation

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

REDWOOD CENTER
for Theoretical Neuroscience

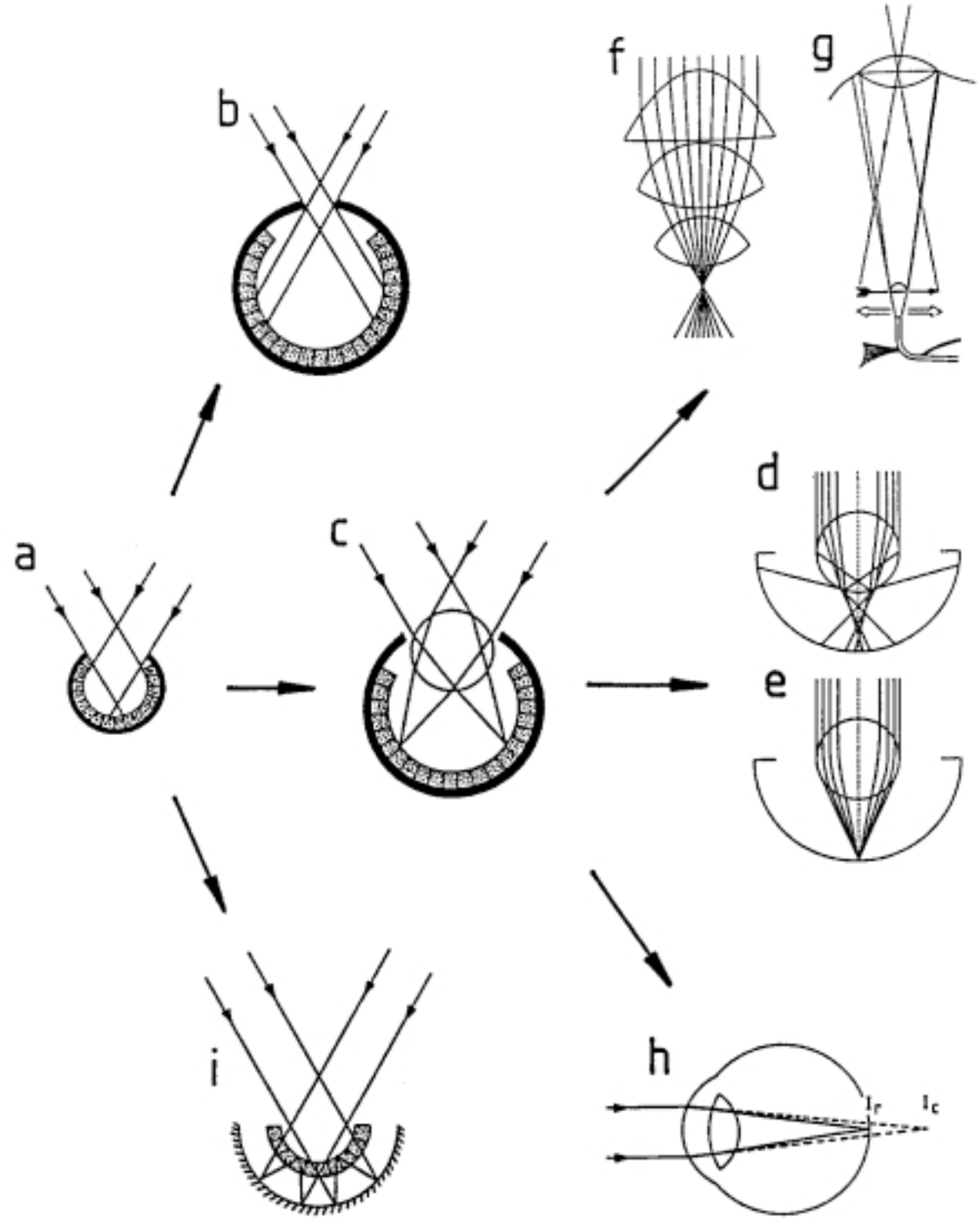
Redwood Center for Theoretical Neuroscience - April 2016

Redwood Center for Theoretical Neuroscience - April 2008

Principles of optics govern the design of eyes

THE EVOLUTION OF EYES

Michael F. Land *Russell D. Fernald*



What principles govern the design of nervous systems?

- Efficient Coding (Barlow)
- Inferential Computation (Helmholtz)

The efficient coding hypothesis
(Barlow 1961; Attneave 1954)

Nervous systems should exploit the statistical dependencies contained in sensory signals

Redundancy Reduction as a Strategy for Unsupervised Learning

A. Norman Redlich

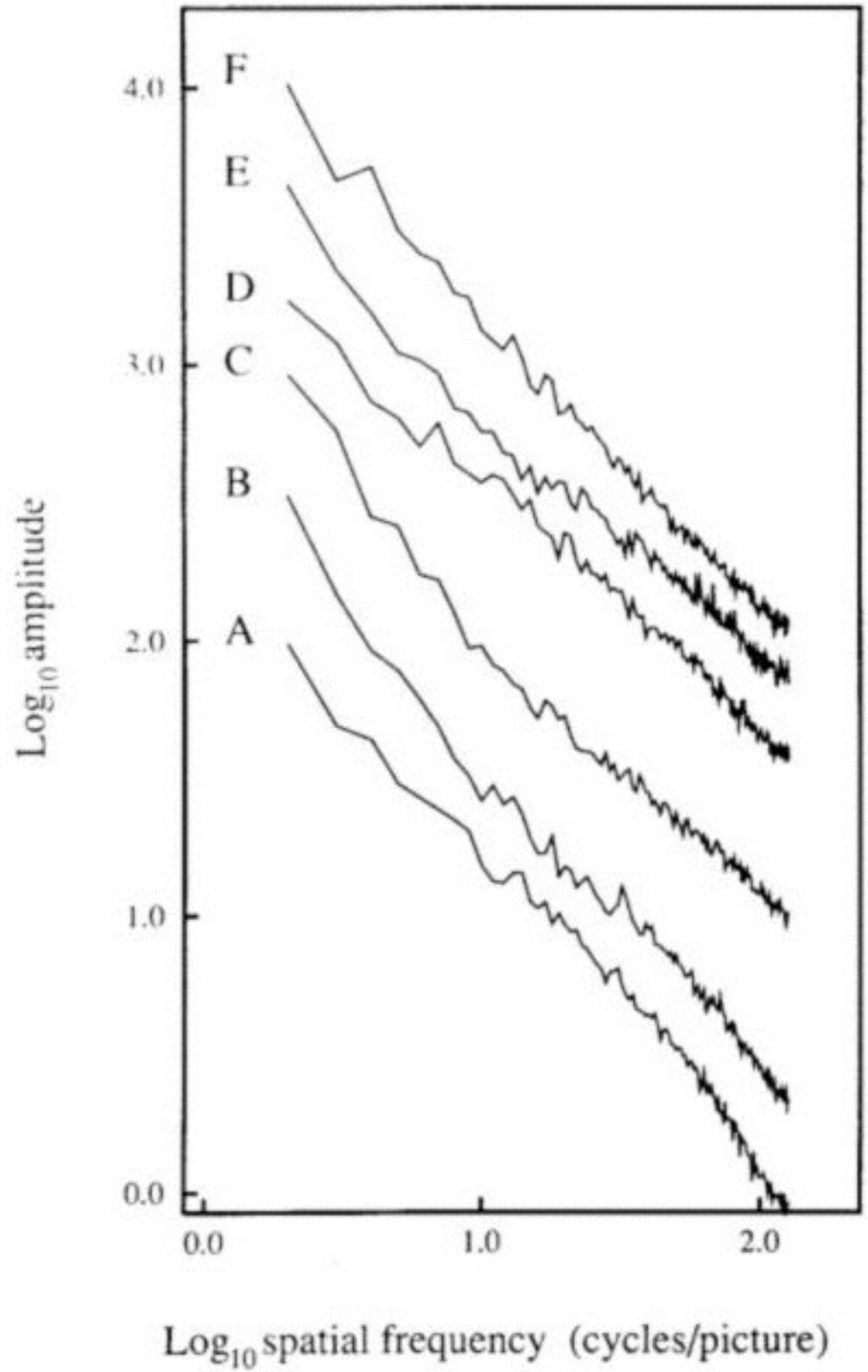
alicewasbeginningtogetverytiredofsittingbyhersisteron
thebankandofhavingnothingtodoonceortwiceshehadpeep
edintothebookhersisterwasreadingbutithadnopicturesor
conversationsinitandwhatistheuseofabookthoughtalicew
ithoutpicturesorconversationssoshewasconsideringinhe
rownmindaswellasshecouldforthehotdaymadeherfeelvery
sleepyandstupidwhetherthepleasureofmakingadaisychai
nwouldbeworththetroubleofgettingupandpickingthedaisi
eswhensuddenlyawhiterabbitwithpinkeyesranclosebyher
therewasnothingsoveryremarkableinthatnordidalicethi
nkitsoverymuchoutofthewaytohearthetabbitsaytoitselfo
hdearohdear

(Neural Computation, 1993)

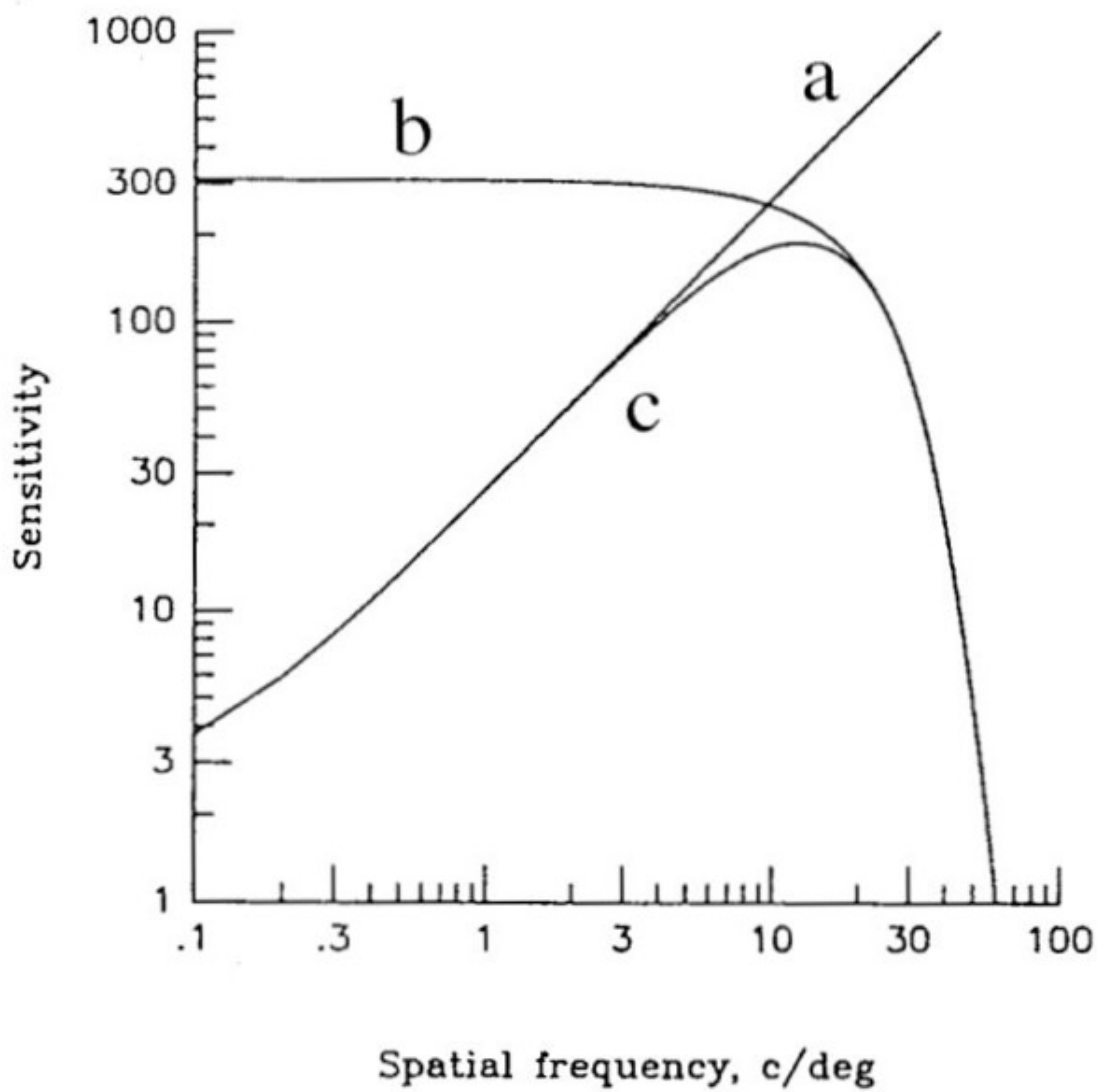
alice was b e g in n ing t o g e t very t i r e d o f s i t t ing b y her s i s t e r o n the b a n k and o f h a v ing n o th ing t o d o o n c e o r t w i c e she h a d p e e p e d in t o the b o o k her s i s t e r was r e a d ing but i th a d n o p i c t u r e s o r c o n ver s a t i o n s in i t and w h a t i s the u s e o f a b o o k th ough t alice with ou t p i c t u r e s o r c o n ver s a t i o n s s o she was c o n s i d e r ing in her o w n m in d a s w e l l a s she c ould for the h o t d a y m a d e her f e e l very s l e e p y and s t u p i d w he the r the p l e a s u r e o f m a k ing a d a i s y c h a in w ould b e w o r th the t r ou b l e o f g e t t ing u p and p i c k ing the d a i s i e s w he n s u d d e n l y a w h i t e r a b b i t with p in k e y e s r a n c l o s e b y her the r e was n o th ing s o very r e m a r k a b l e in that n o r d i dalic e th in k i t s o very m u c h ou t o f the w a y t o he a r the r a b b i t s a y t o i t s e l f o h d e a r o h d e a r

alice was beginning toget verytiredof sitting by hersister ontheban k and of having nothingtodo onceortwice shehad peeped intothe book hersister was read ing but ithad no pictures or conversation s in it and what is theuseof a book thoughtalice without pictures or conversation s so shewas consider ing in her ow n mind aswell asshecould for the hot day made her feel very sleepy and stupid whether the plea sure of making ad a is y ch a in wouldbe wor th the trouble of gettingupand p i c king the d a is ies when suddenly a whiterabbit with p in k eyes r an close by her therewas nothing so very remark able in that n or did alice think it so very much outoftheway to hear therabbit saytoitself ohdear ohdear

Power spectrum of natural images (Field 1987)

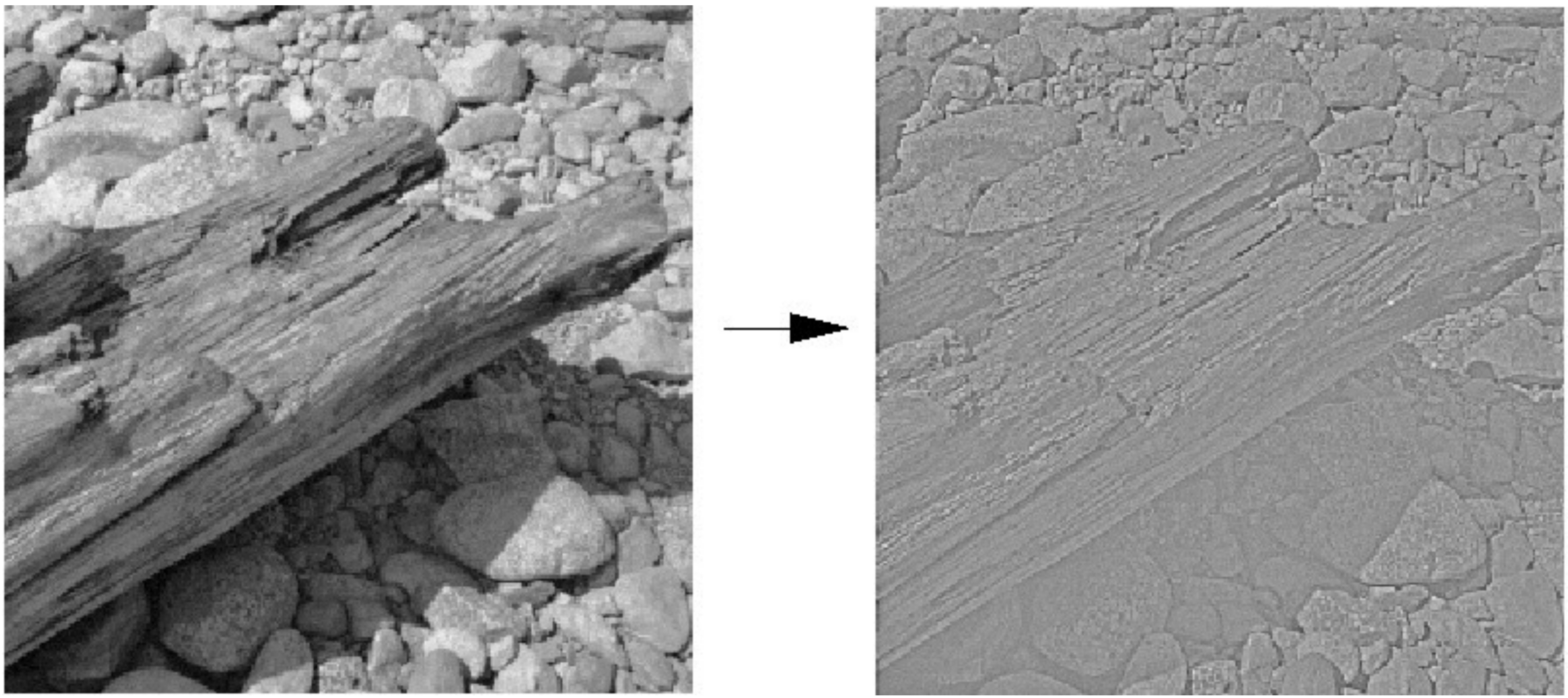


'Whitening' (Atick & Redlich, 1990)

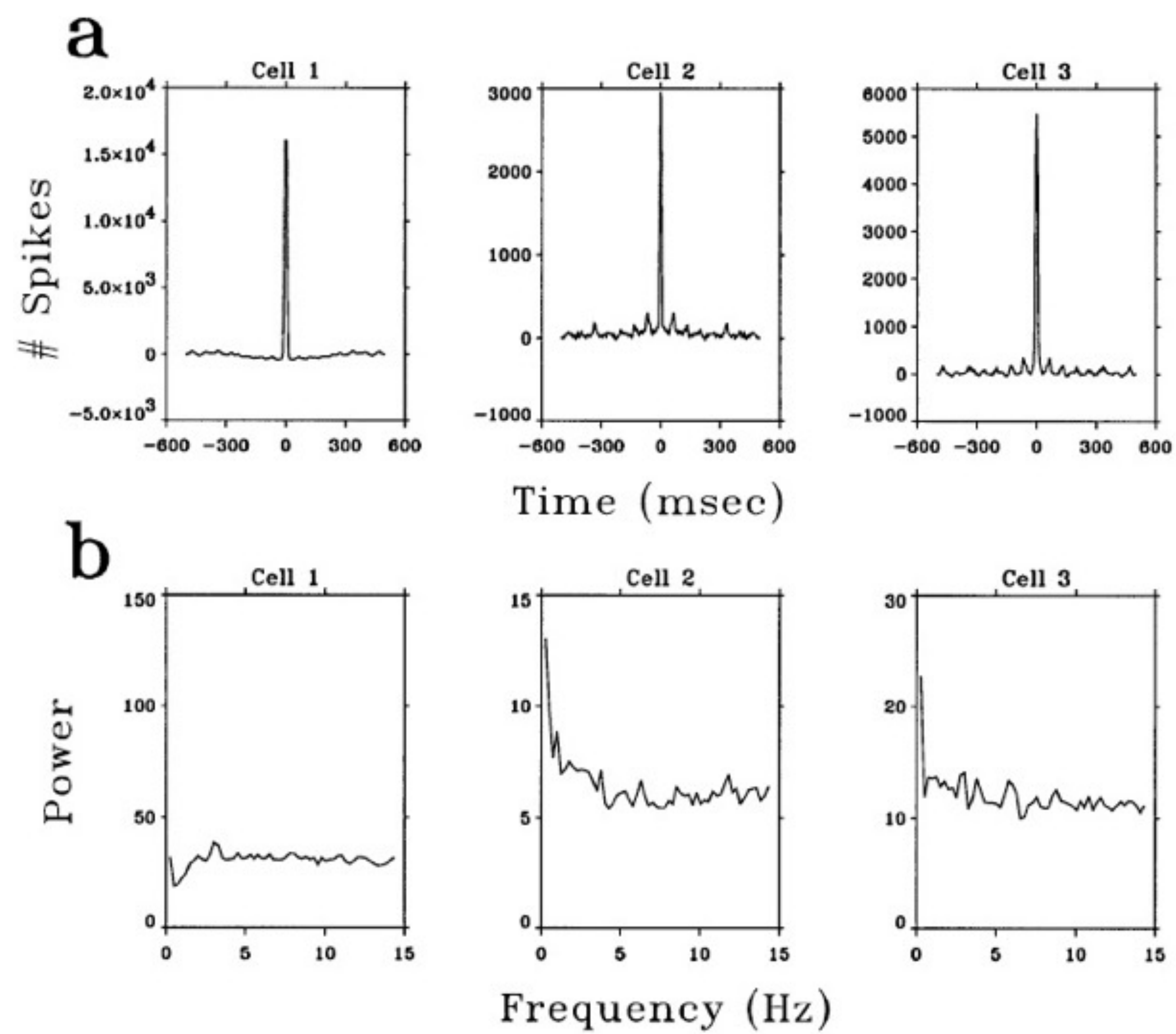


Whitening

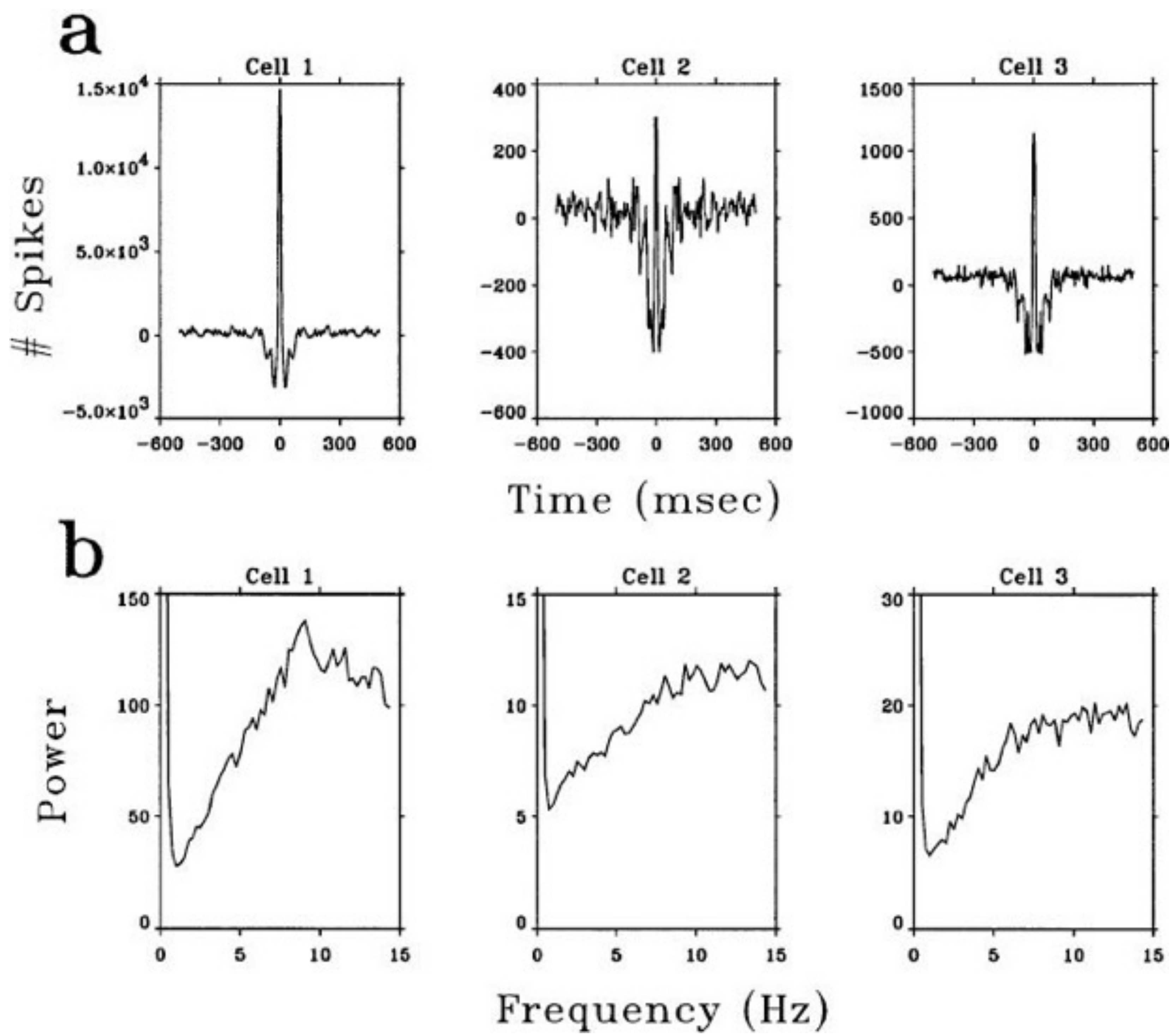
before after



Evidence for whitening in the LGN
(Dan et al., 1996)



Evidence for whitening in the LGN
(Dan et al., 1996)



Redundancy reduction revisited

Horace Barlow

Physiological Laboratory, Downing Site, Cambridge CB2 3EG, UK

E-mail: hbb10@cam.ac.uk

Received 31 November 2000

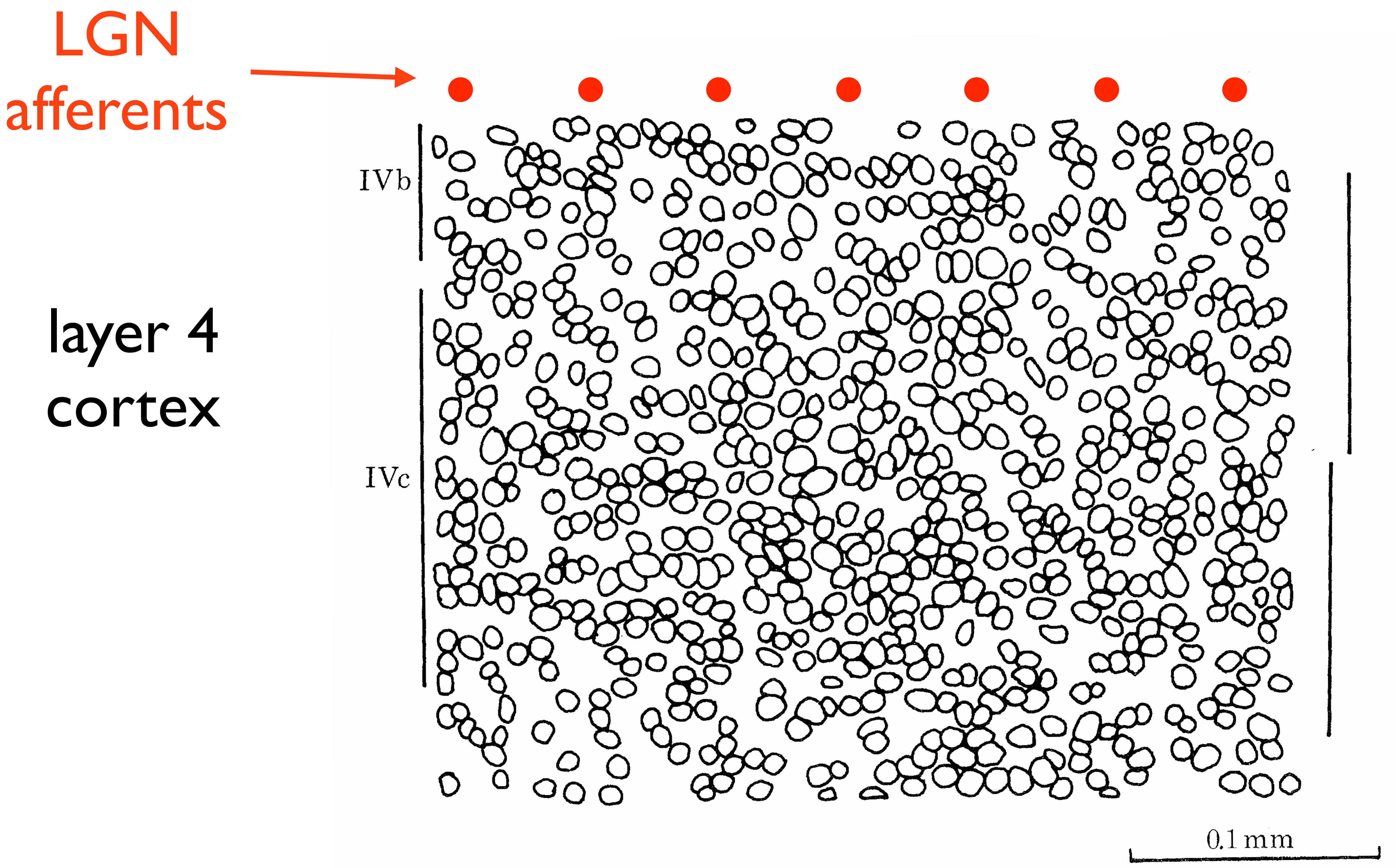
Abstract

Soon after Shannon defined the concept of redundancy it was suggested that it gave insight into mechanisms of sensory processing, perception, intelligence

This paper argues that the original hypothesis was wrong in over-emphasizing the role of compressive coding and economy in neuron numbers, but right in drawing attention to the importance of redundancy.

non-random probabilities and interdependences of objects and events signalled by sensory messages. These are particularly relevant for Bayesian calculations of the optimum course of action. Instead of thinking of neural representations as transformations of stimulus energies, we should regard them as approximate estimates of the probable truths of hypotheses about the current environment, for these are the quantities required by a probabilistic brain working on Bayesian principles.

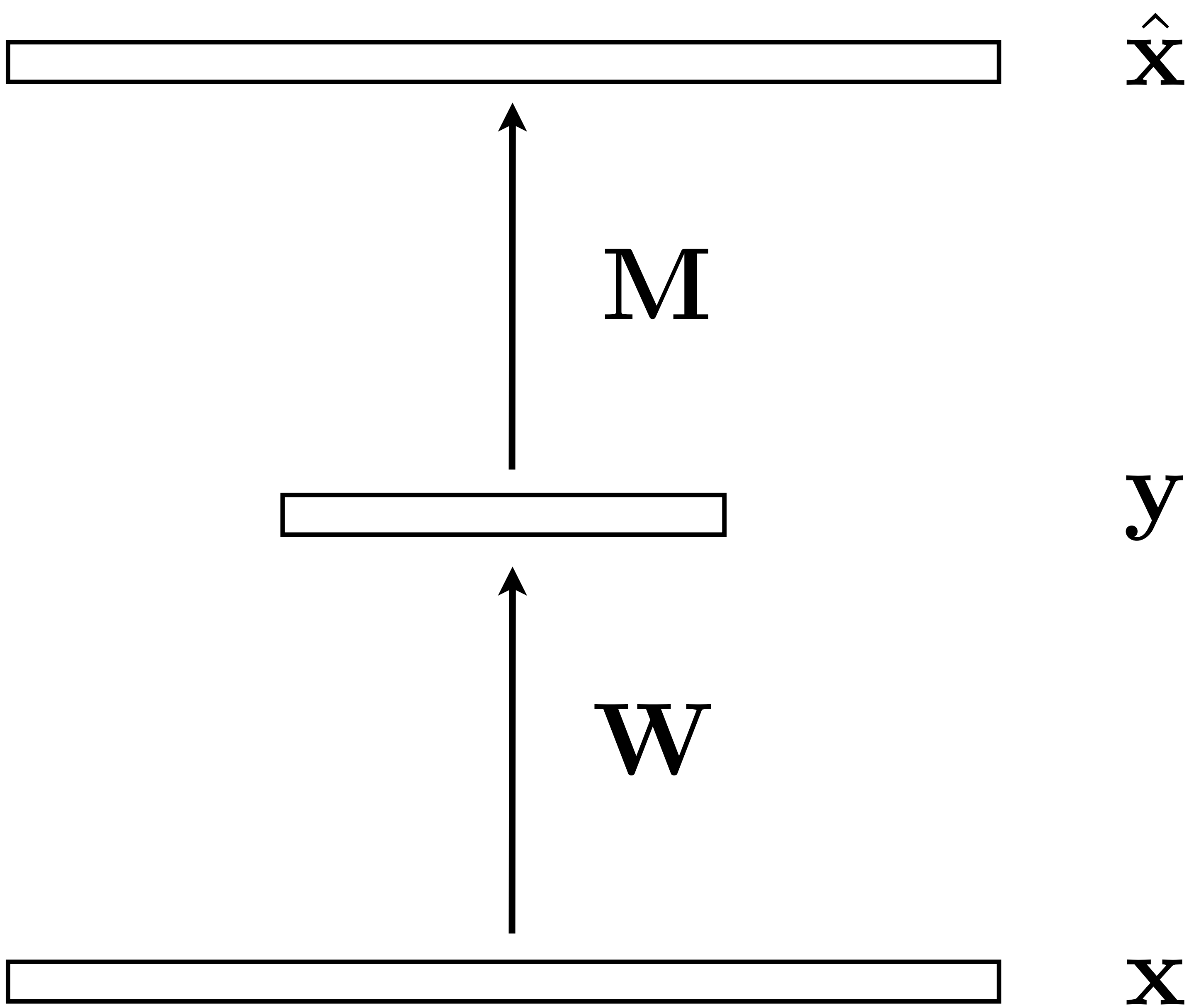
V1 is highly overcomplete



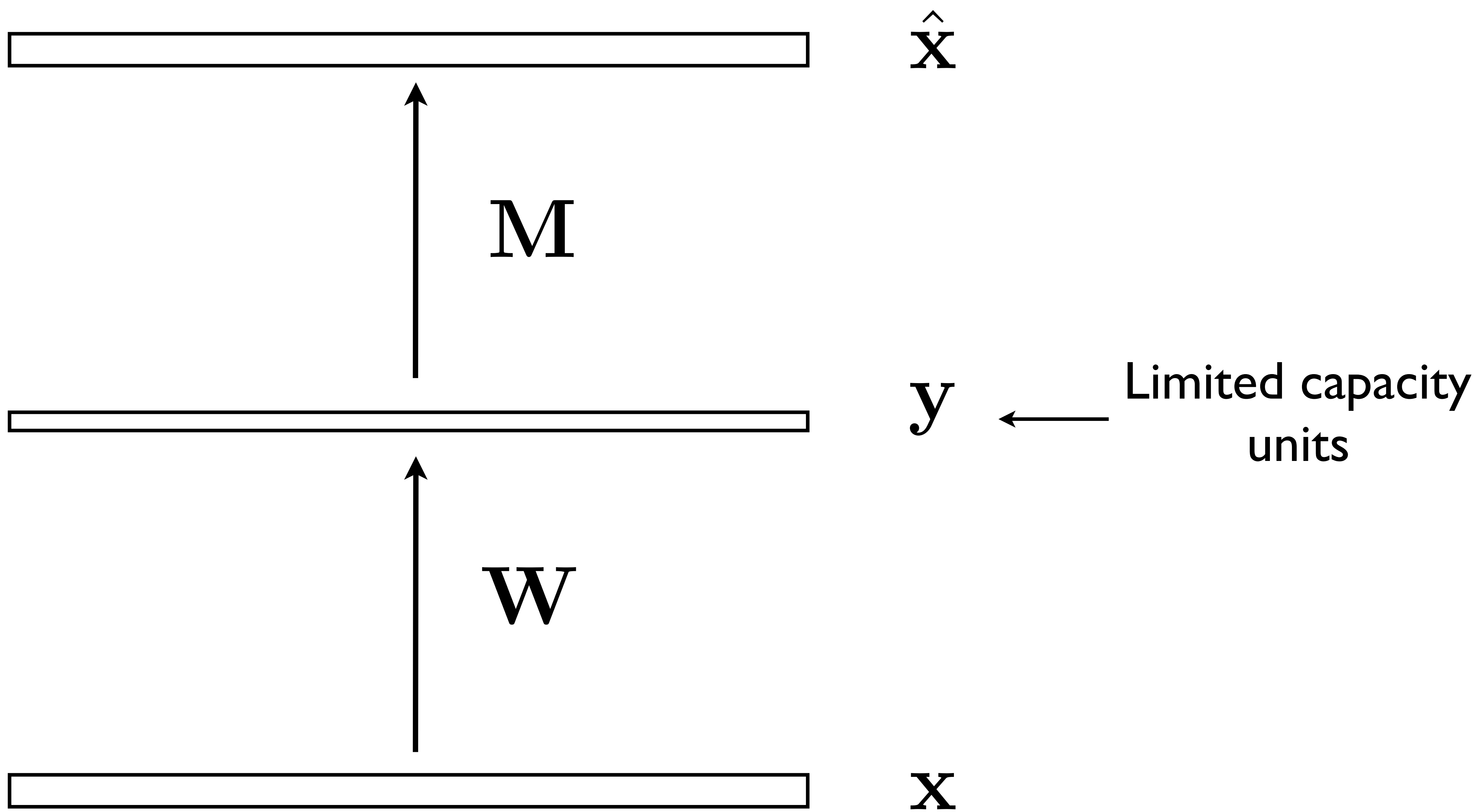
Barlow (1981)

Autoencoder networks

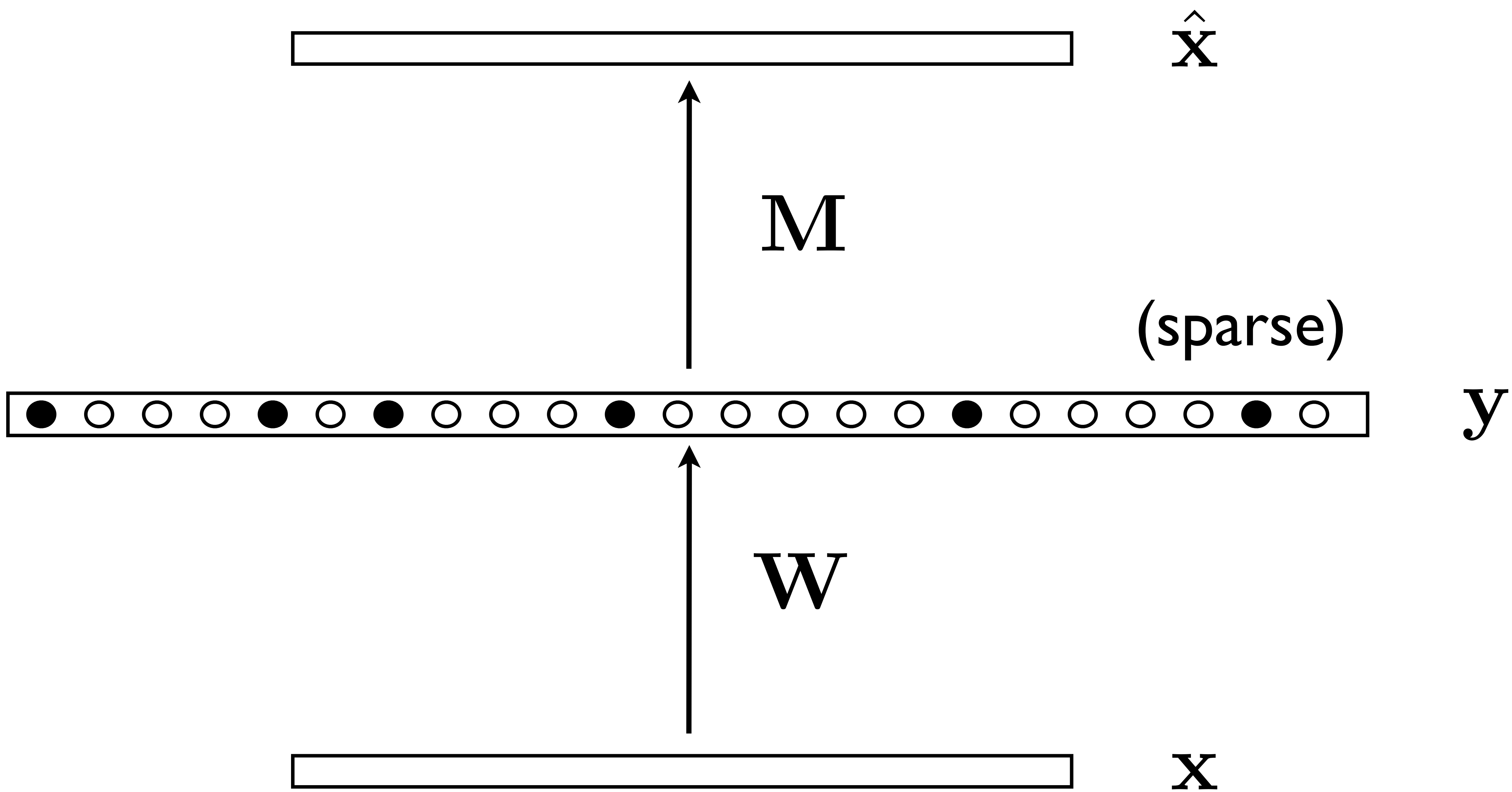
$$\min_{\mathbf{W},\mathbf{M}} |\mathbf{x} - \hat{\mathbf{x}}|^2$$



Bottleneck may also be in the form of limited capacity units. Optimal strategy in this case is to whiten.



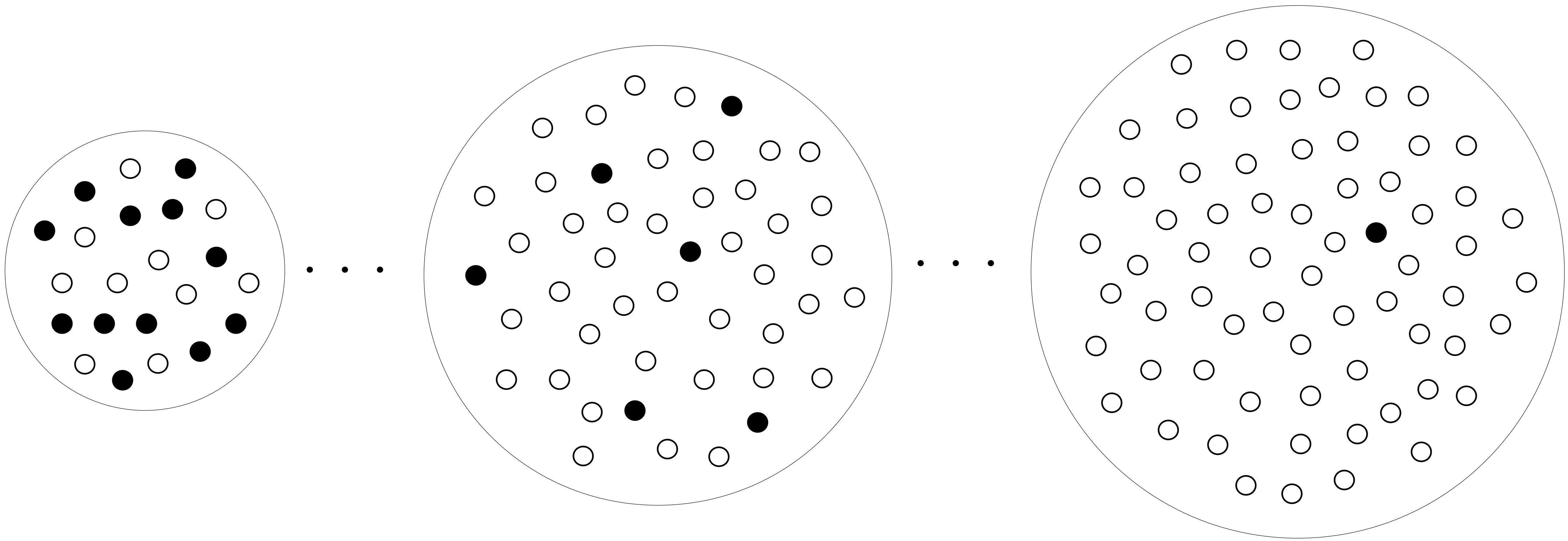
Sparse codes impose a different type of bottleneck by limiting the number of active units



Dense codes
(e.g., ascii)

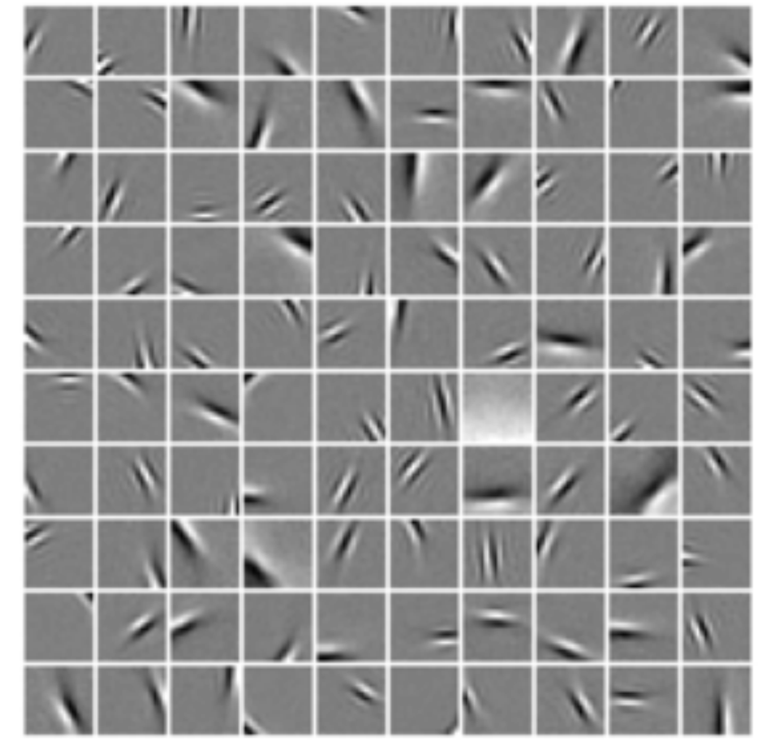
Sparse, distributed codes

Local codes
(e.g., grandmother cells)



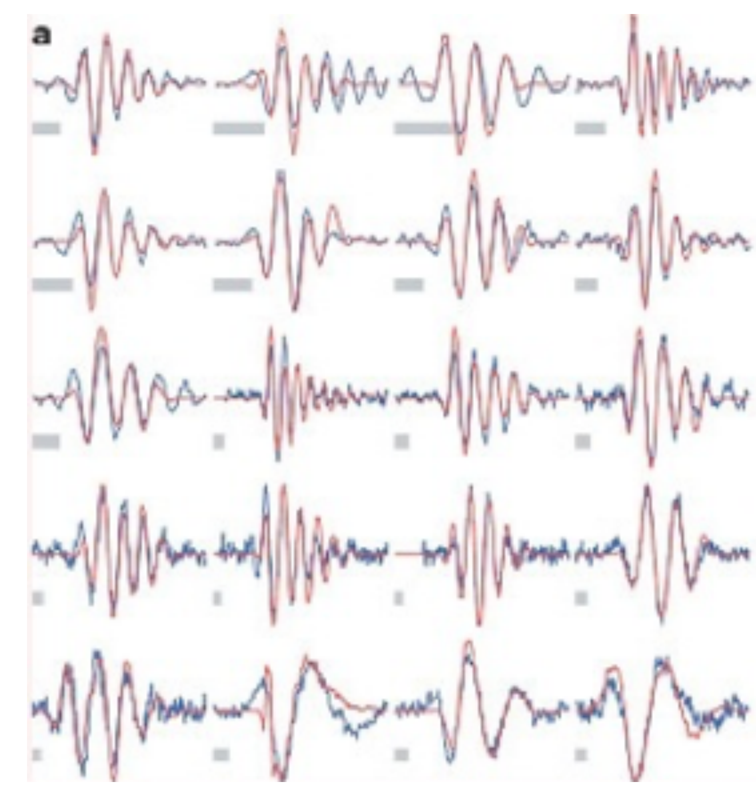
2^N . . . $\binom{N}{K}$ . . . N

natural images → sparse coding/ICA →



V1 receptive fields

natural sounds → sparse coding/ICA →

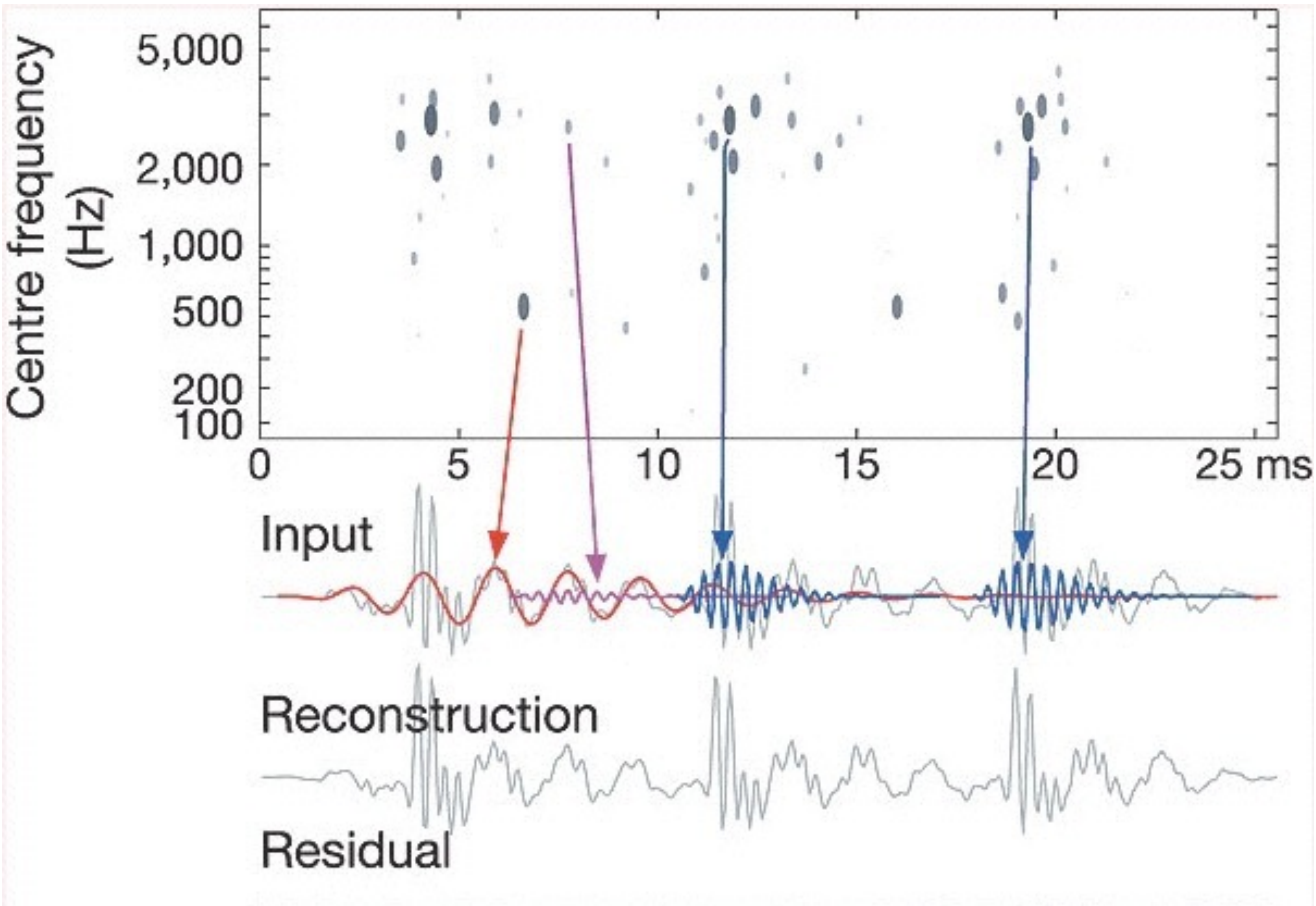


auditory nerve filters

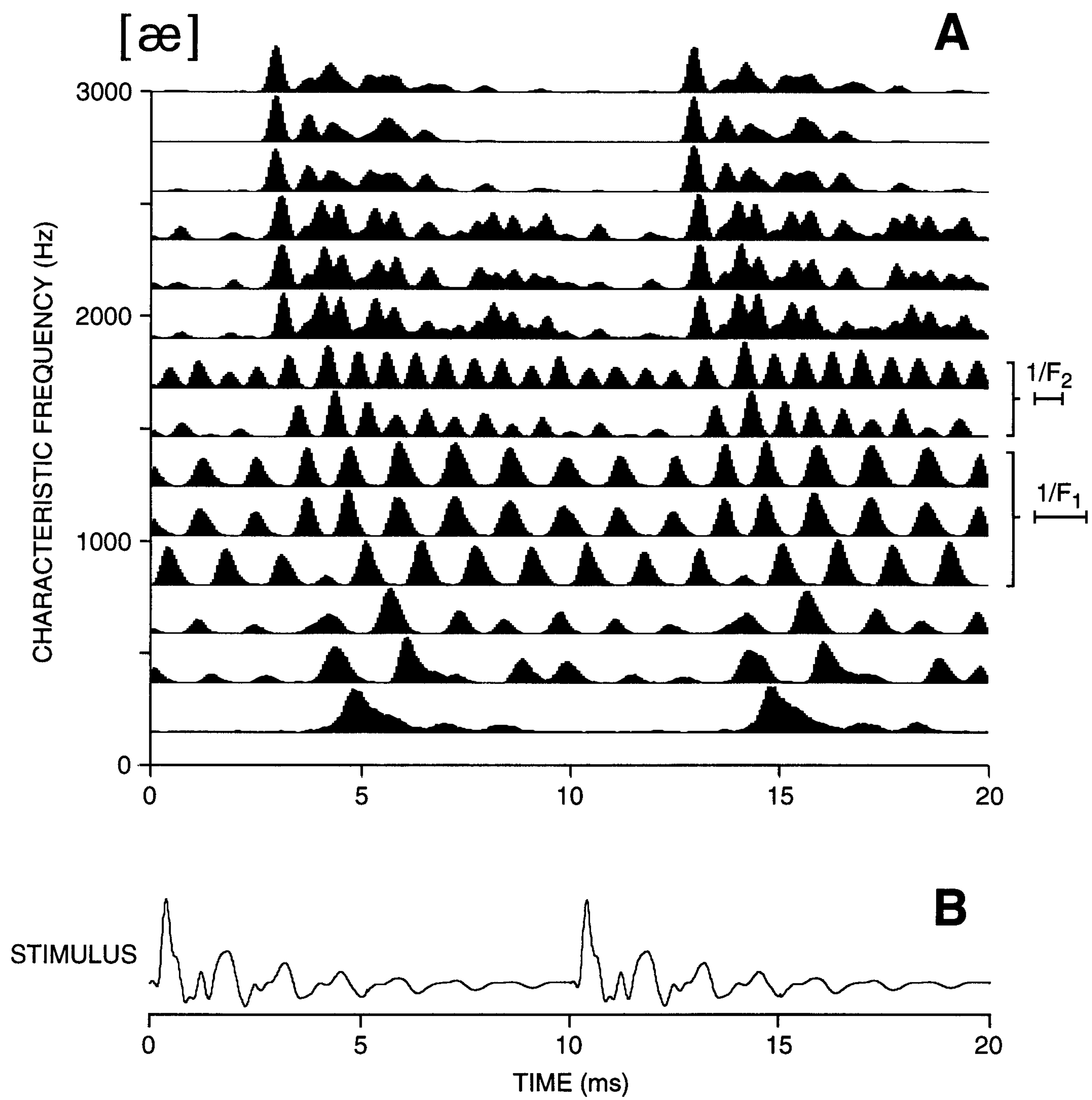
Sparse coding of natural sounds
(Smith & Lewicki 2006)

$a_i(t)$

$s(t)$

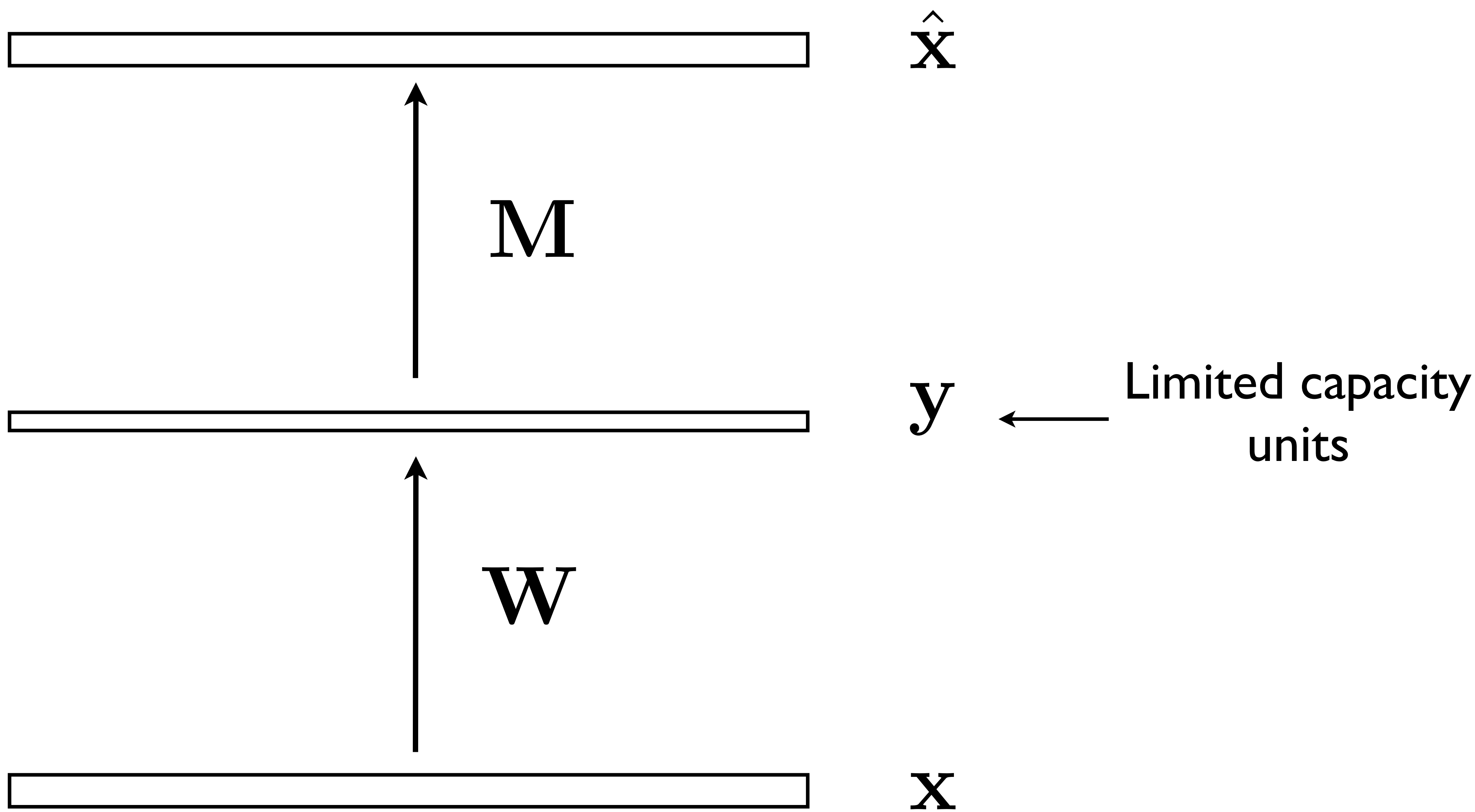


But this doesn't really look like sparse coding



Delgutte (1997)

Let's go back to that bottleneck idea…



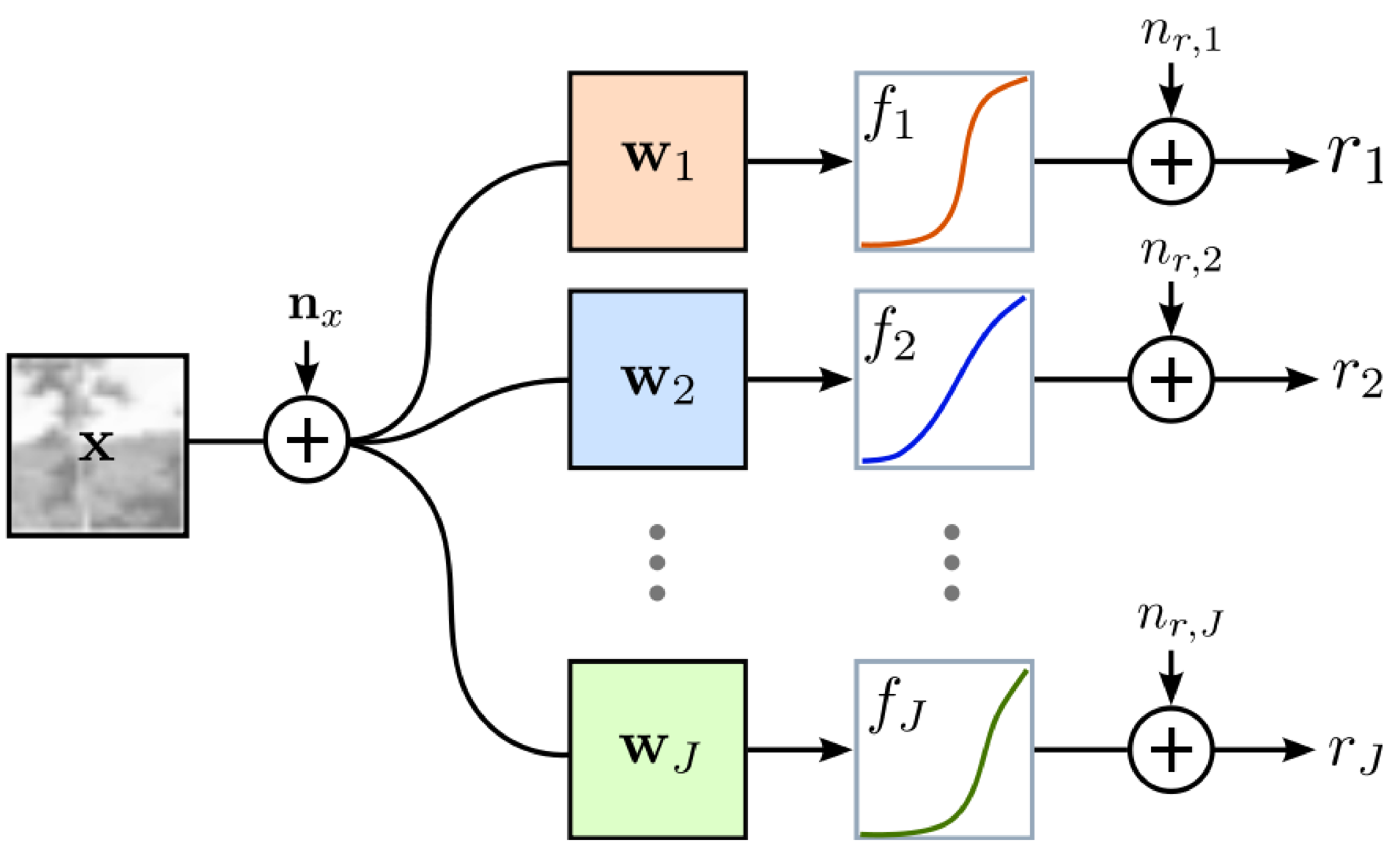
From: D.J.C. MacKay, *Information Theory, Inference, and Learning Algorithms*

• Exercise 4.1.[2, p.69] – *Please work on this problem before reading Chapter 4.*

You are given 12 balls, all equal in weight except for one that is either heavier or lighter. You are also given a two-pan balance to use. In each use of the balance you may put any number of the 12 balls on the left pan, and the same number on the right pan, and push a button to initiate the weighing; there are three possible outcomes: either the weights are equal, or the balls on the left are heavier, or the balls on the left are lighter. Your task is to design a strategy to determine which is the odd ball *and* whether it is heavier or lighter than the others *in as few uses of the balance as possible.*

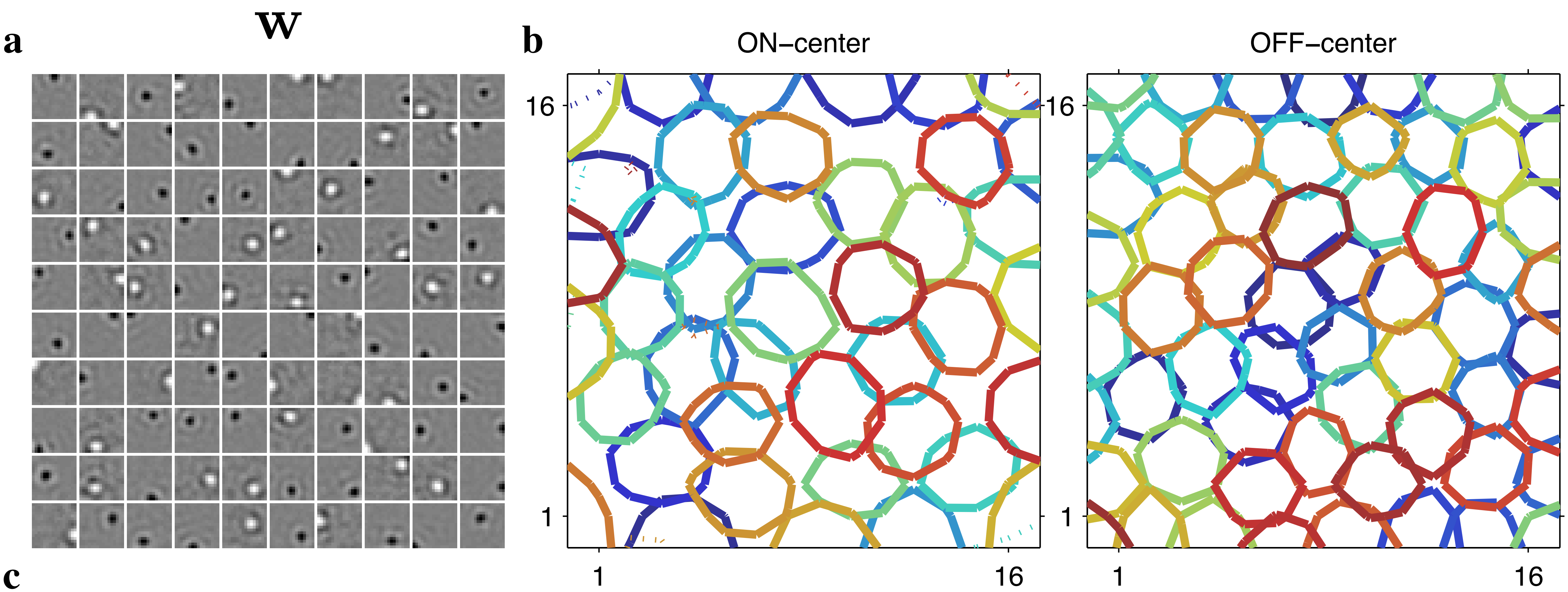
Efficient coding model of retina

(Karklin & Simoncelli 2012)

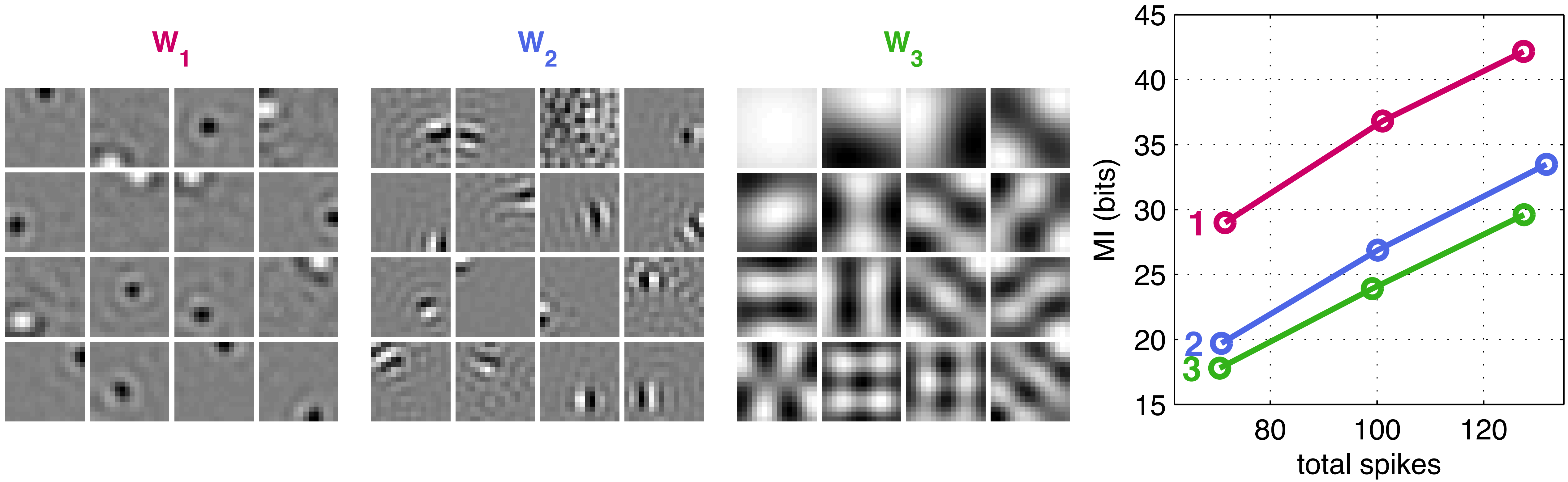


Objective function: $I(X;R) - \sum_j \lambda_j \langle r_j \rangle$

Efficient coding model of retina
(Karklin & Simoncelli 2012)



Efficient coding model of retina
(Karklin & Simoncelli 2012)



THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 56 | April 1977 | Number 4

Information in the Zero Crossings of Bandpass Signals

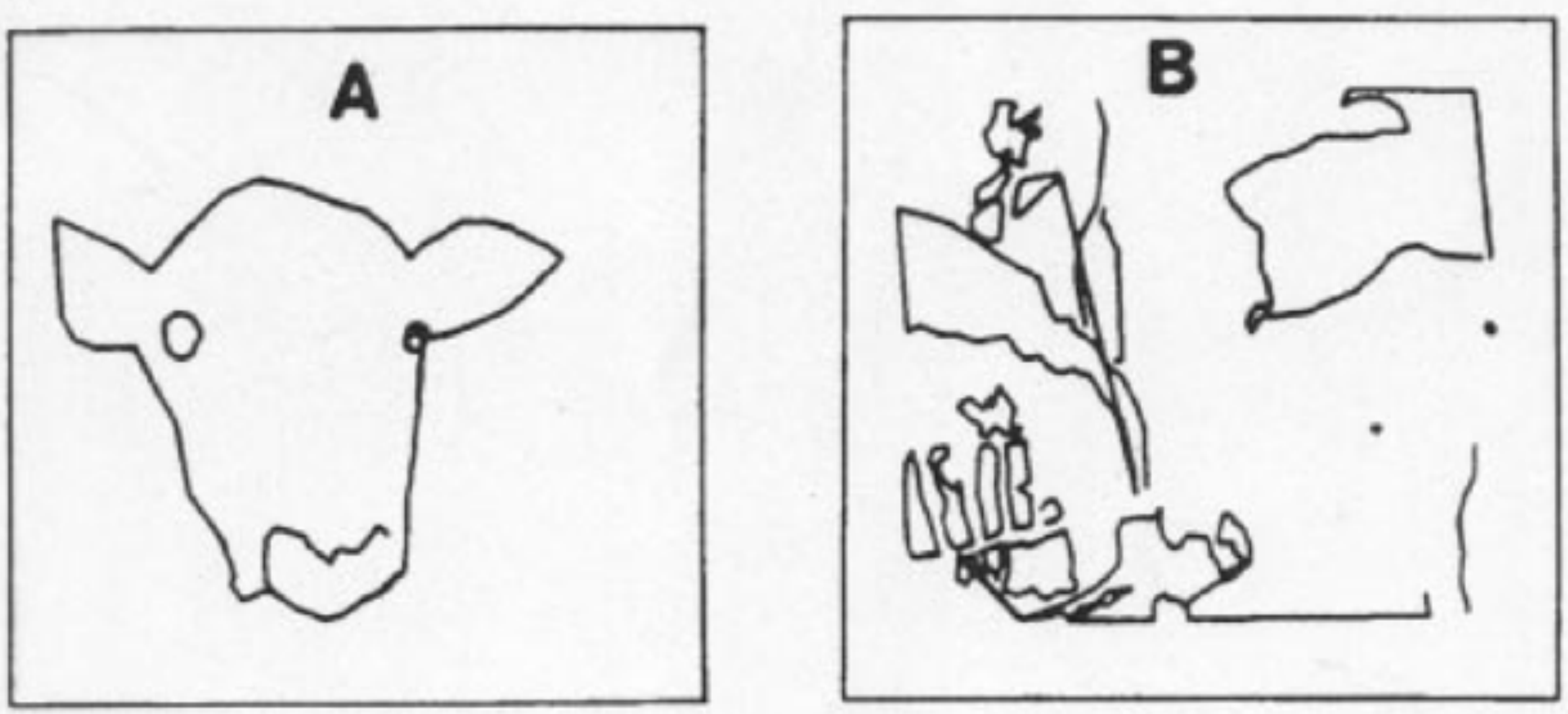
By B. F. LOGAN, JR.

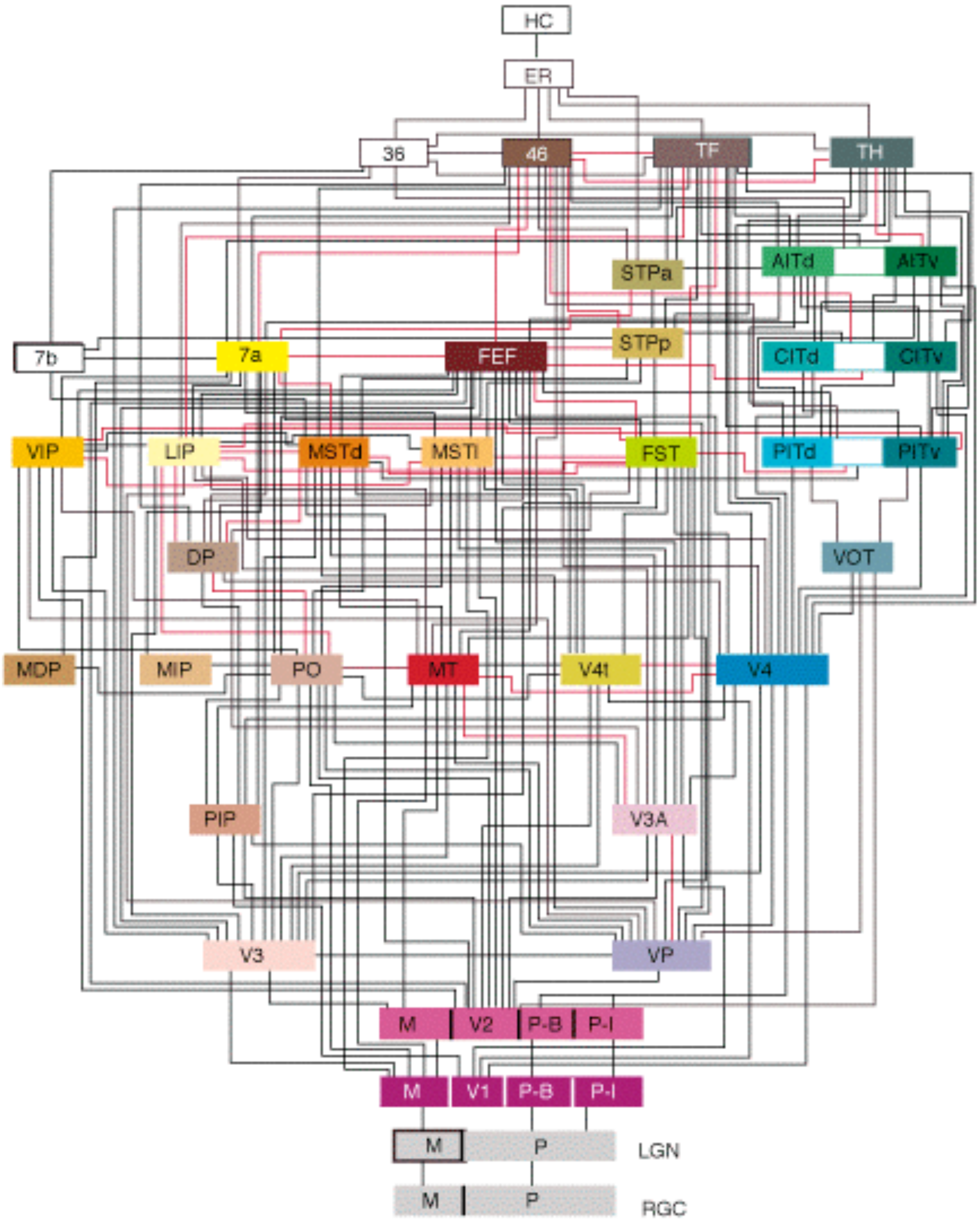
(Manuscript received October 4, 1976)

An interesting subclass of bandpass signals $\{h\}$ is described wherein the zero crossings of h determine h within a multiplicative constant. The members may have complex zeros, but it is necessary that h should have no zeros in common with its Hilbert transform $\hat{h}$ other than real simple zeros. It is then sufficient that the band be less than an octave in width. The subclass is shown to include full-carrier upper-sideband signals (of less than an octave bandwidth). Also it is shown that full-carrier lower-sideband signals have only real simple zeros (for any ratio of upper and lower frequencies) and, hence, are readily identified by their zero crossings. However, under the most general conditions for uniqueness, the problem of actually recovering h from its sign changes appears to be very difficult and impractical.

Perception as inference

b





place cells
grid cells

.

.

face cells

.

invariant repr.
complex motion

.

.

?

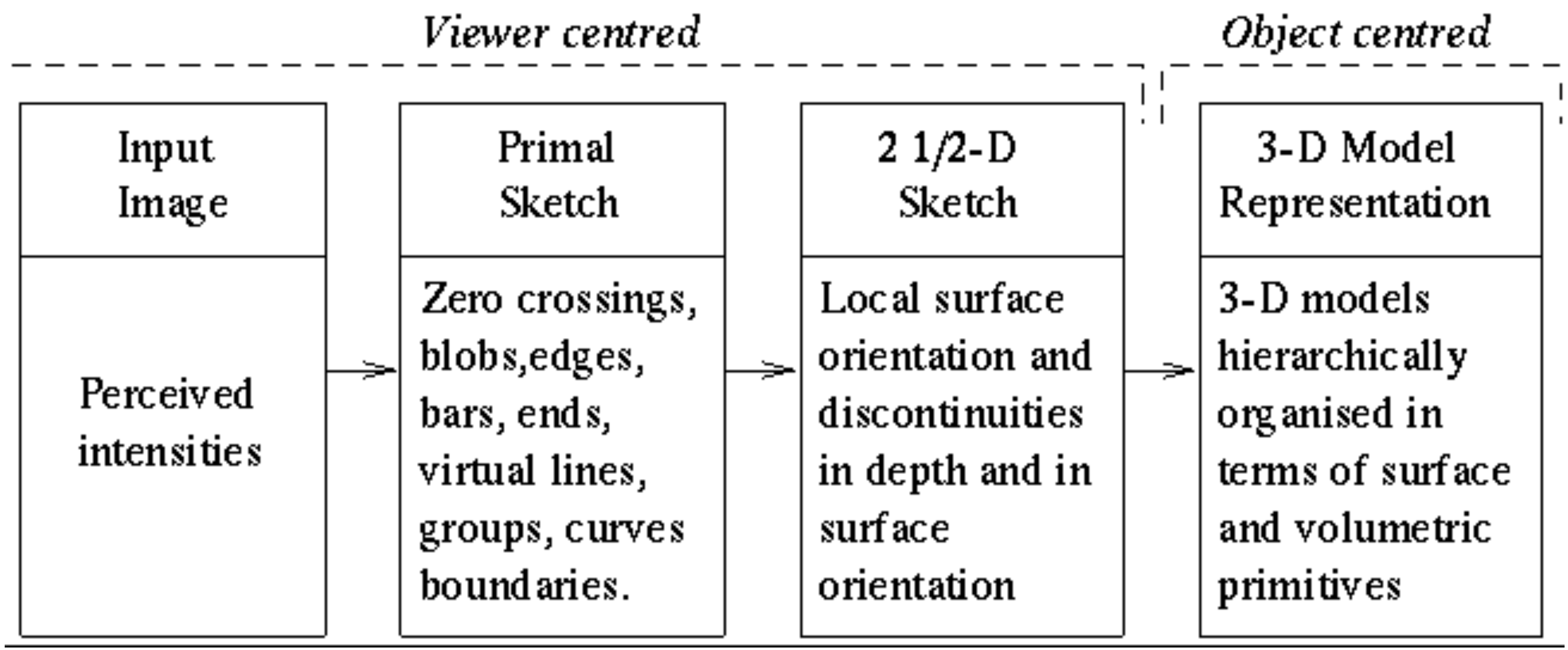
.

.

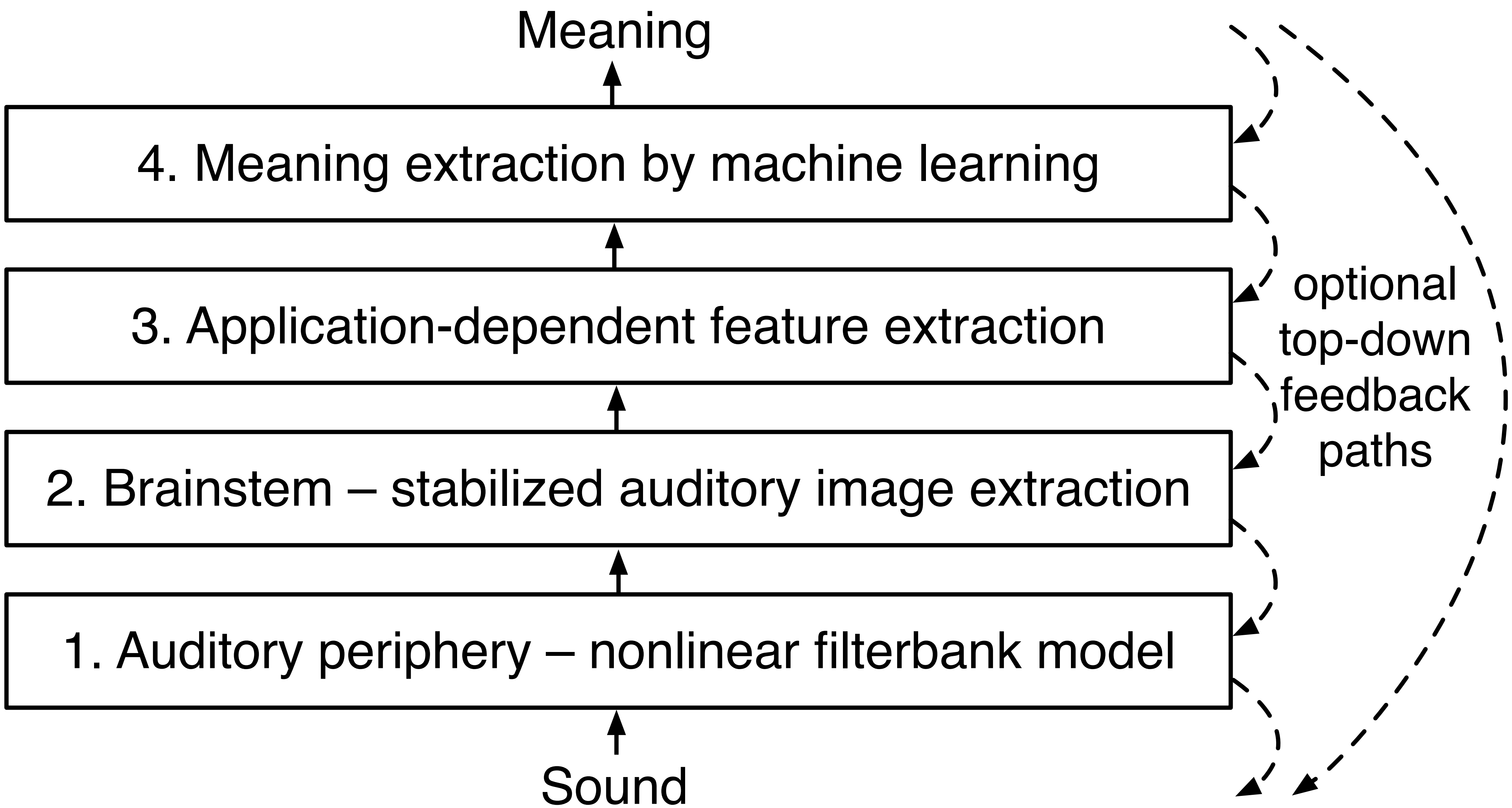
.

'Gabor filters'

The approach of David Marr



The approach of Dick Lyon



‘Deep learning’

(Hinton, Ng, Bengio, Lecun, Google brain, etc.)

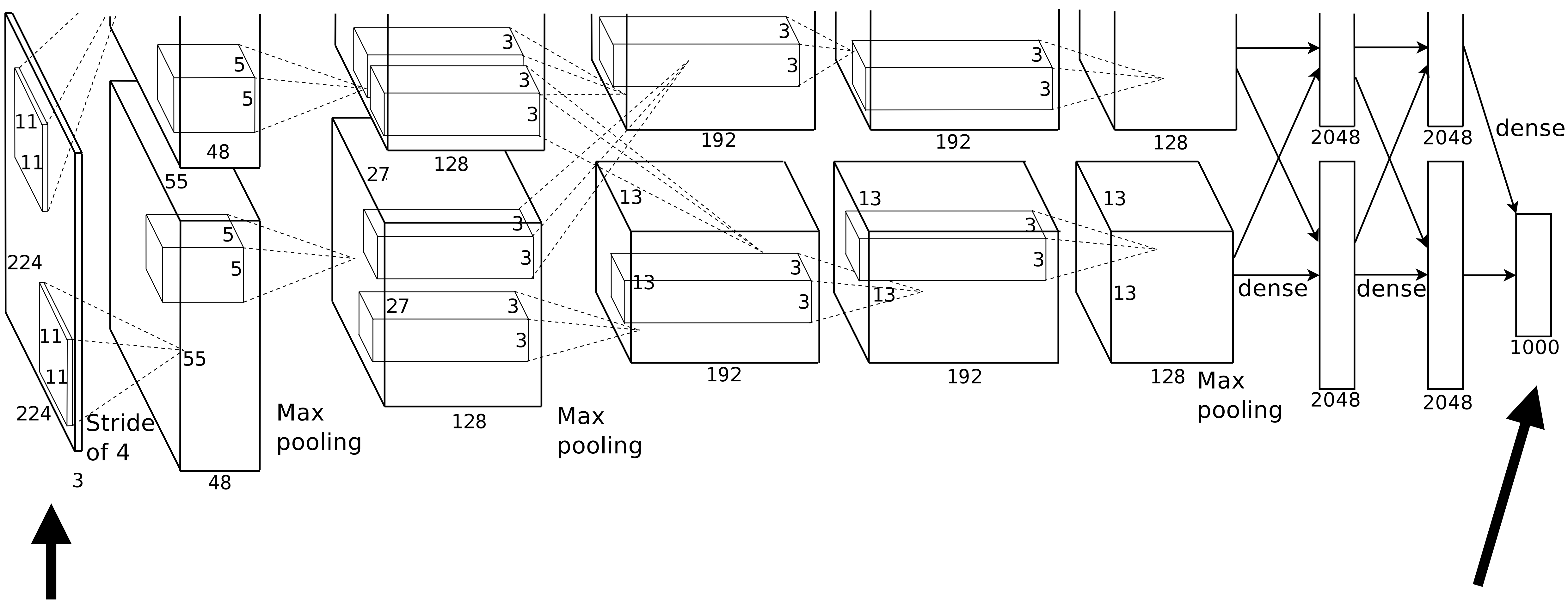
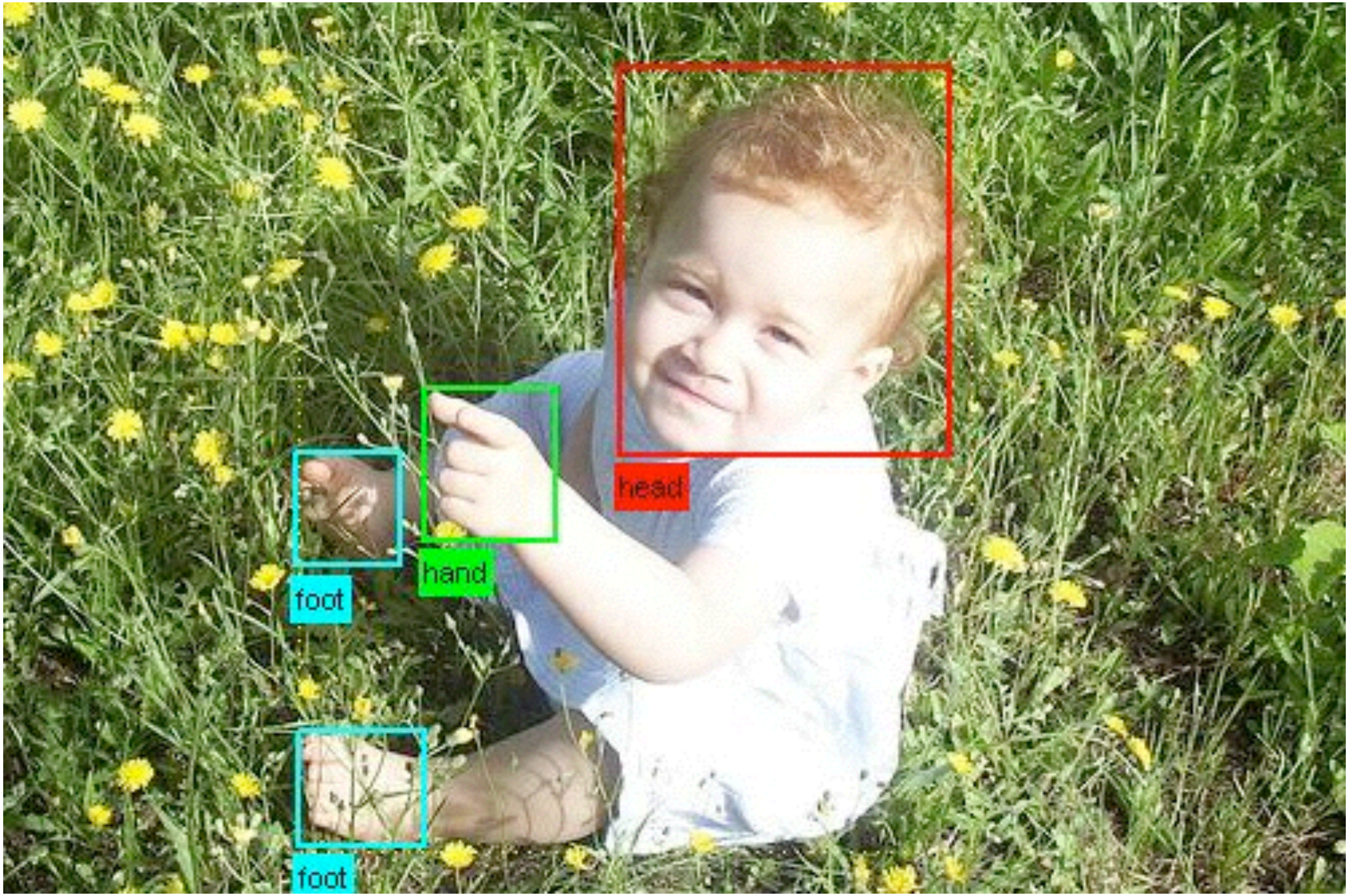
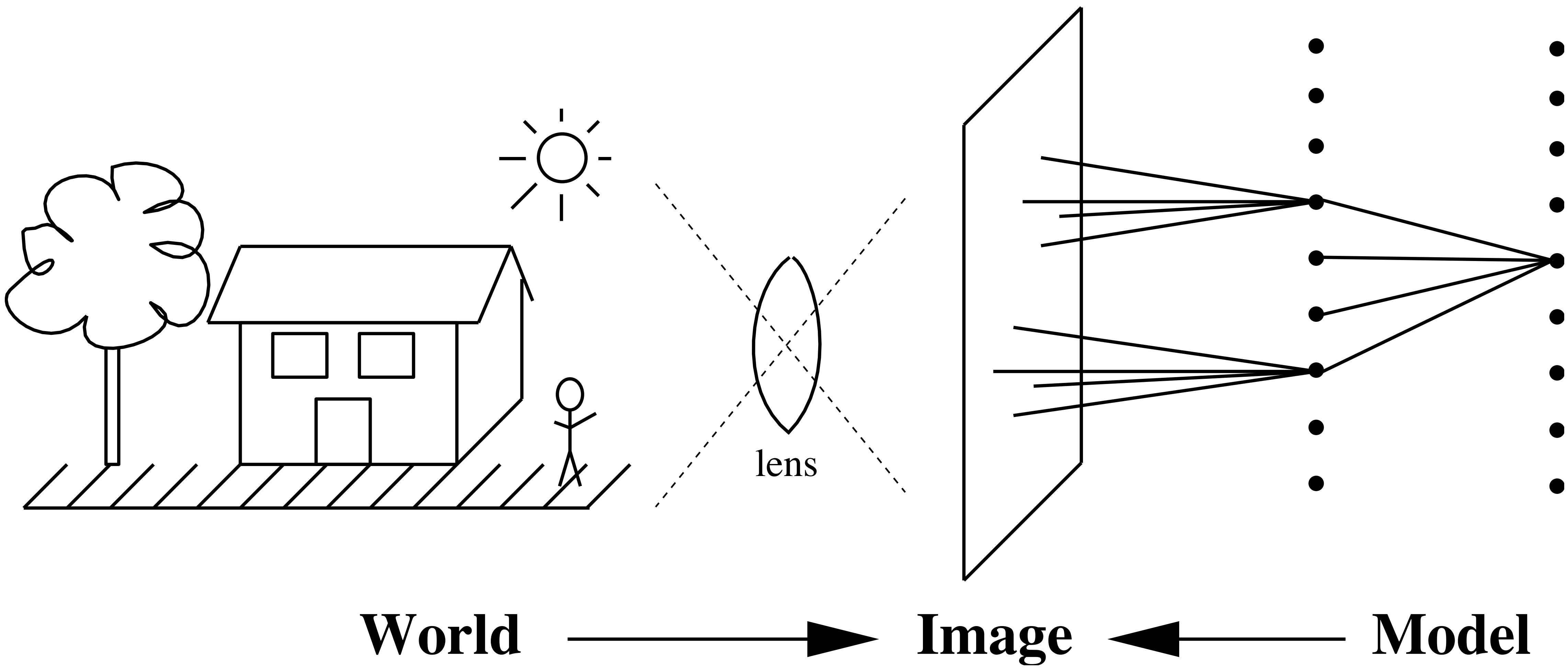


image feature extraction and pooling classification

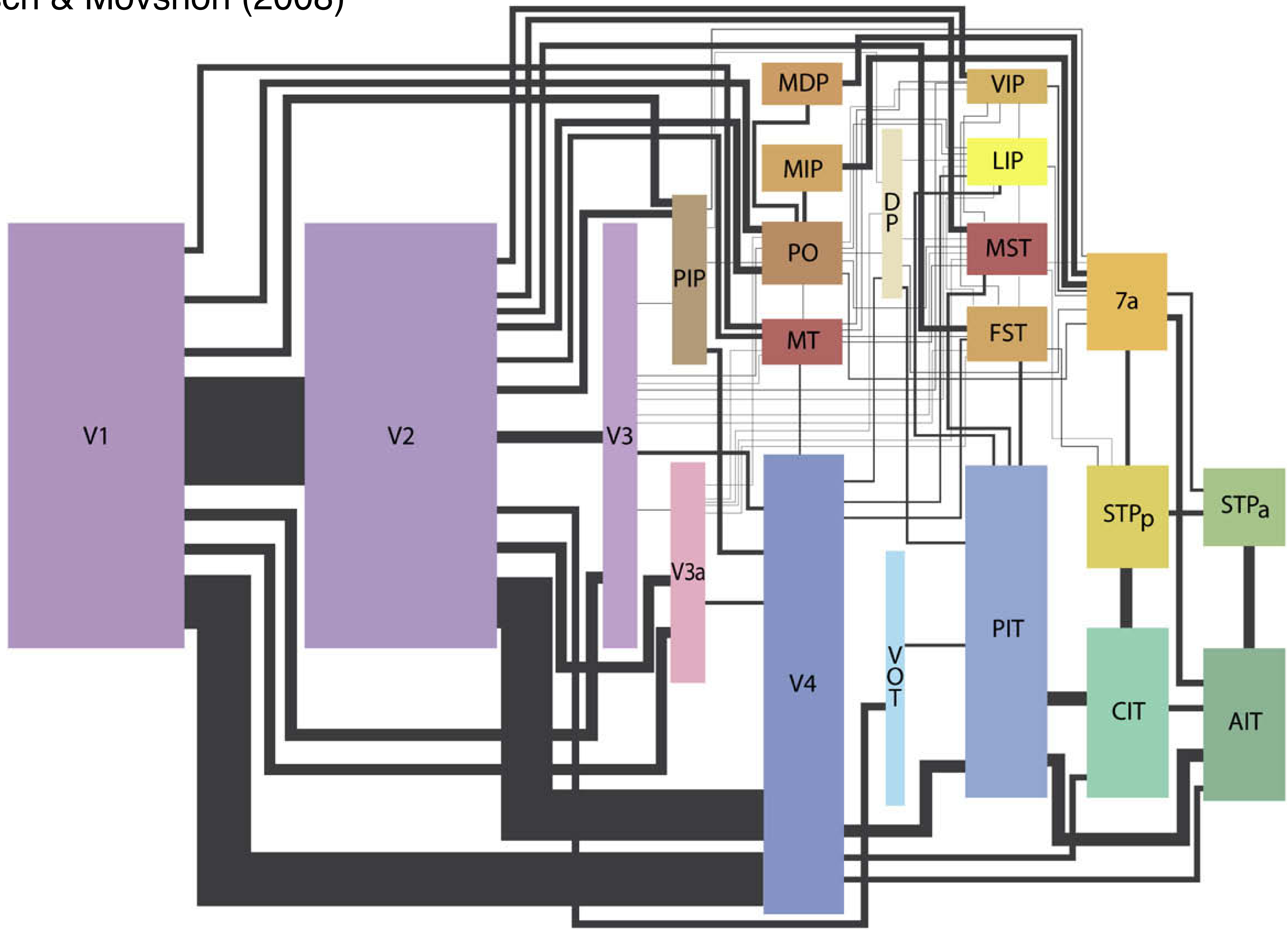


Is this perception?

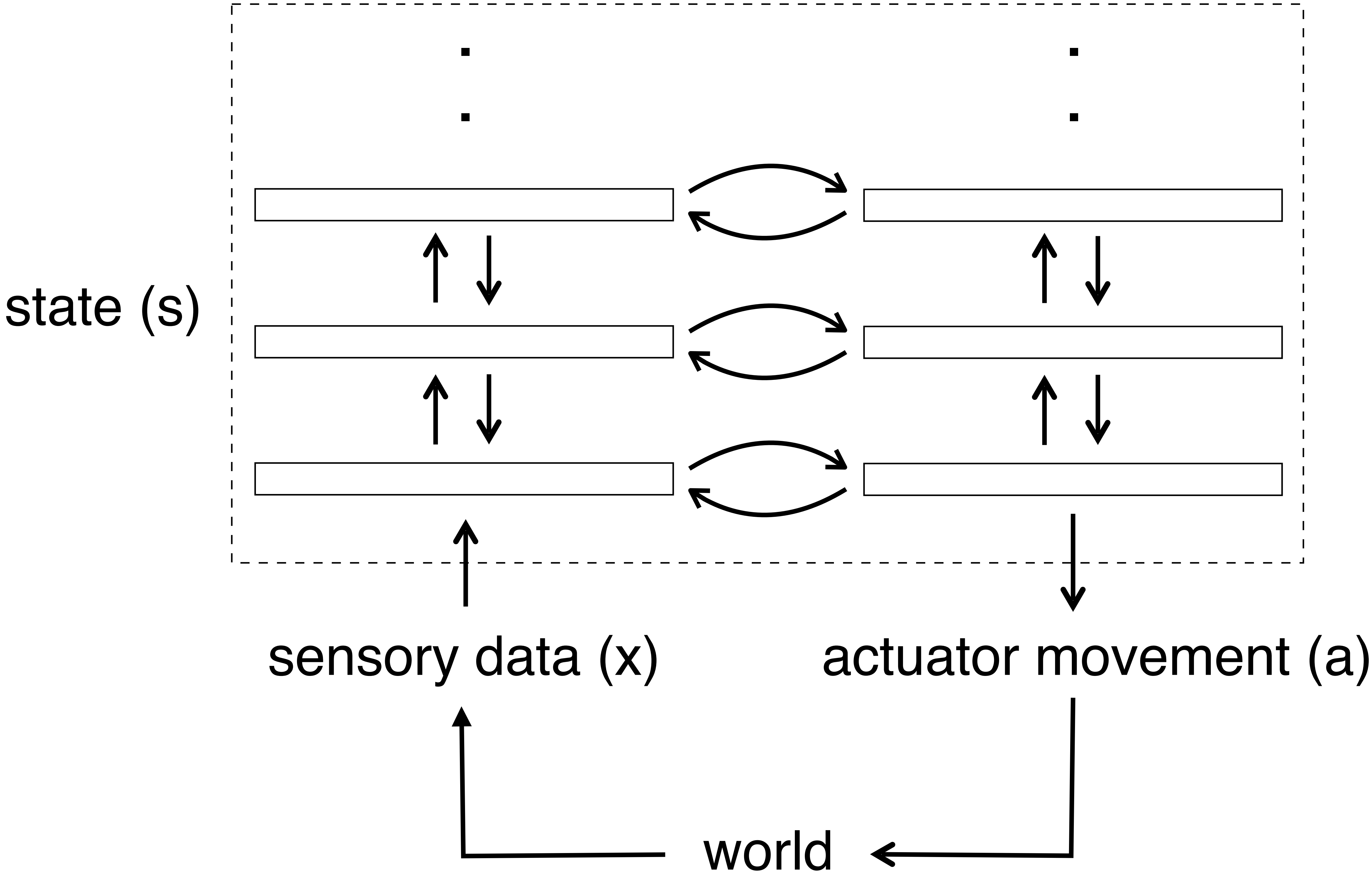
Vision as inference



Wallisch & Movshon (2008)



‘Gabor filters’ · · ? · · objects · faces



state (s)
sensory data (x)
actuator movement (a)
world

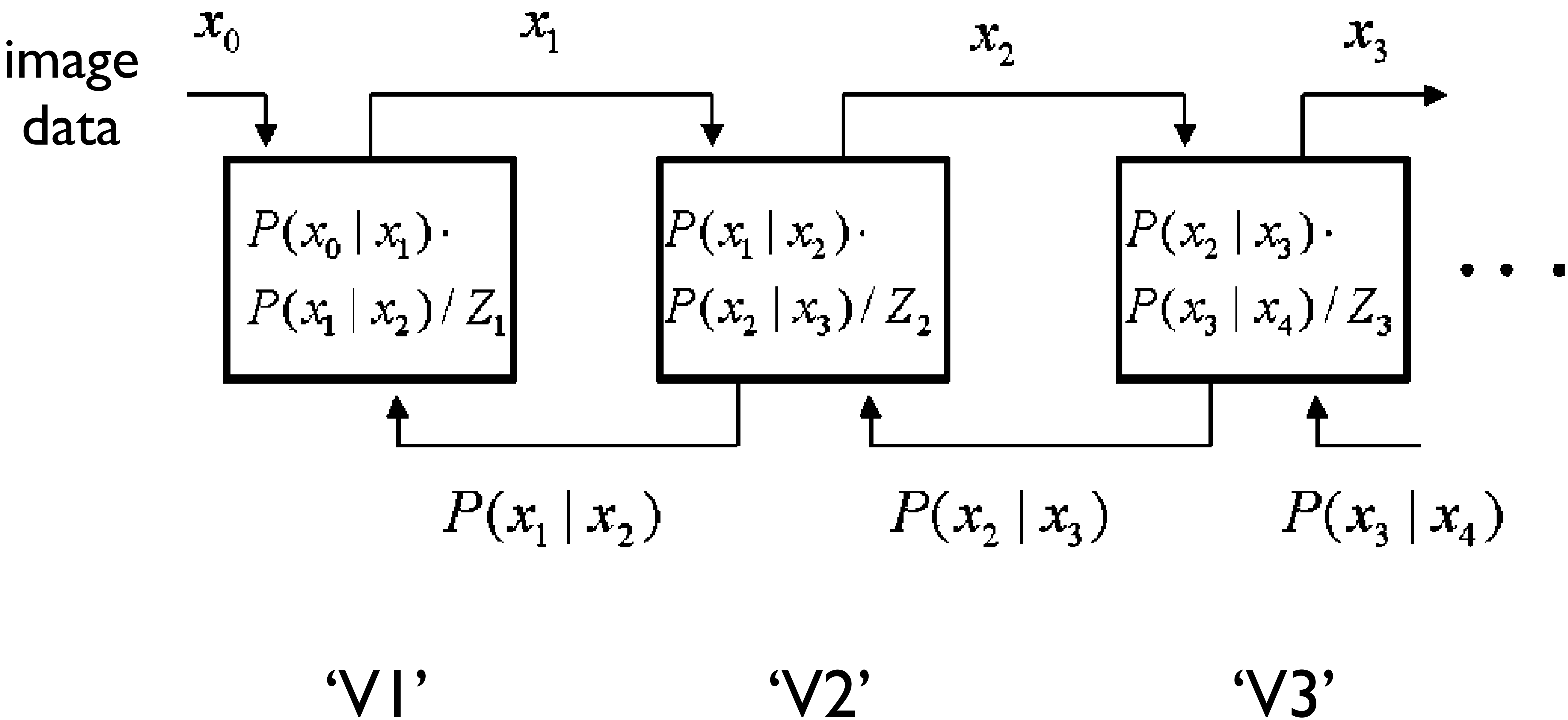
Bayes' rule

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

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

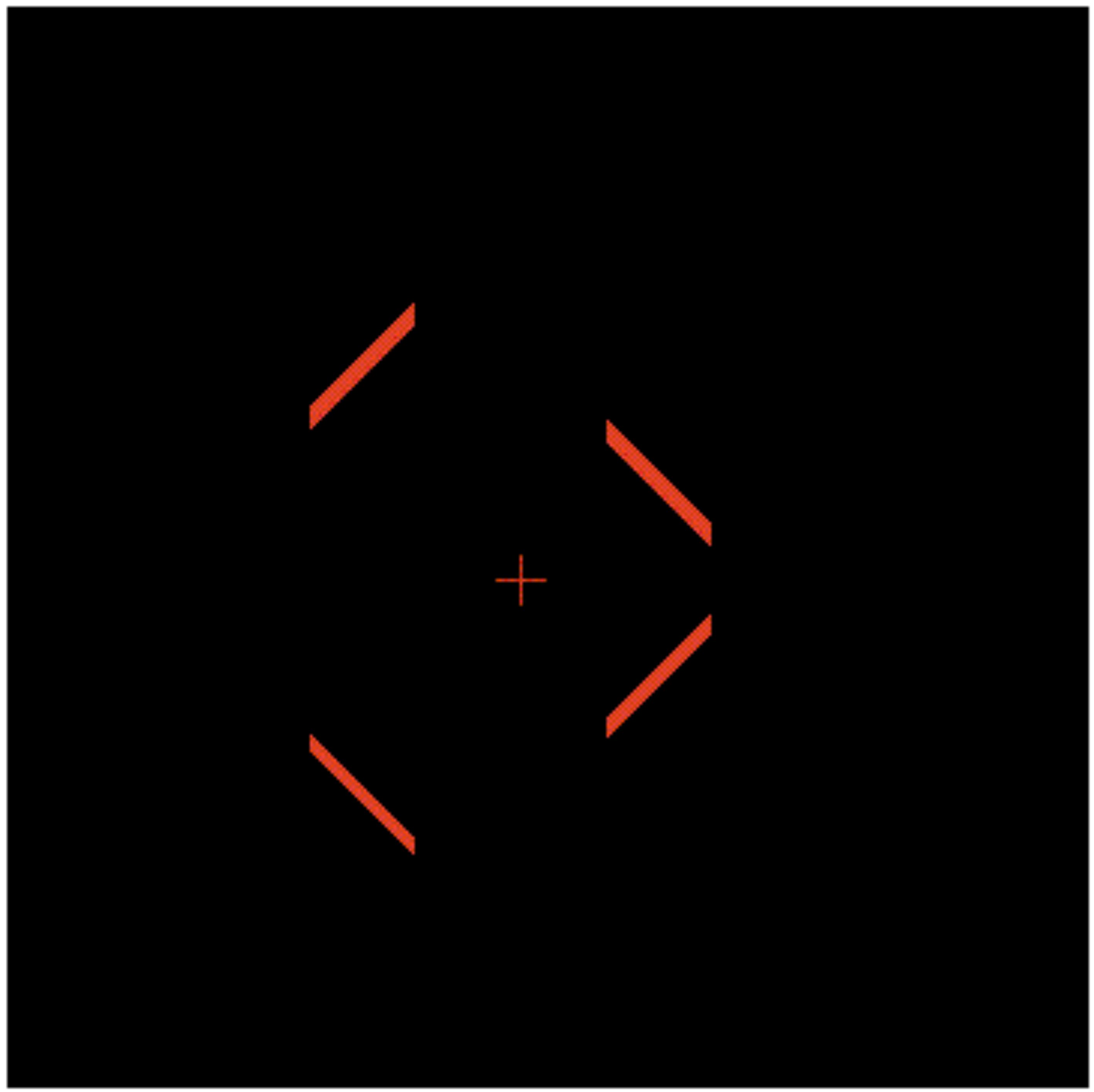
Hierarchical Bayesian inference in visual cortex

(Lee & Mumford, 2003)



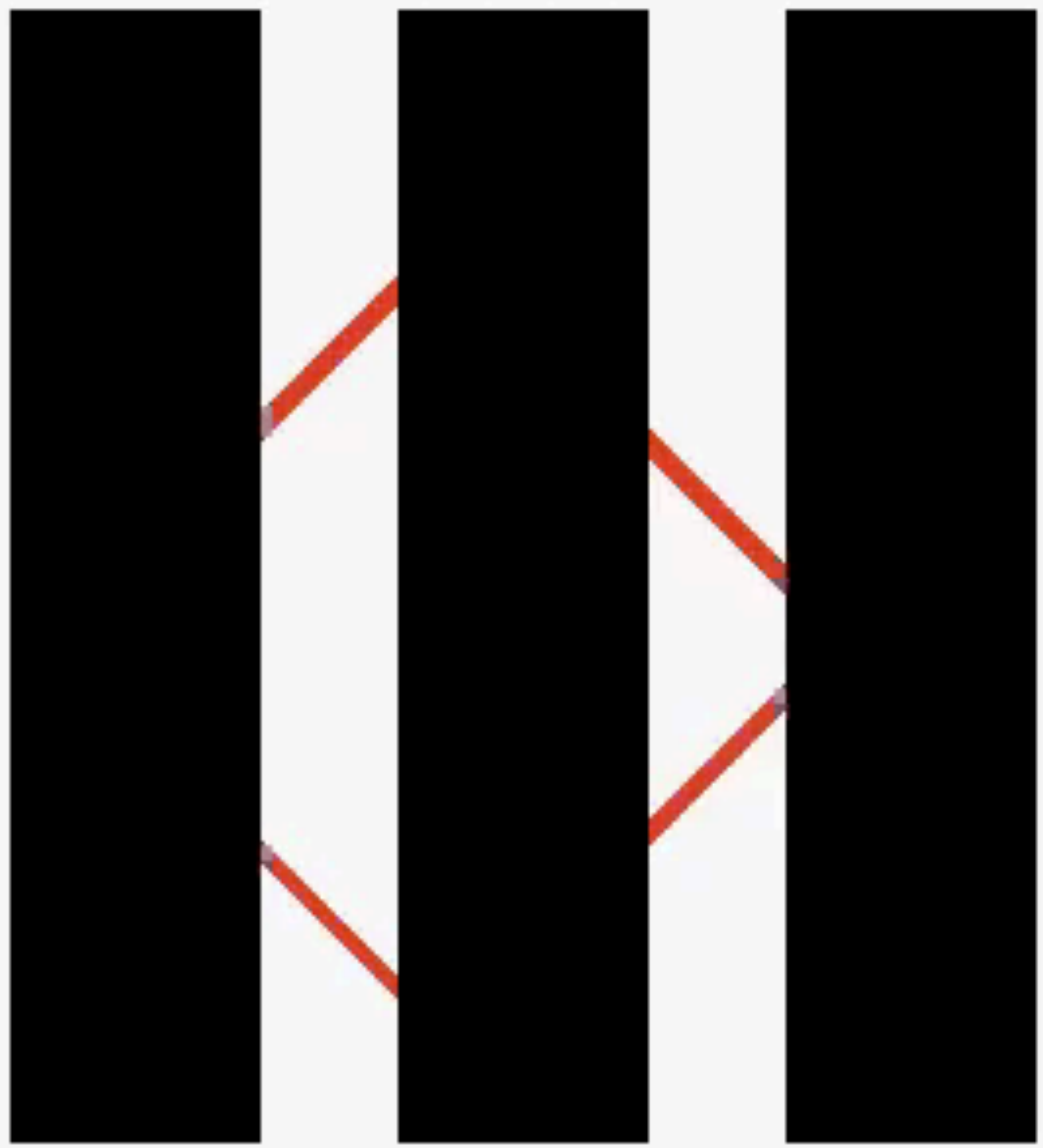
What do you see?

How do neurons in V1 encode this?

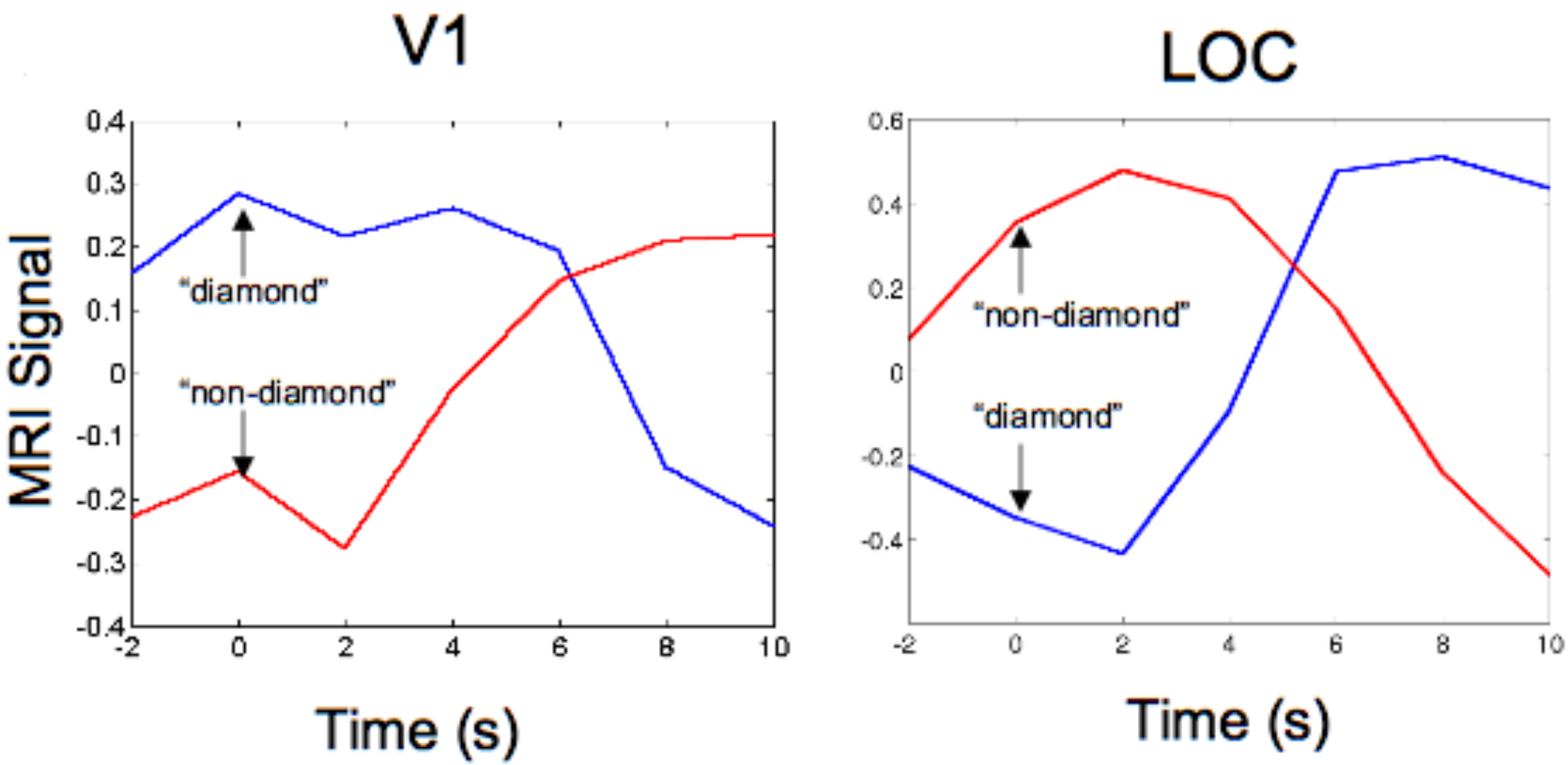


Murray, Kersten, Schrater, Olshausen, Woods, *PNAS* 2002.

(easy version)

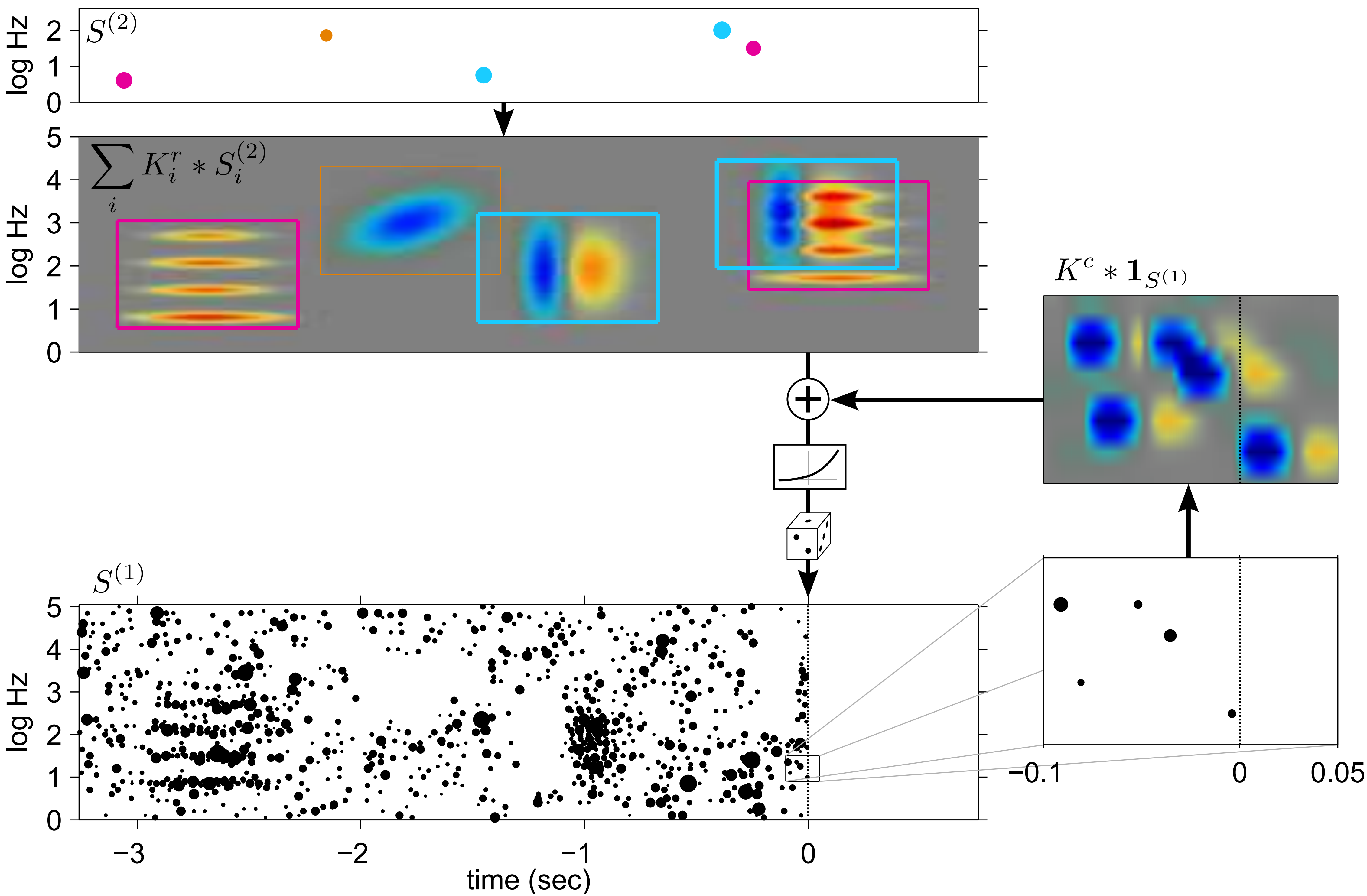


BOLD signal in V1 and LOC



Hierarchical spike coding of sound

(Karklin, Ekanadham & Simoncelli, 2012)



Main points

- The theory of efficient coding has provided important insights about neural representations in the early visual pathway.
- Attempts to apply efficient coding models to the auditory pathway thus far are incomplete and need to be revisited.
- Perception is the problem of inferring a model of the world sufficient for guiding behavior - not well described as a simple input-output chain.
- Hierarchical Bayesian inference provides a promising framework for studying and understanding cortical information processing.