The Sparse Manifold Transform

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

with **Yubei Chen** (EECS) and **Dylan Paiton** (Vision Science)

NSF 1718991 RI: Extracting and understanding
sparse structure in spatiotemporal data, PI: Sommer

REDWOOD CENTER
for Theoretical Neuroscience

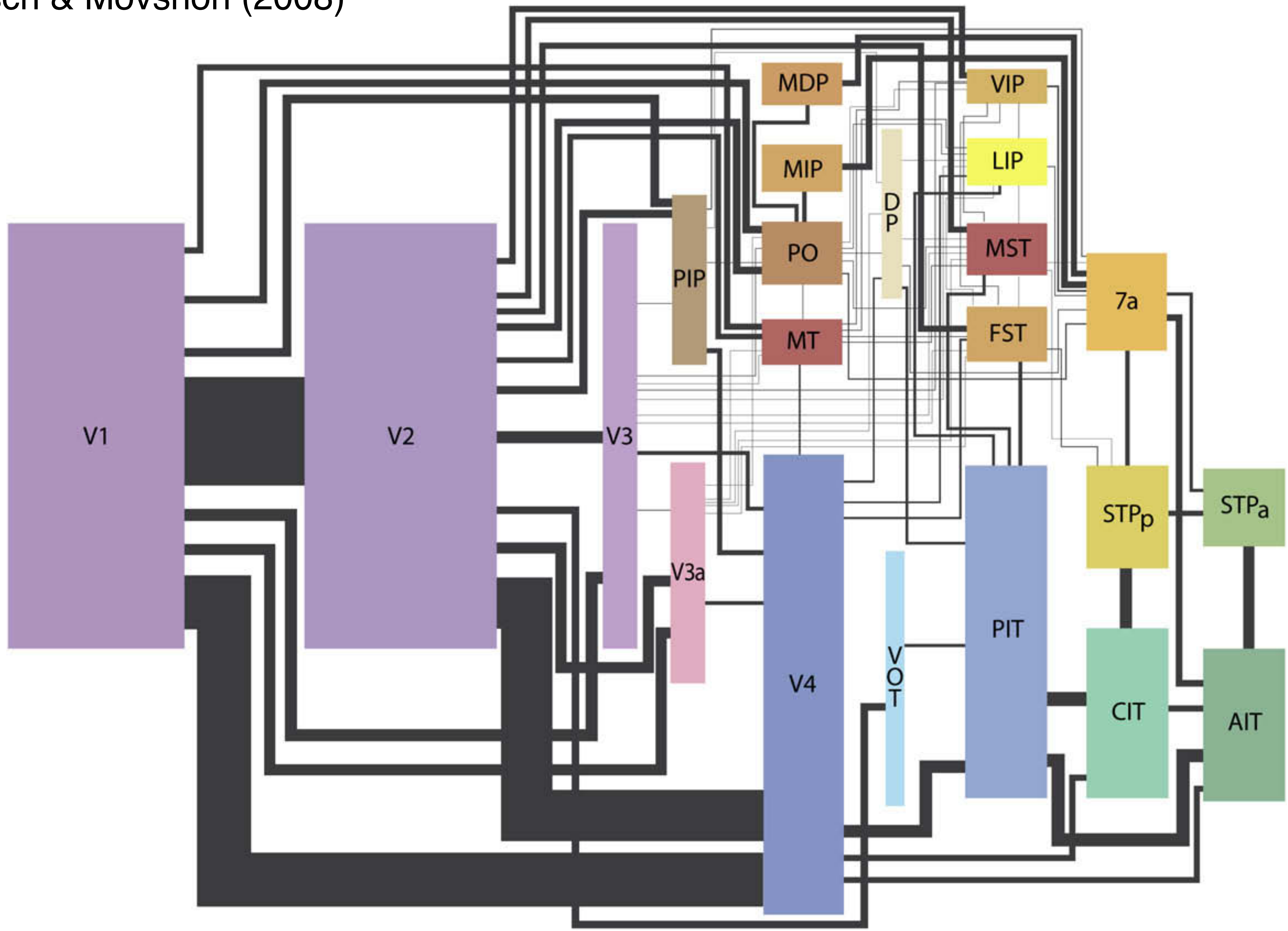
Redwood Center for Theoretical Neuroscience - April 2016

Are there principles?

"God is a hacker"
– Francis Crick

"Individual nerve cells were formerly thought to be unreliable… This was quite wrong, and we now realise their apparently erratic behavior was caused by our ignorance, not the neuron's incompetence."
– H.B. Barlow (1972)

Wallisch & Movshon (2008)



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

Three principles of unsupervised learning

Sparse coding ➔ feature selectivity

Manifold flattening ➔ equivariance

Persistence ➔ invariance

Sparse coding

The 'Ratio Club' (1952)

Harold Shipton. John Bates. W.E. Hick. John Pringle. Donald Sholl. John Westcott. Donald Mackay.

Giles Brindley. Tom McLardy. Ross Ashby. Thomas Gold. Albert Uttley.

Alan Turing. Gurney Sutton. William Rushton. George Dawson. Horace Barlow.

Perception, 1972, volume 1, pages 371-394

Single units and sensation: A neuron doctrine for perceptual psychology?

H B Barlow
Department of Physiology-Anatomy, University of California, Berkeley, California 94720
Received 6 December 1972

Abstract. The problem discussed is the relationship between the firing of single neurons in sensory pathways and subjectively experienced sensations. The conclusions are formulated as the following five dogmas:
1. To understand nervous function one needs to look at interactions at a cellular level, rather than either a more macroscopic or microscopic level, because behaviour depends upon the organized pattern of these intercellular interactions.

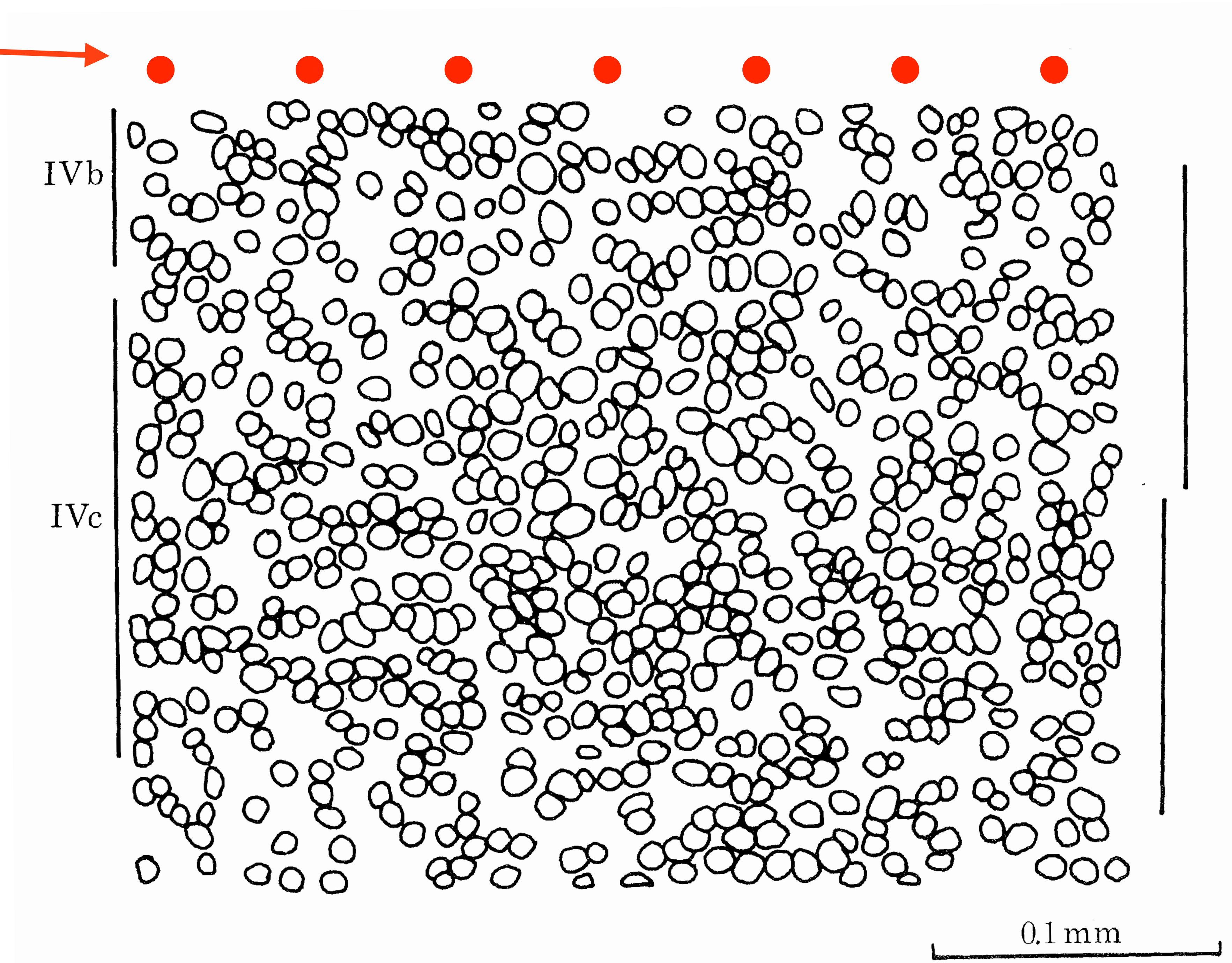
2. The sensory system is organized to achieve as complete a representation of the sensory stimulus as possible with the minimum number of active neurons.

neurons, each of which corresponds to a pattern of external events of the order of complexity of the events symbolized by a word.
5. High impulse frequency in such neurons corresponds to high certainty that the trigger feature is present.
The development of the concepts leading up to these speculative dogmas, their experimental basis, and some of their limitations are discussed.

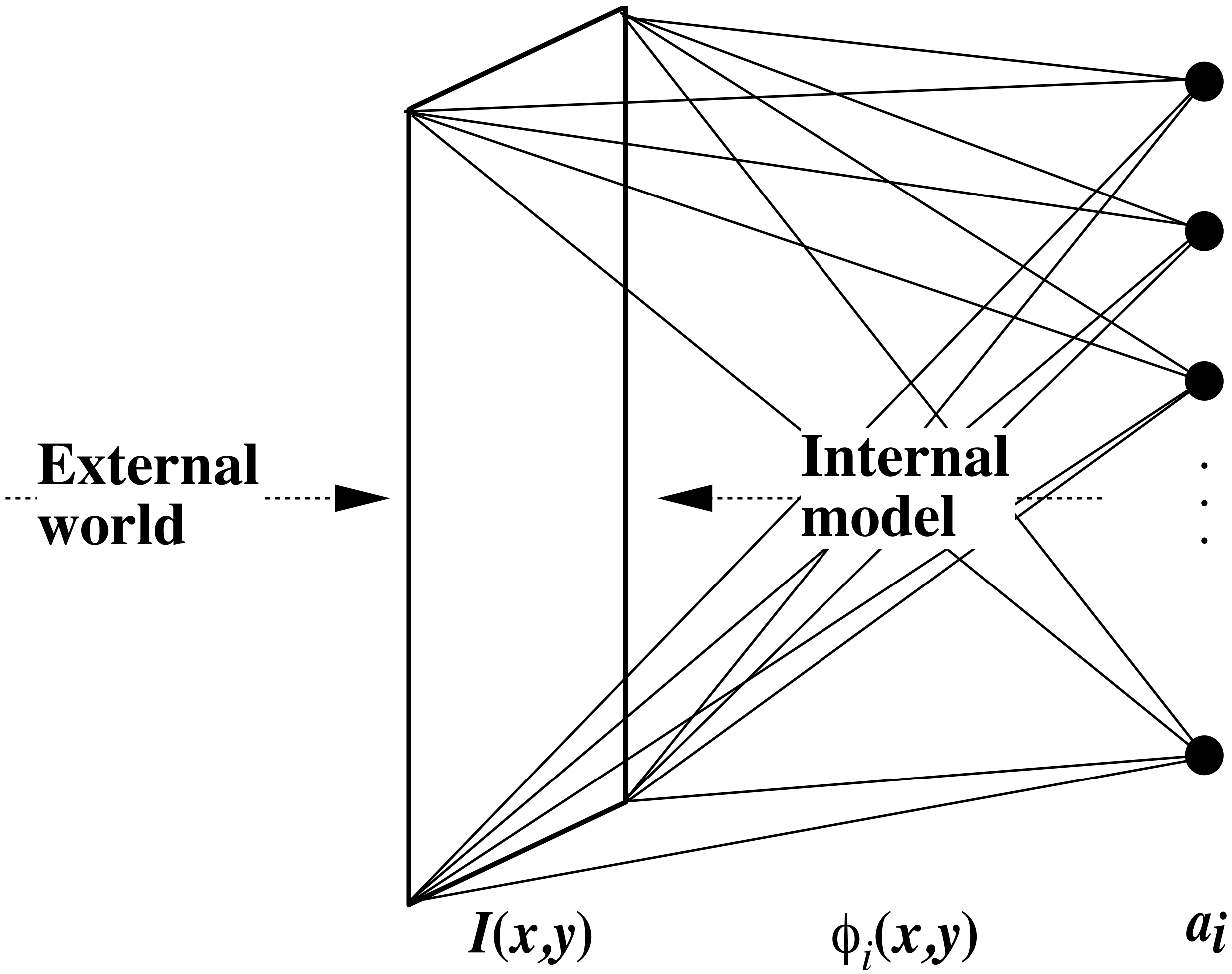
V1 is highly overcomplete

LGN afferents

layer 4 cortex



Barlow (1981)



Sparse coding image model

(Olshausen & Field, 1996; Chen, Donoho & Saunders 1995)

$$I(\vec{x}) = \sum_{i=1}^{M} a_i \phi_i(\vec{x}) + \epsilon(\vec{x})$$

image

neural activities (sparse)

features

other stuff

Energy function

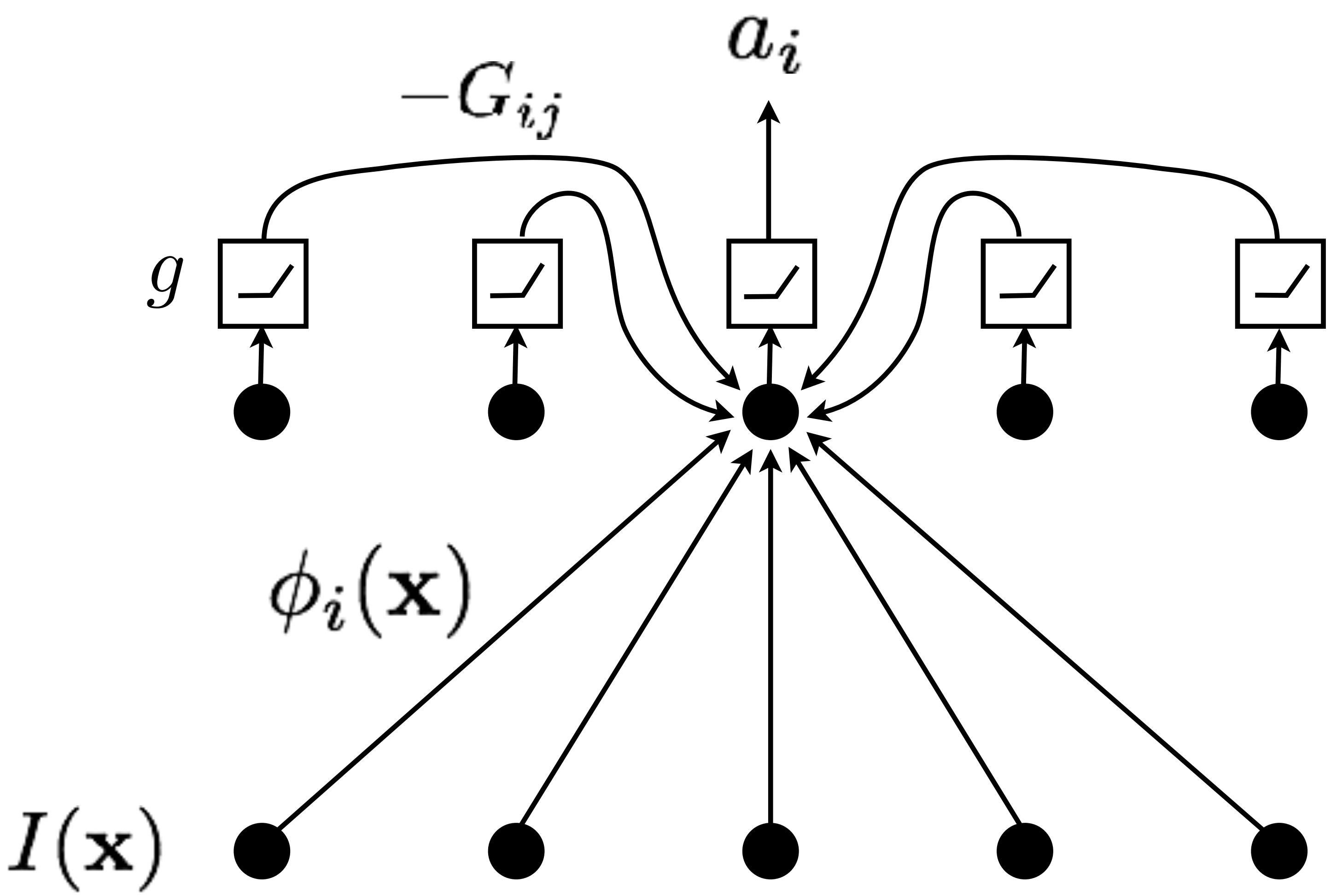
$$E = \frac{1}{2}|\mathbf{I} - \Phi\,\mathbf{a}|^2 + \lambda \sum_i C(a_i)$$

preserve information

be sparse

Locally Competitive Algorithm (LCA) minimizes the energy function

(Rozell, Johnson, Baraniuk & Olshausen, 2008)

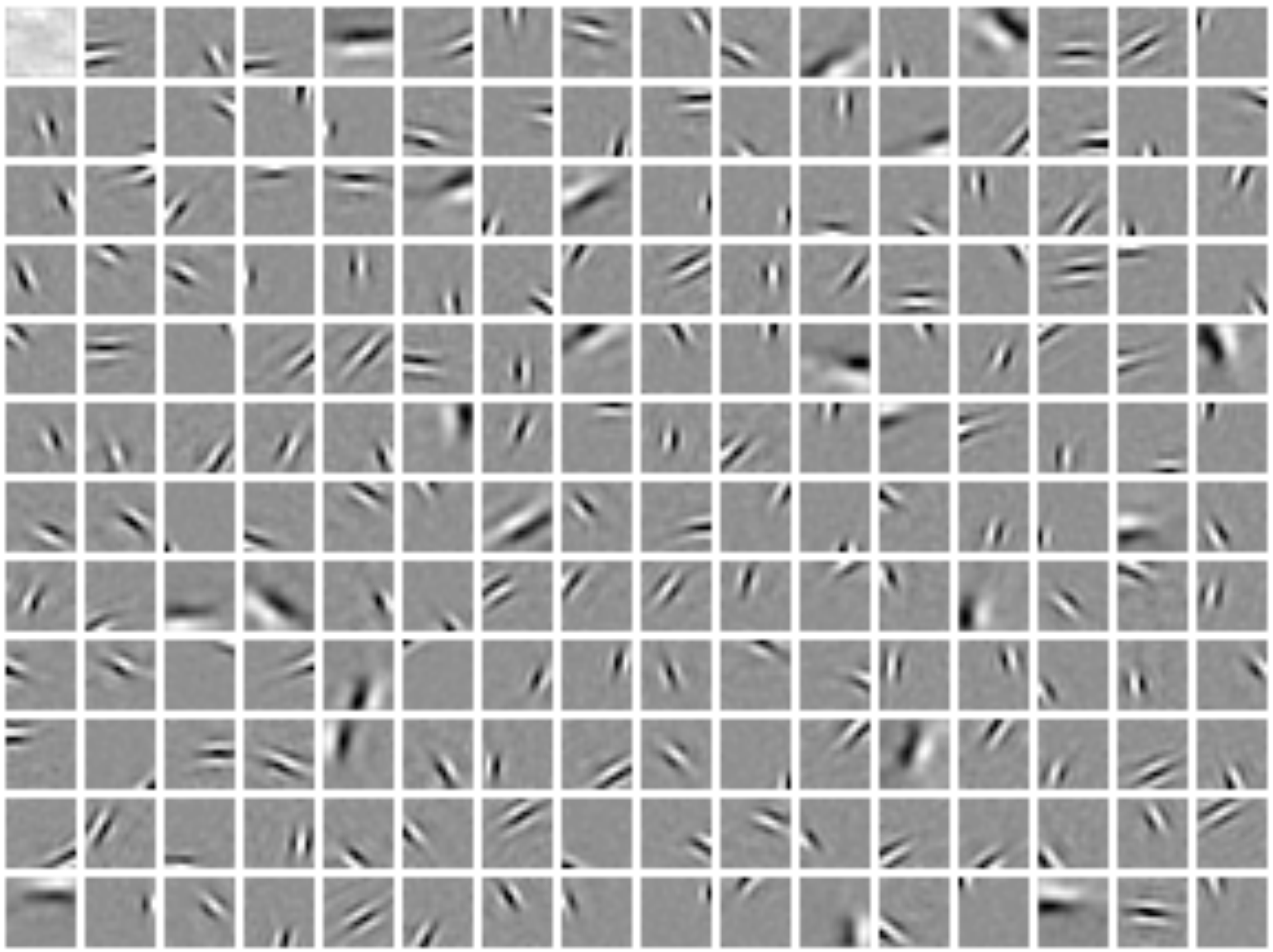


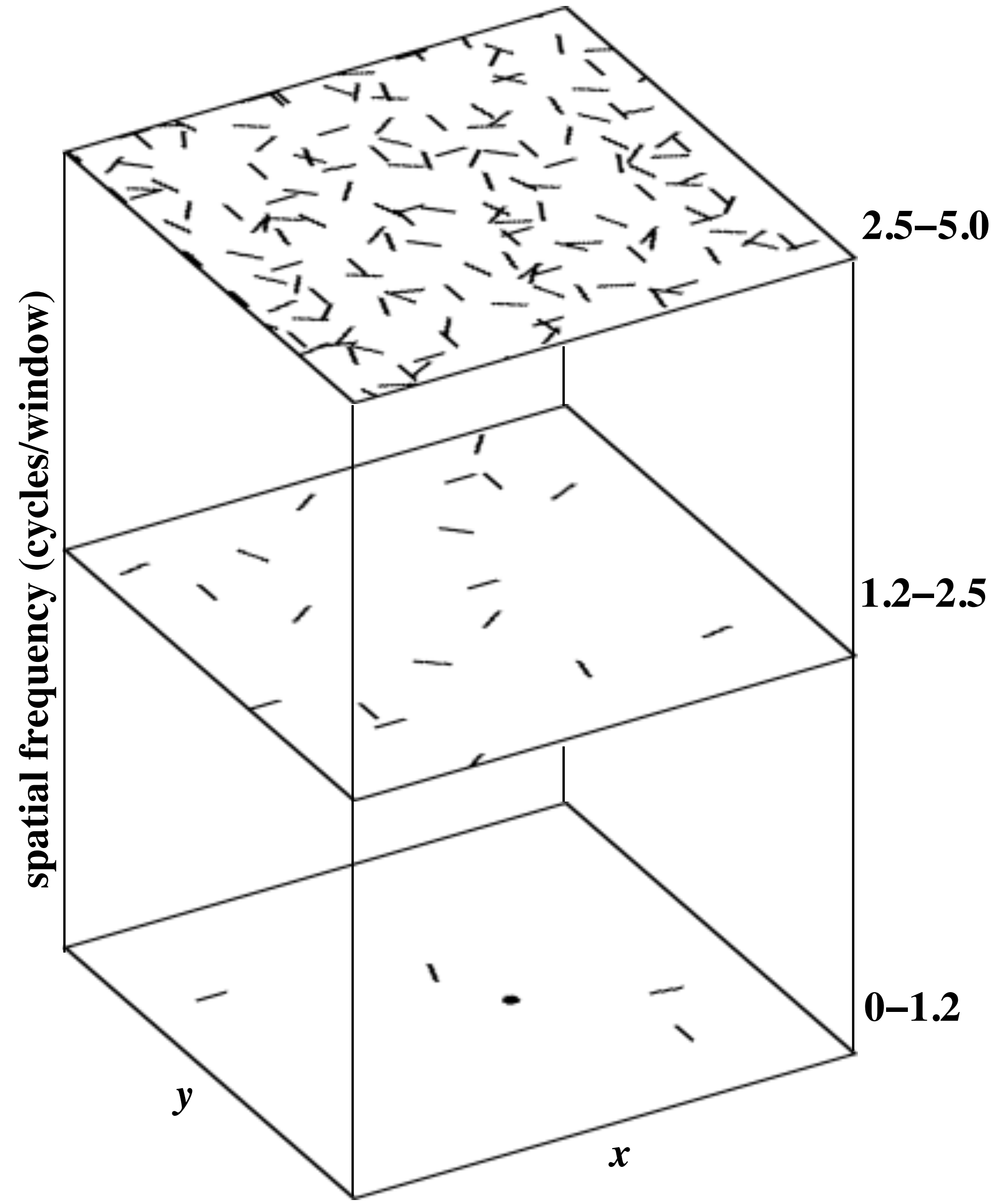
$$\begin{aligned} \tau\, \dot{u}_i + u_i &= b_i - \sum_{j \neq i} G_{ij}\, a_j \\ a_i &= g(u_i) \end{aligned}$$

$$b_i = \sum_{\mathbf{x}} \phi_i(\mathbf{x})\, I(\mathbf{x})$$

$$G_{ij} = \sum_{\mathbf{x}} \phi_i(\mathbf{x})\, \phi_j(\mathbf{x})$$

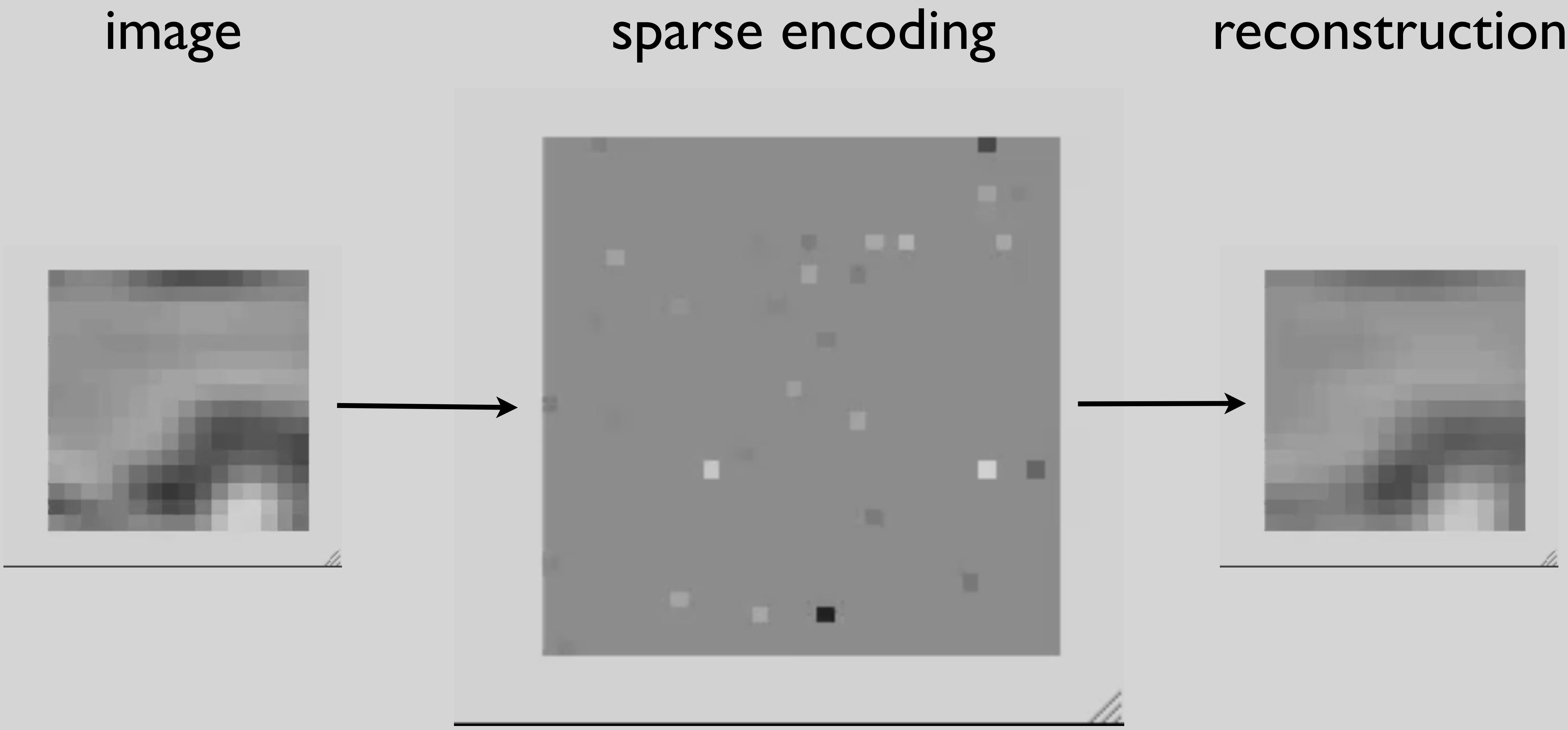
Learned dictionary $\{\phi_i\}$



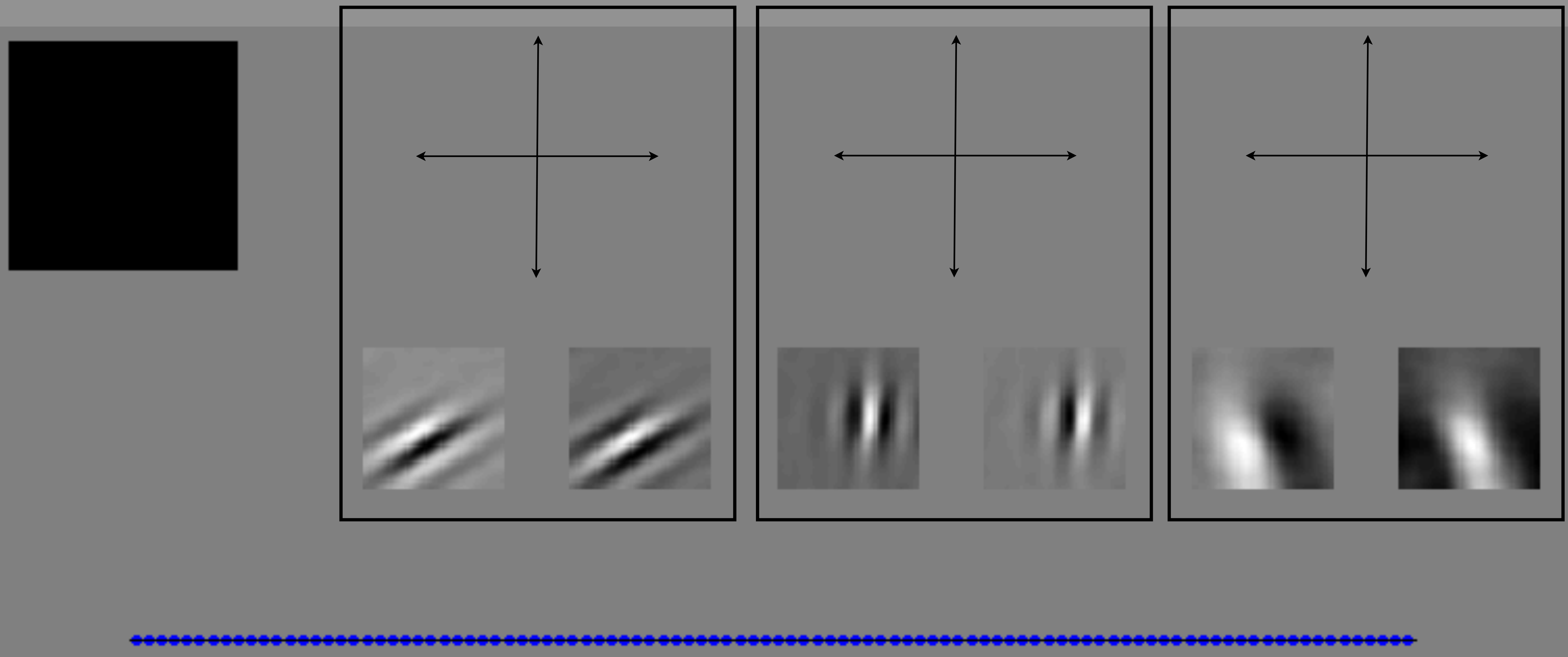


Olshausen & Field (1996)

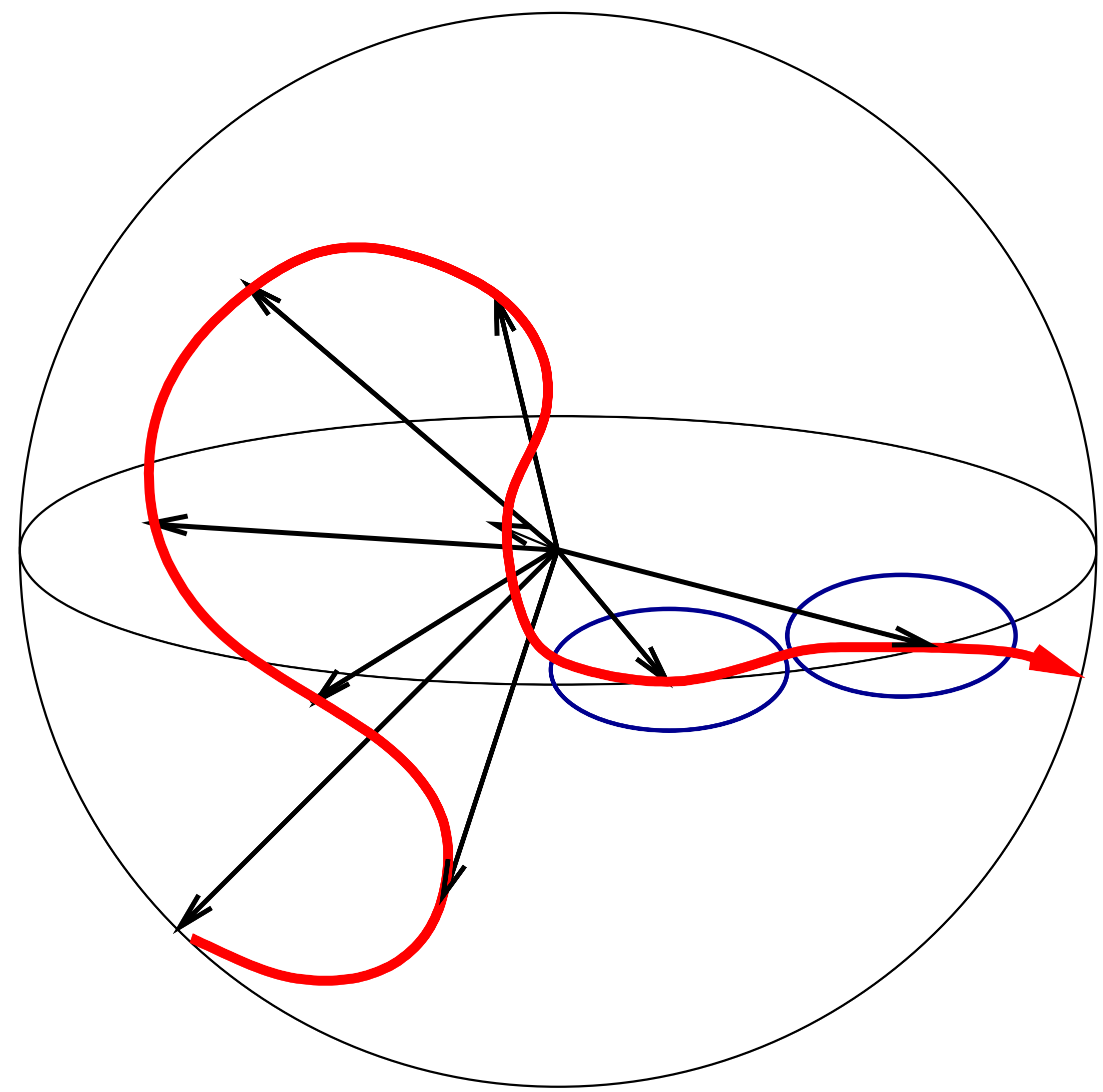
Sparse encoding of a time-varying image



Sparse coefficient activations form smooth trajectories in particular local subspaces



Basis functions tile the manifold of natural images in such a way that data points along the manifold are spanned by a small number of basis vectors.



Manifold flattening

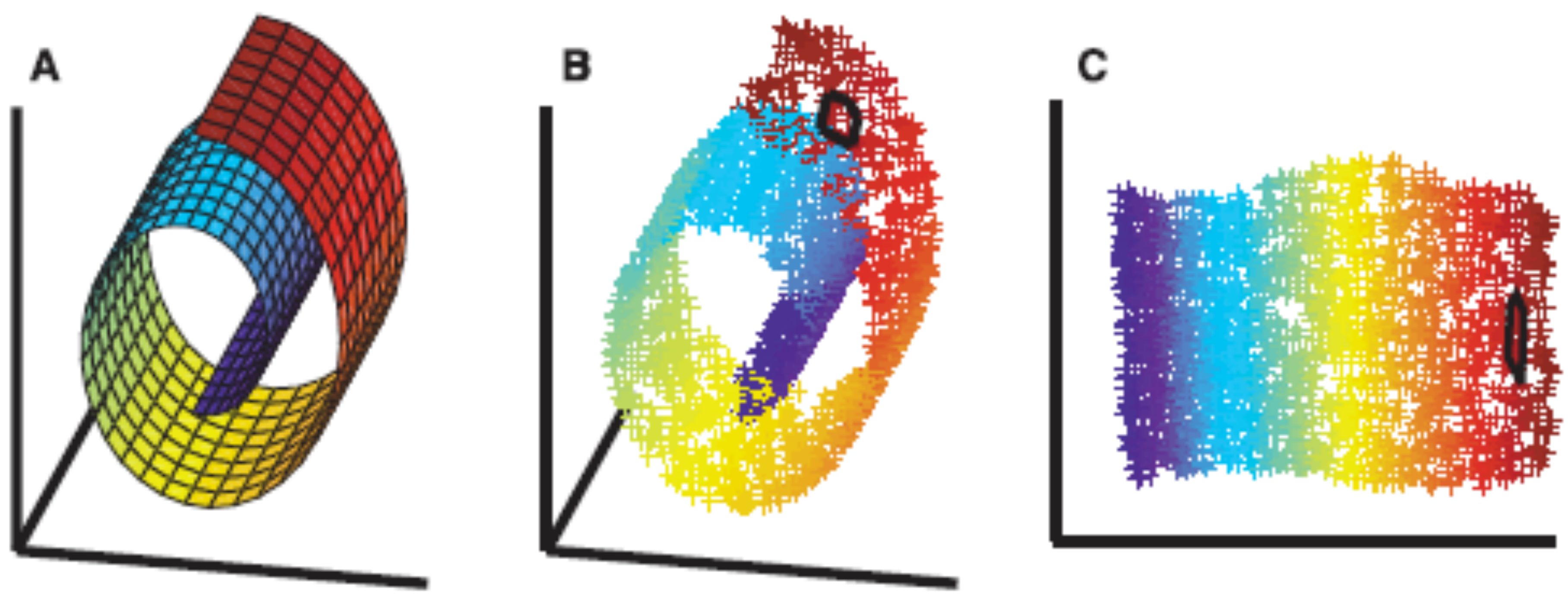
Nonlinear Dimensionality Reduction by Locally Linear Embedding

Sam T. Roweis[1] and Lawrence K. Saul[2]

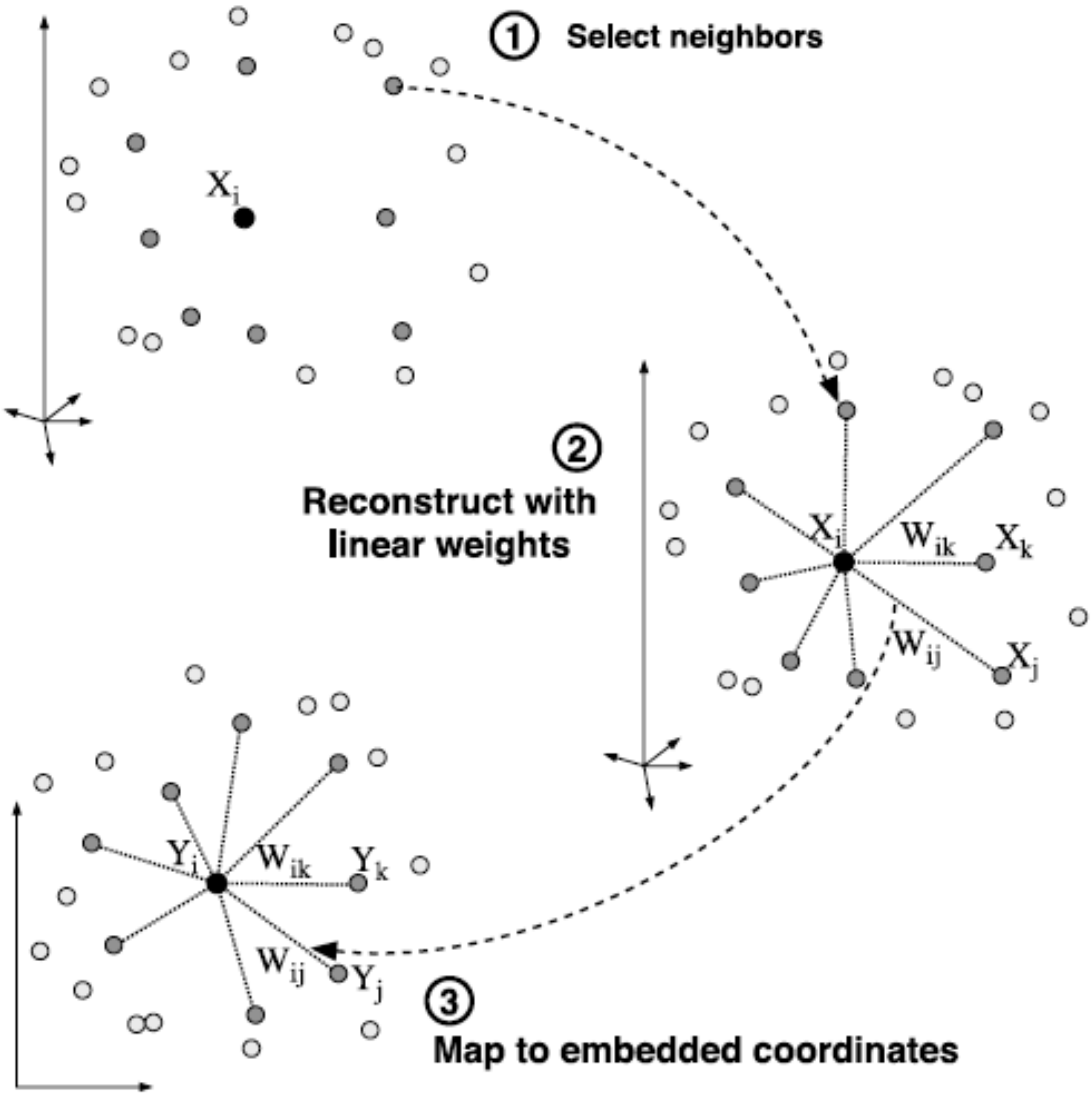
A Global Geometric Framework for Nonlinear Dimensionality Reduction

Joshua B. Tenenbaum,[1*] Vin de Silva,[2] John C. Langford[3]

Science, 22 Dec. 2000



Local Linear Embedding (LLE)

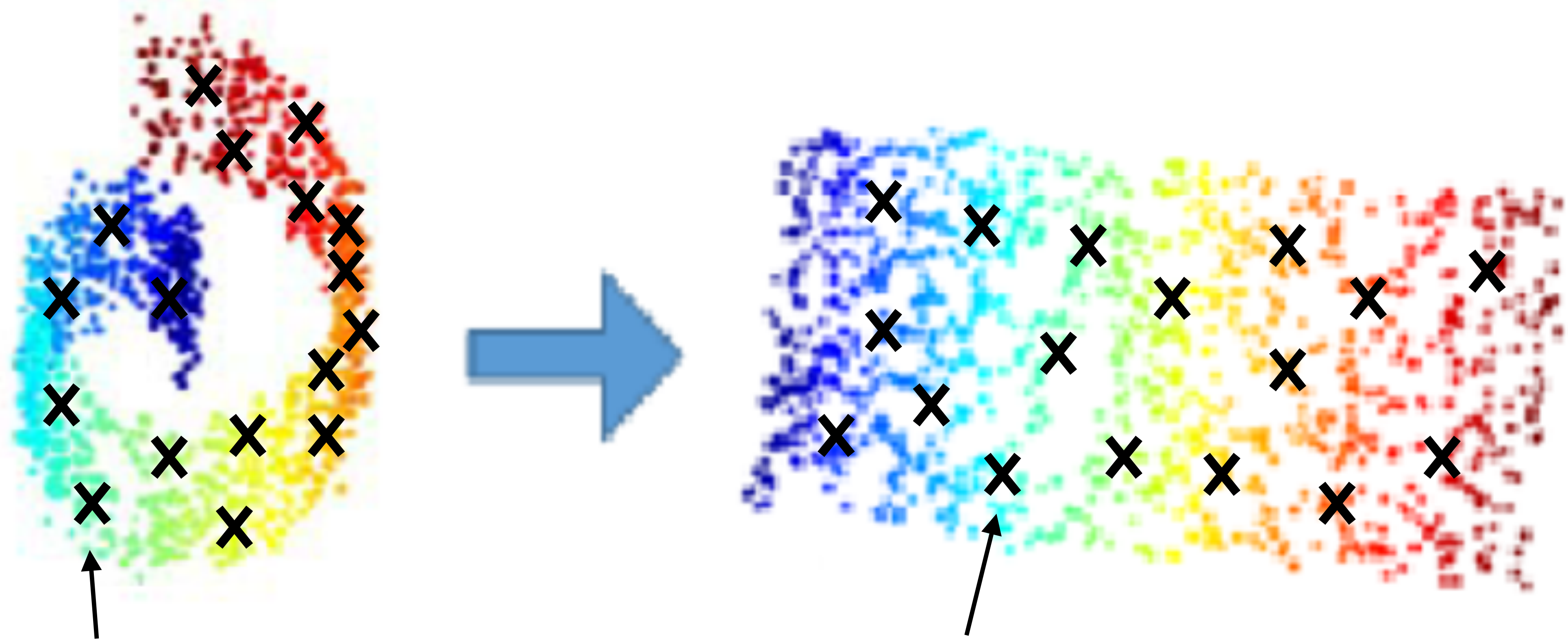


$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \Sigma_j W_{ij} \vec{X}_j \right|^2$$

$$\Phi(Y) = \sum_i \left| \vec{Y}_i - \Sigma_j W_{ij} \vec{Y}_j \right|^2$$

Local Linear Landmarks (LLL)

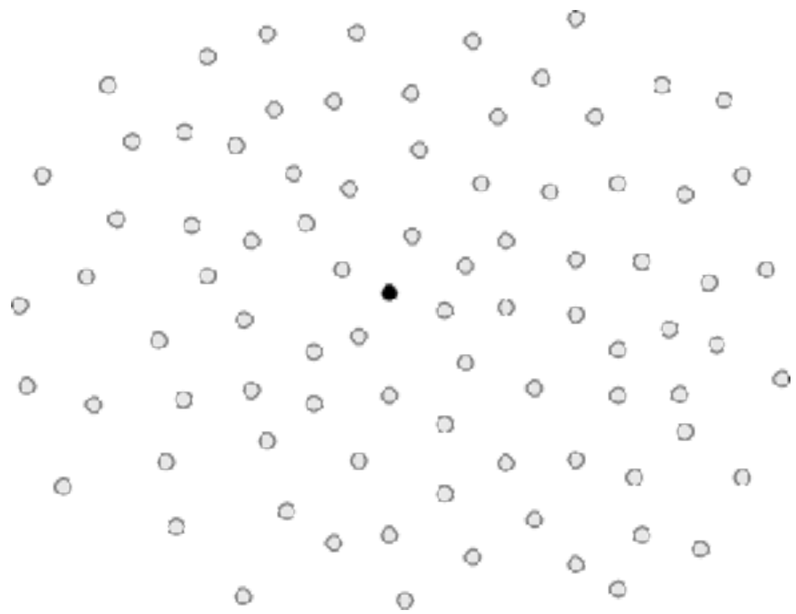
(Vladymyrov & Carreira-Perpinán, 2013)



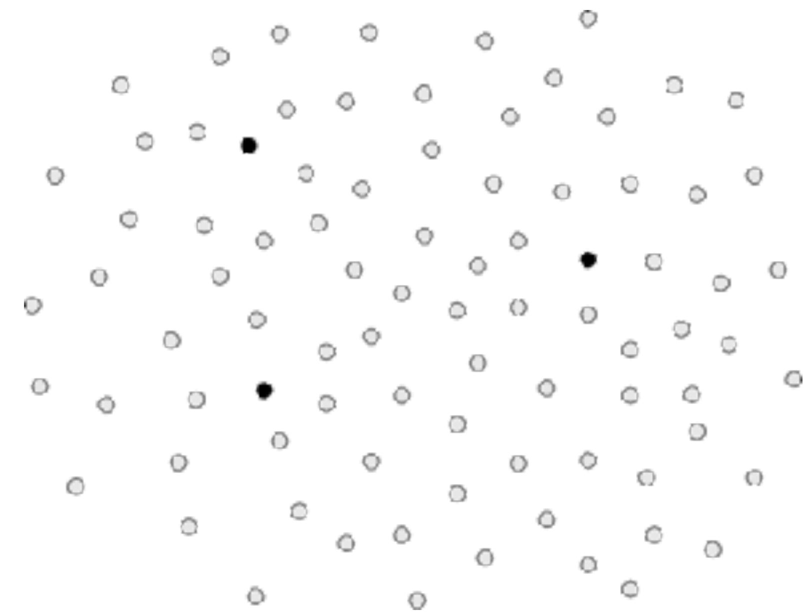
$\Phi_i \longrightarrow P_i$

$$x = \Phi\,\alpha + n \longrightarrow y = P\,\alpha$$

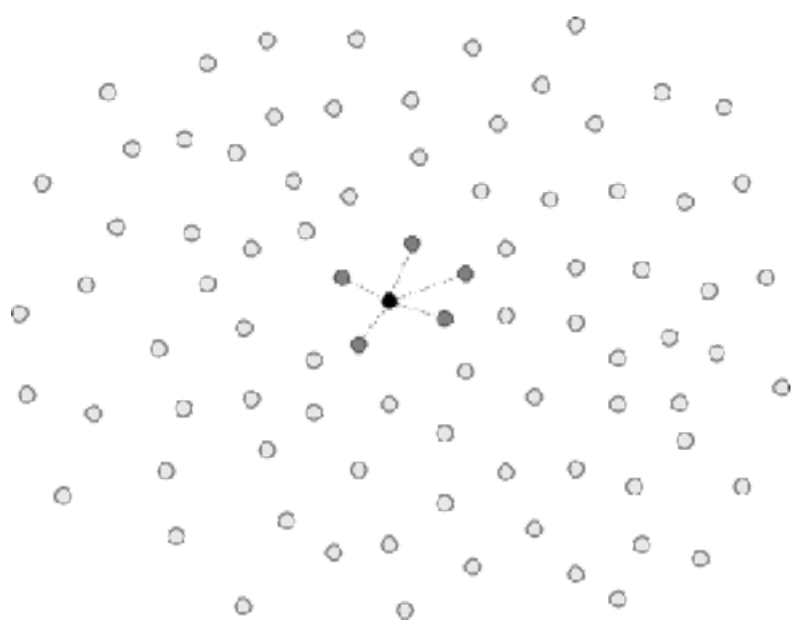
From 1-sparse to k-sparse



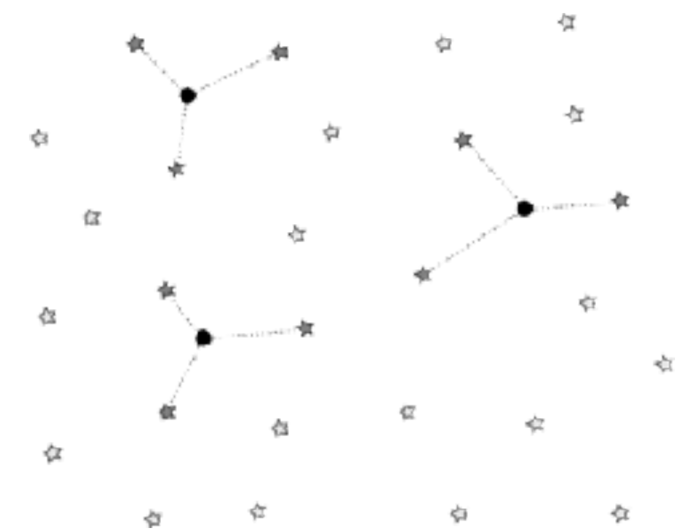
1-sparse



k-sparse

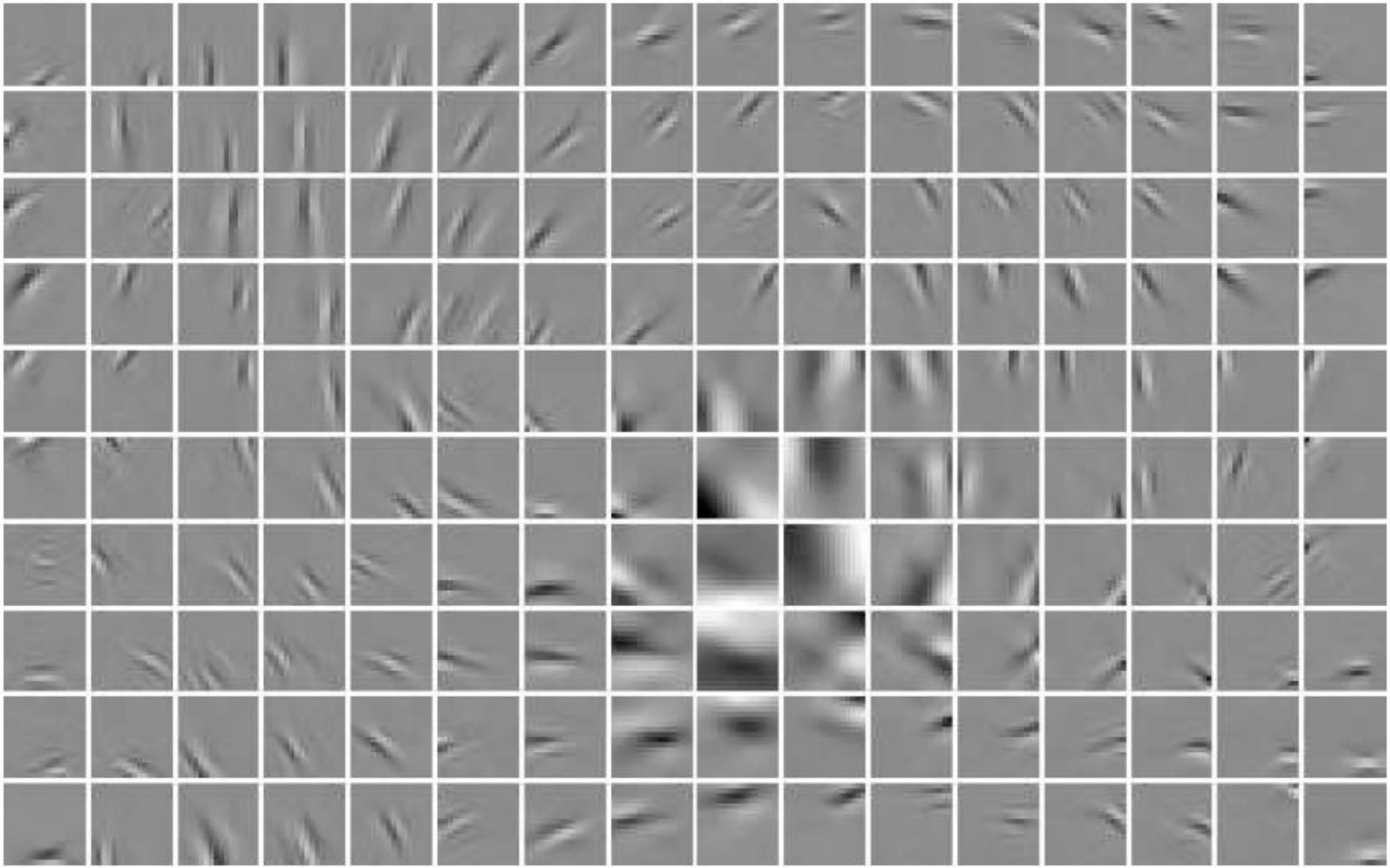


Interpolation by KNN data vectors



Interpolation by landmarks or dictionary ✔

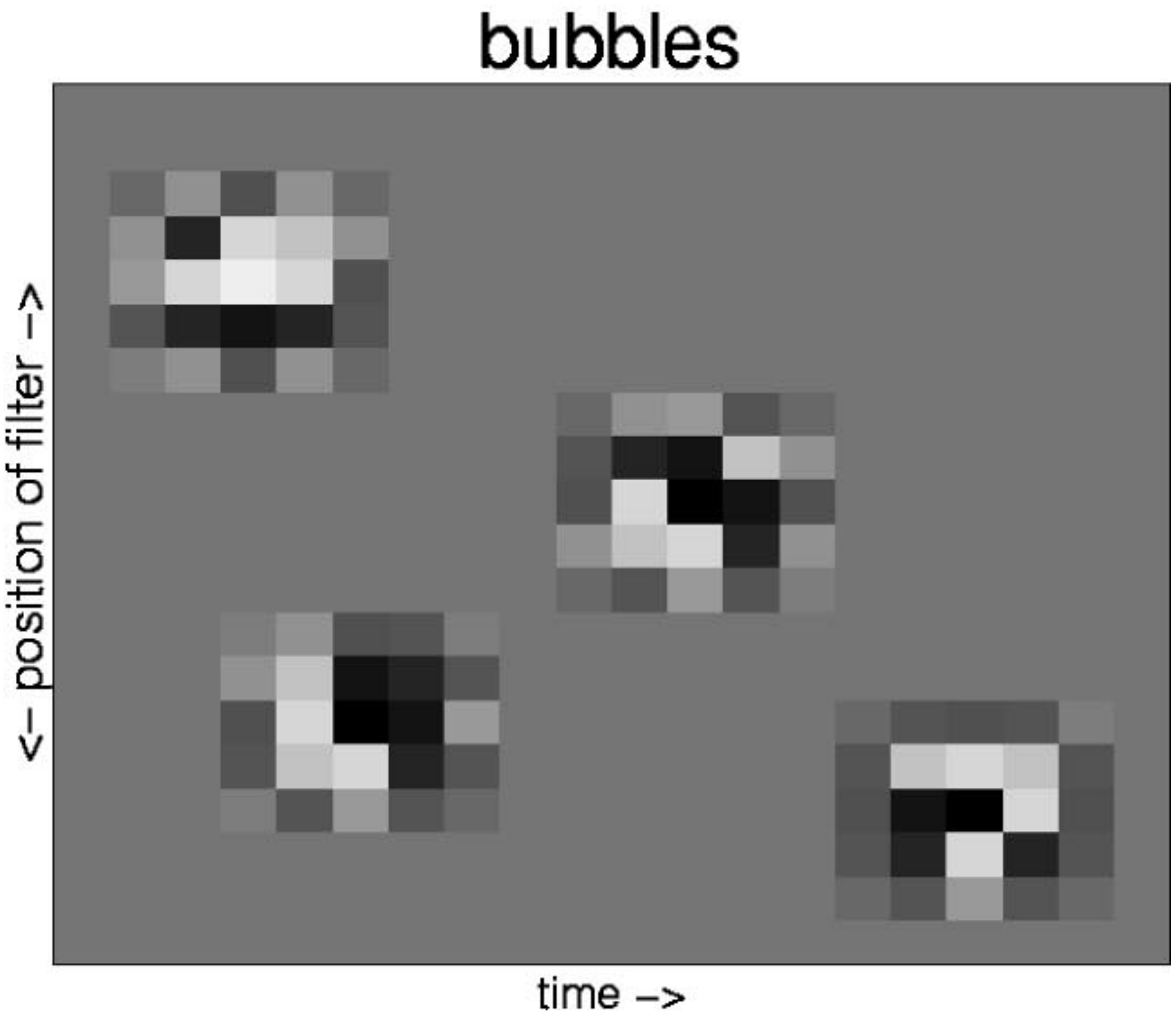
‘Topographic ICA’
(Hyvärinen & Hoyer 2001)



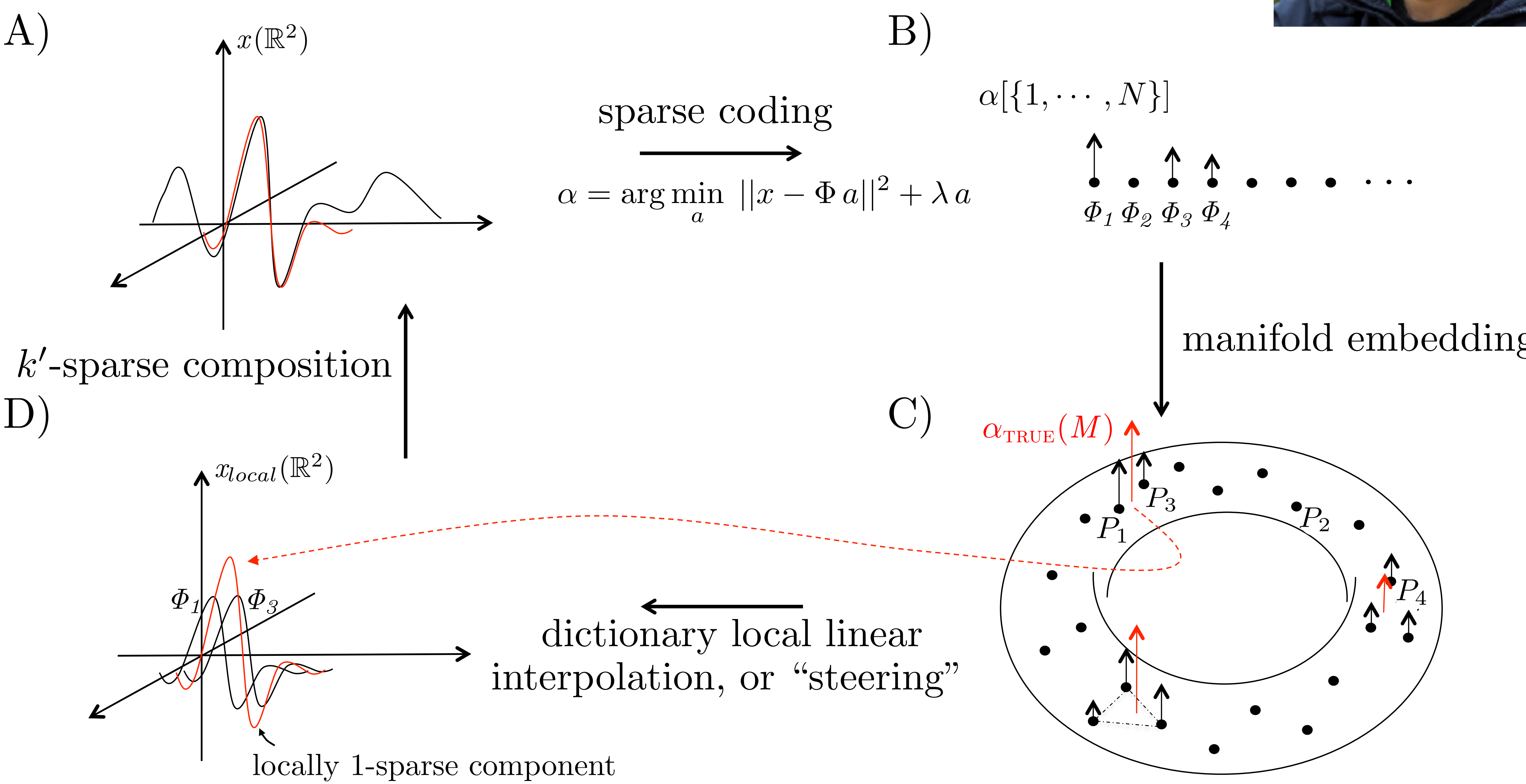
Bubbles: a unifying framework for low-level statistical properties of natural image sequences

Aapo Hyvärinen, Jarmo Hurri, and Jaakko Väyrynen

Neural Networks Research Centre, Helsinki University of Technology, P.O. Box 9800, FIN-02015 HUT, Finland

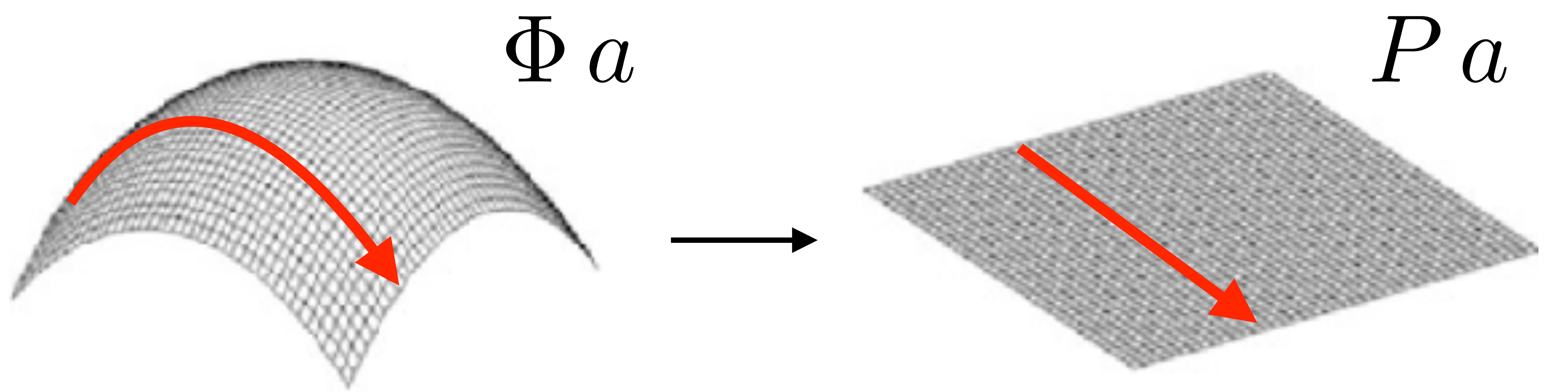


Sparse Manifold Transform
(Yubei Chen, Ph.D. thesis)



Persistence

We seek a geometric mapping $f : \Phi \rightarrow P$, s.t. each of the dictionary elements is mapped to a new vector, $P_j = f(\Phi_j)$, where P_j is the j[th] column of *P*. Continuous temporal transformations in the input should have a linear flow on *M* and also in the geometrical embedding space.



We desire: $$P a_t \approx \frac{1}{2} P a_{t-1} + \frac{1}{2} P a_{t+1}$$

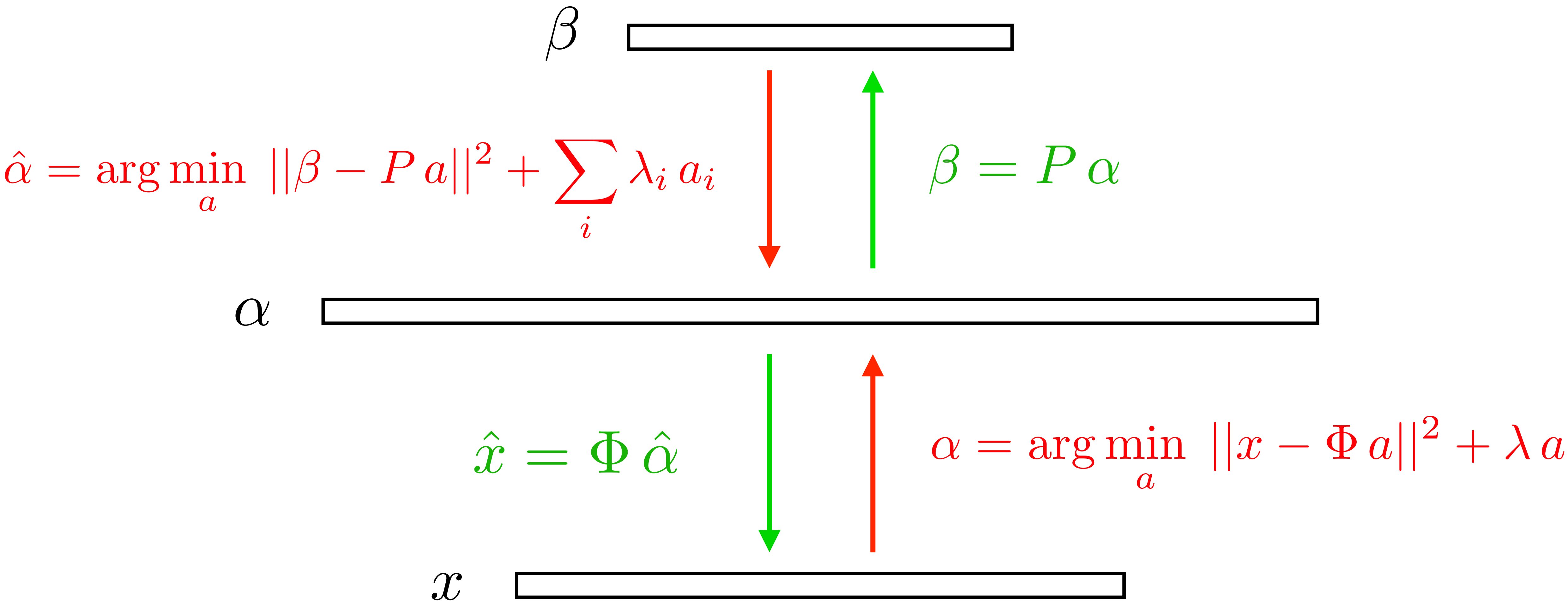
Objective function: $$\min_P \|P\,A\,D\|_F^2$$

s.t. $$P\,V\,P^T = I$$

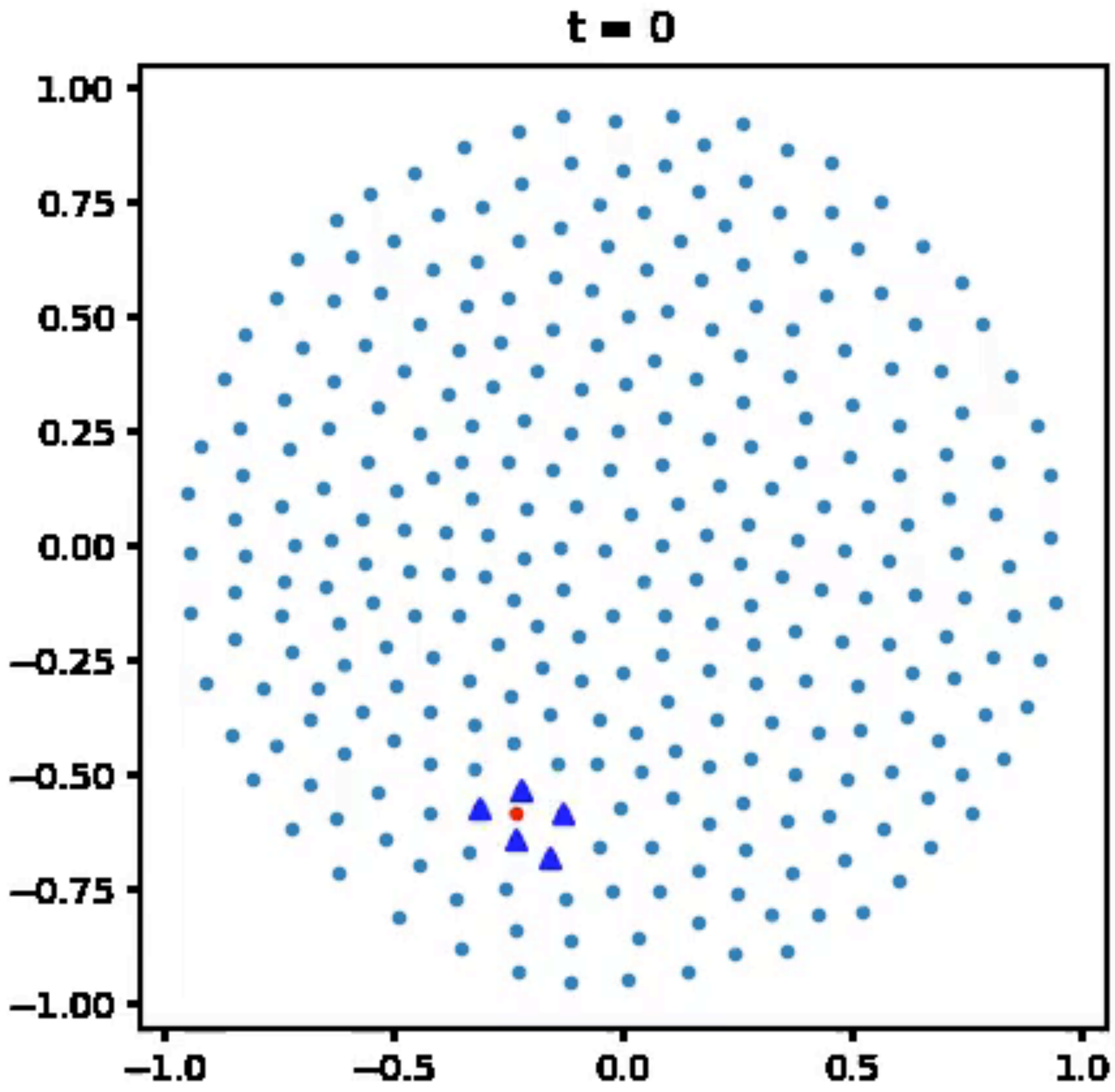
$$V = \mathrm{Cov}(a)$$

$$D = \begin{bmatrix} 1 & -\frac{1}{2} & 0 & 0 & \dots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \dots & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \dots & 0 \\ 0 & 0 & -\frac{1}{2} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix}$$

The sparse manifold transform

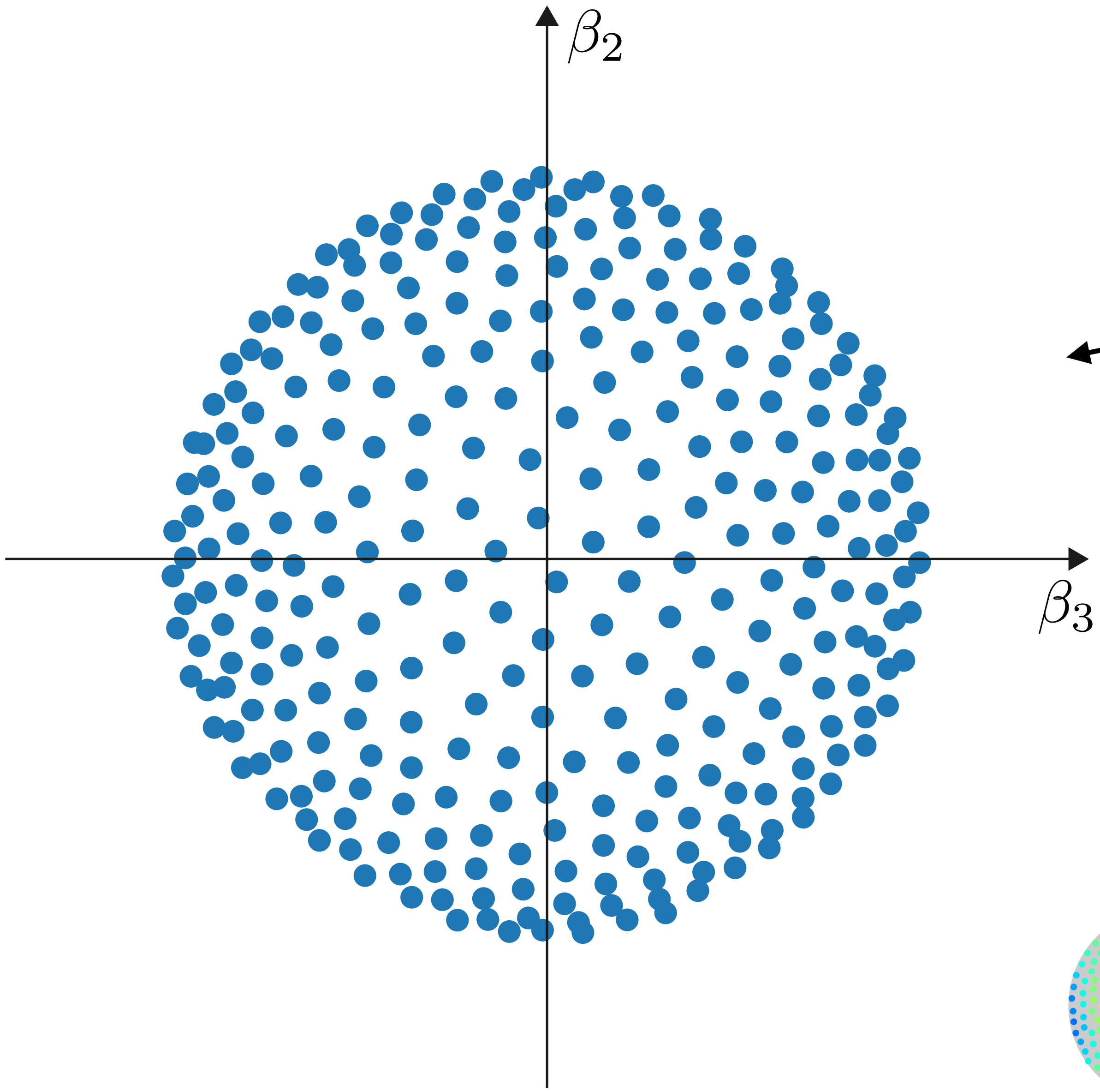


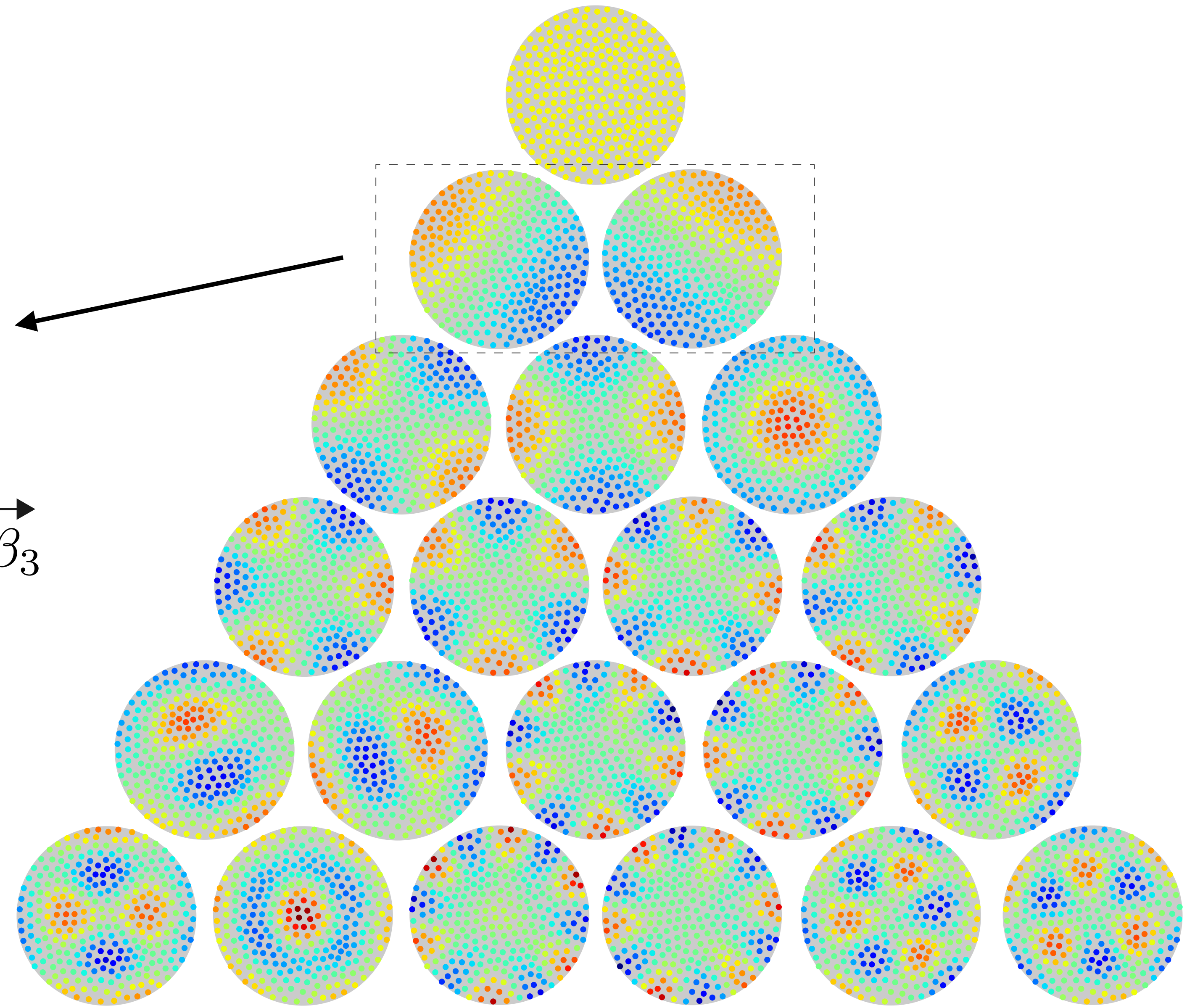
Simple example



Simple example

learned embedding



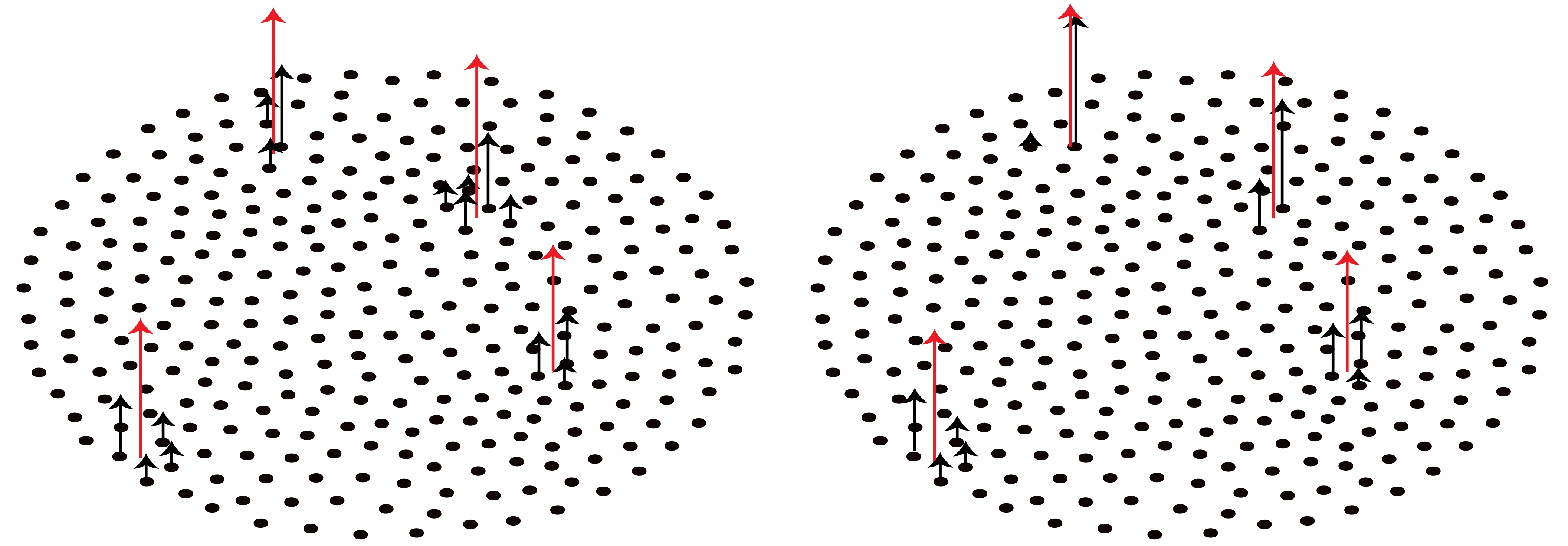


rows of P

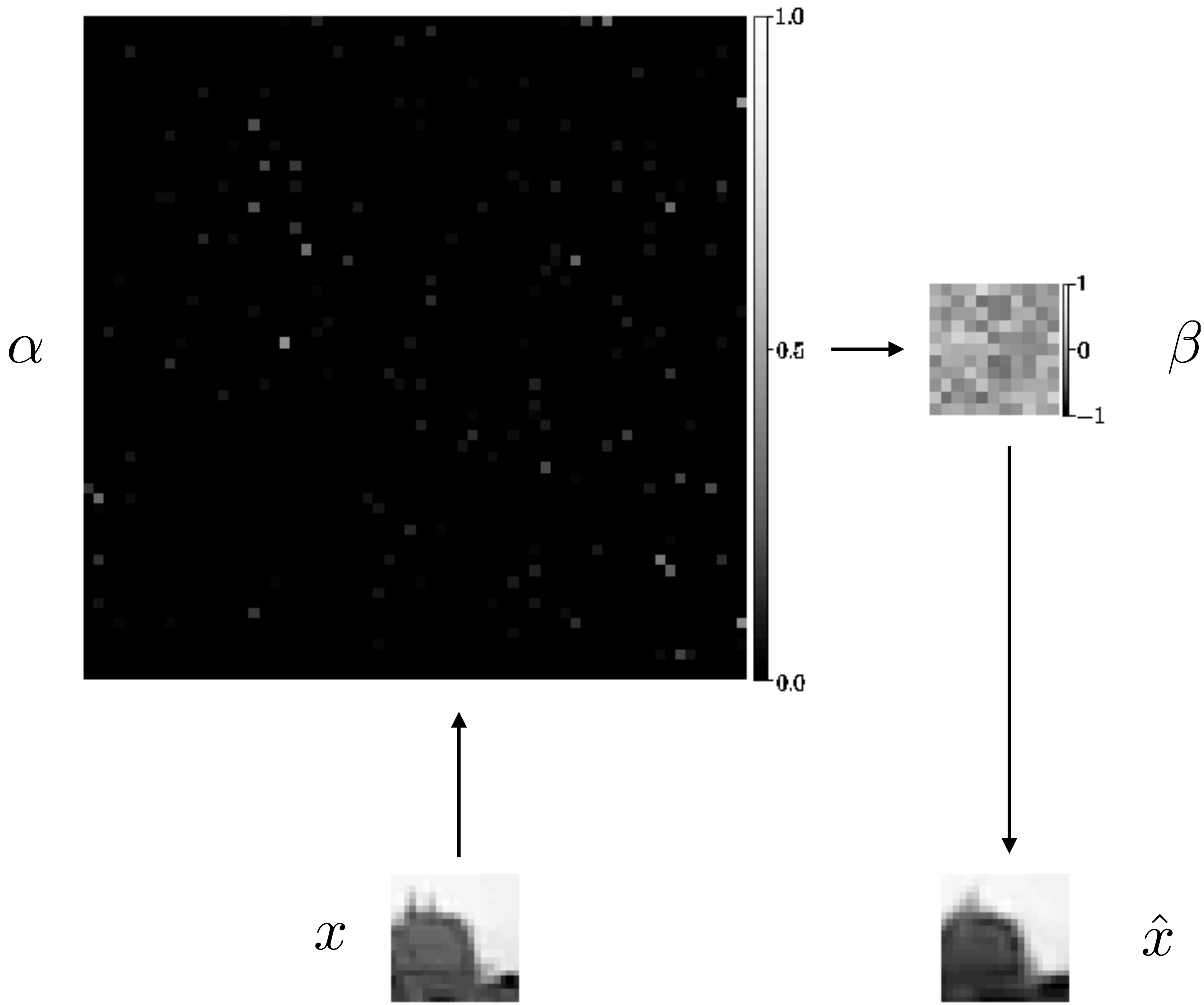
Simple example

k-sparse function embedding and recovery

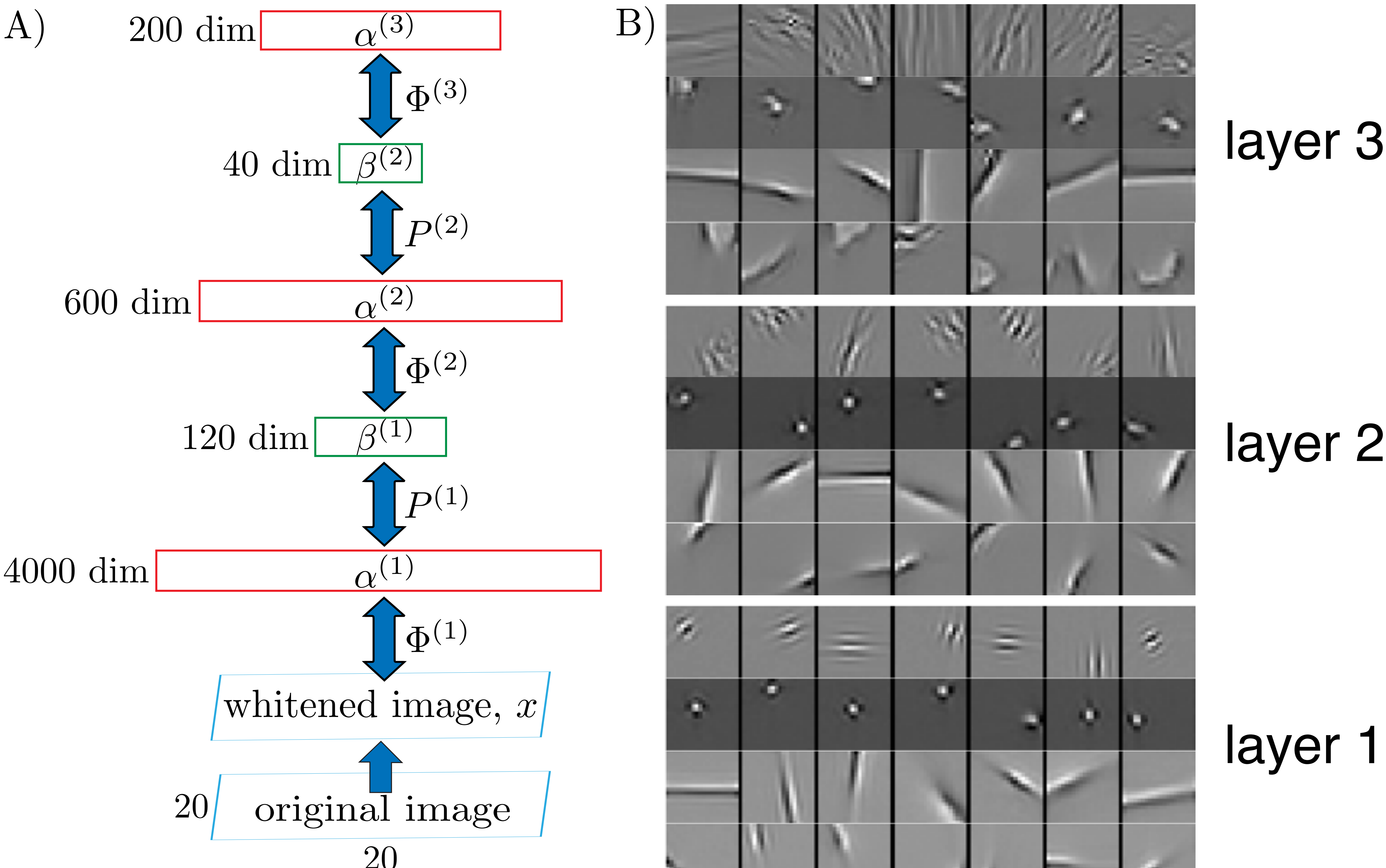
4-sparse function ⟶ β ⟶ recovery



Encoding of a natural video sequence

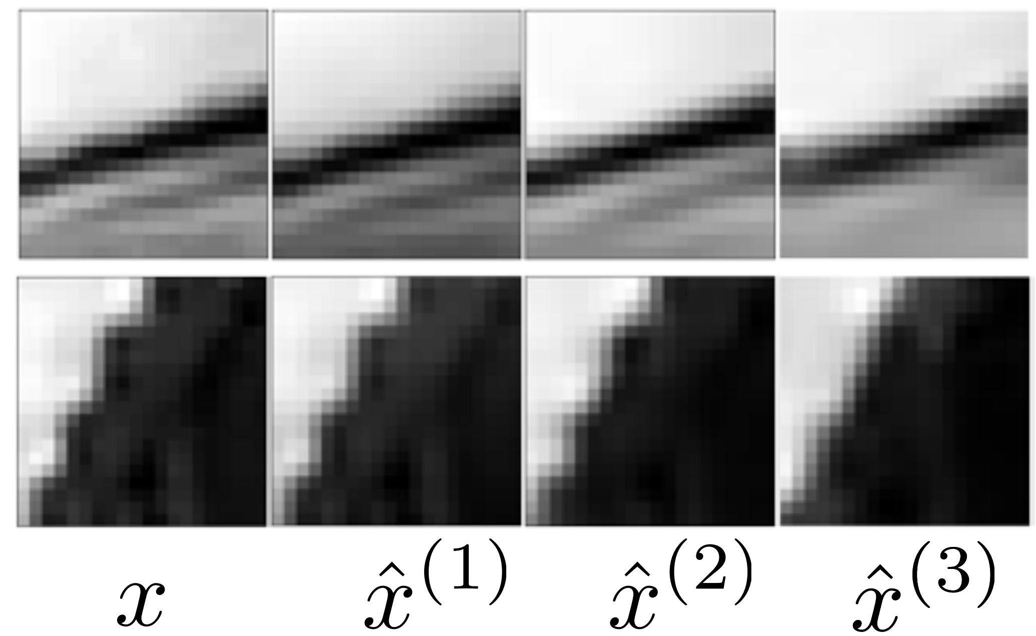


Stacked Sparse Manifold Transform

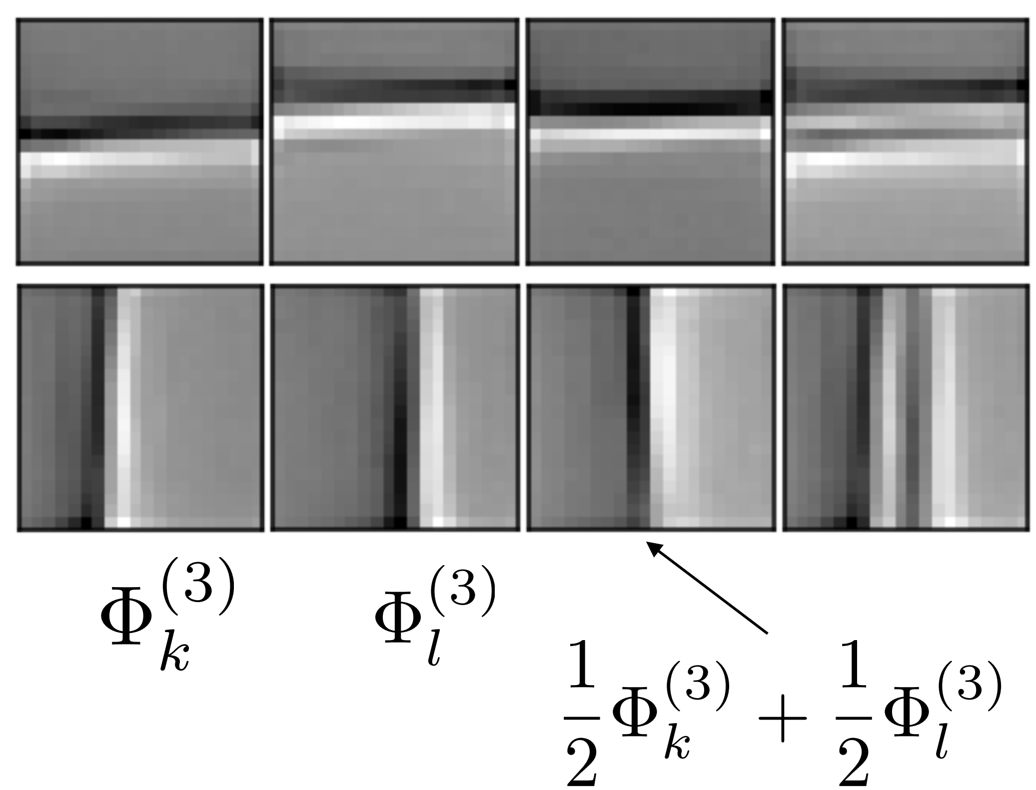


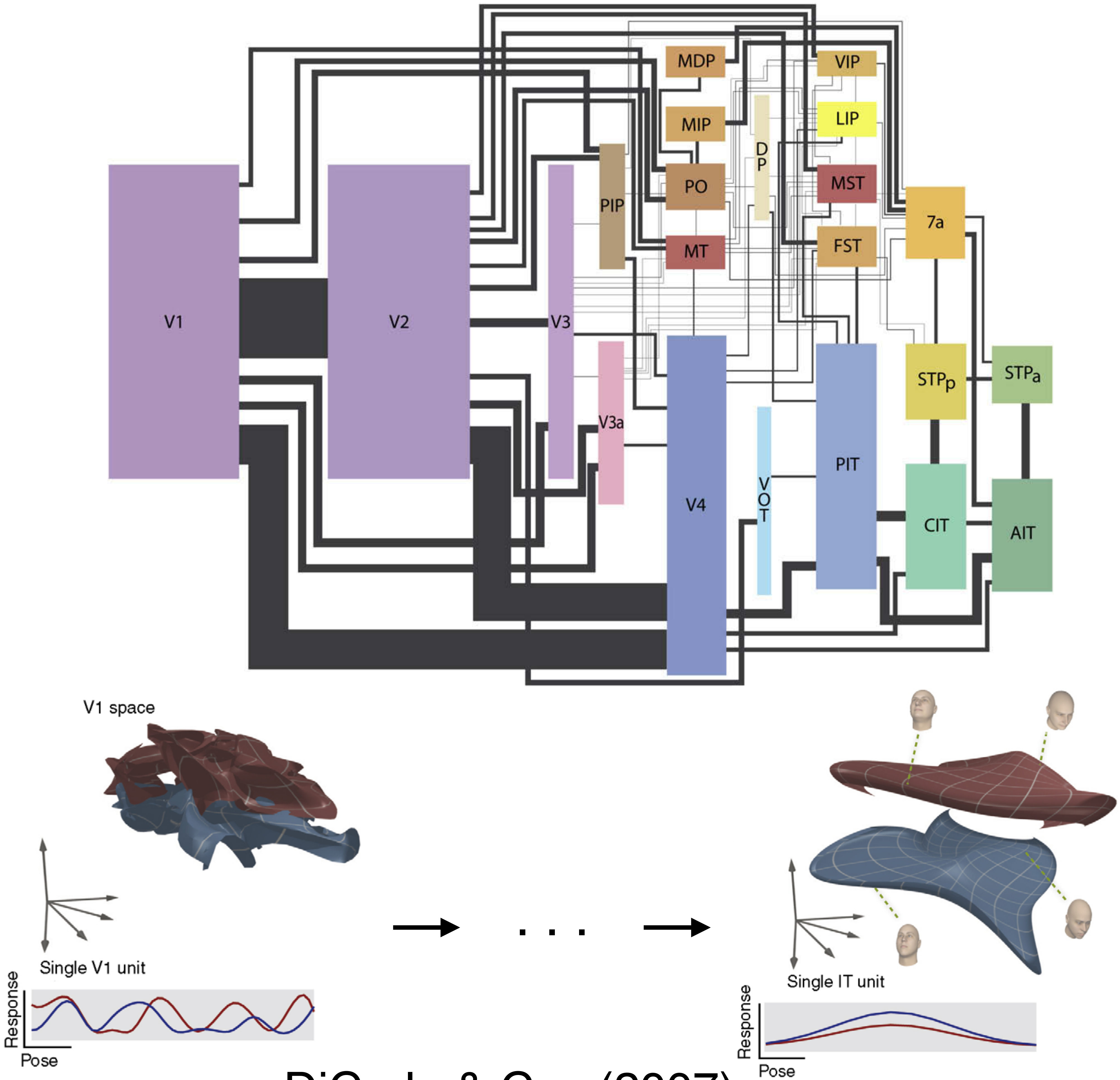
Stacked Sparse Manifold Transform

reconstruction from layers 1,2,3



'flattening' at layer 3





DiCarlo & Cox (2007)

Main points

- Sparse coding provides a foundation for manifold learning.

- A geometrical embedding of the dictionary may be learned by exploiting temporal persistence of structure in the visual world.

- Manifold flattening may be accomplished in a progressive manner by successive stages of sparse coding (dimensionality expansion) and linear projection (dimensionality collapse) in the ventral stream of visual cortex.