视觉神经编码与特征群组

张丽清

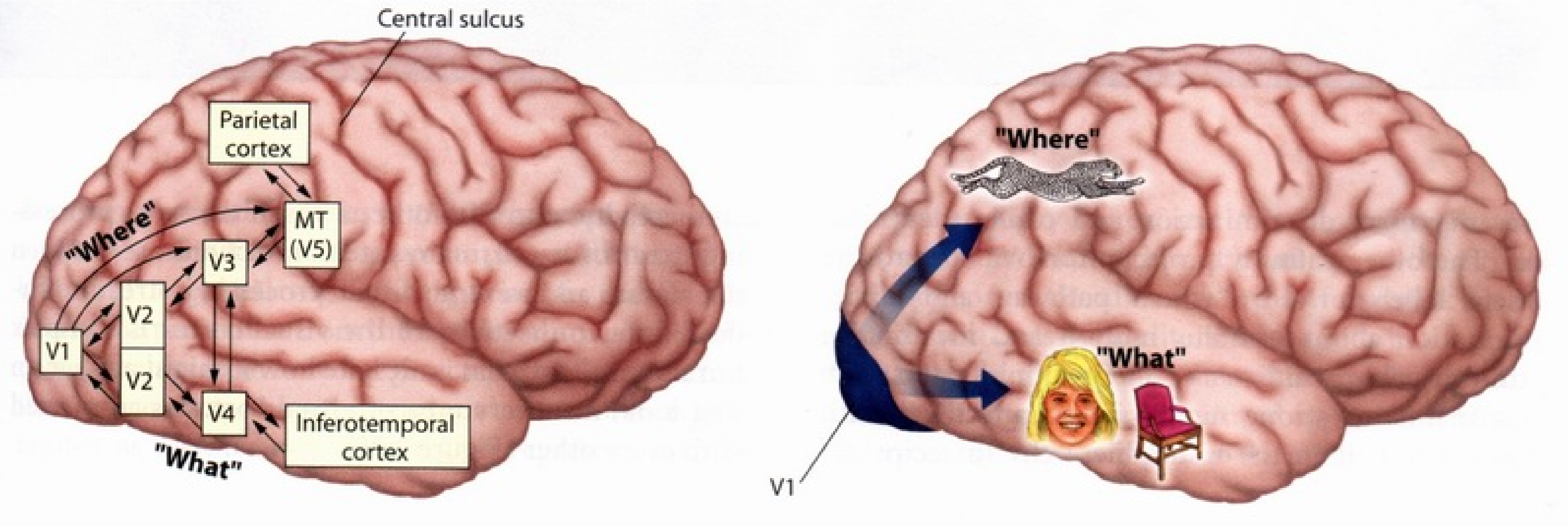
上海交通大学计算机系

感知计算实验室

目录

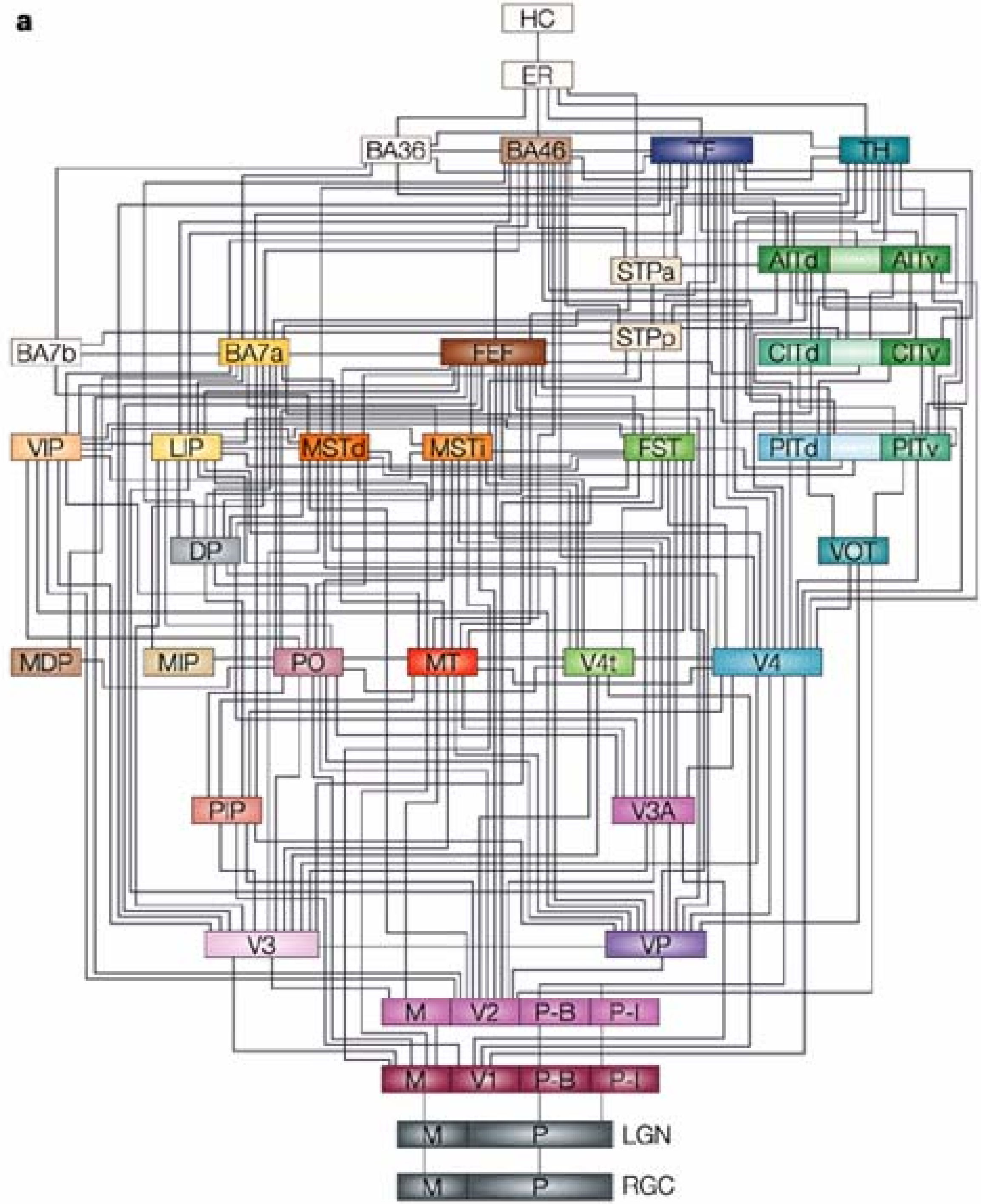
- 视觉编码生物学基础
- 神经编码模型
- 稀疏编码与算法
  - ICA算法
  - 稀疏表象
  - 非对称学习算法
- 视觉EEG信号处理
- 视觉编码的应用

Visual Information Pathways



'Where': the motion and spatial location

'What': the detailed features, form, and object identity

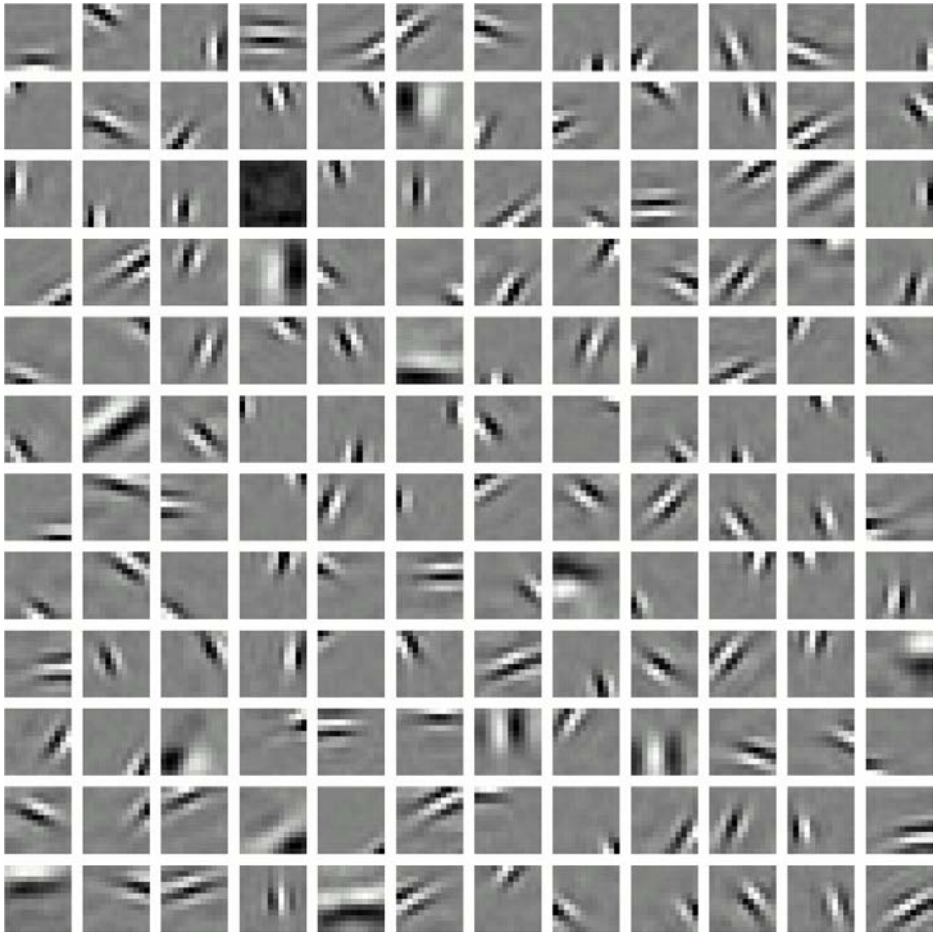

a
HC
ER
BA36
BA46
TF
TH
STPa
AITd
AITv
STPp
BA7b
BA7a
FEF
CITd
CITv
VIP
LIP
MSTd
MSTl
FST
PITd
PITv
DP
VOT
MDP
MIP
PO
MT
V4t
V4
PIP
V3A
V3
VP
M V2 P-B P-I
M V1 P-B P-I
M P LGN
M P RGC

视觉关联区结构示意

Coding Principle

- Efficient Coding-- Attneave (1954)
  - The goal of visual perception is to produce an efficient representation of the input signal.
- Redundancy Reduction -- Barlow (1961)
  - The role of early sensory neurons is to remove statistical redundancy in the sensory input.
- Sparse Coding – Vinje, Gallant (Science 2000)
  - During natural vision, the classical and nonclassical receptive fields function together to form a sparse representation of the visual world.
- Temporal coding--
  - The temporal structure of the firing patterns of single neurons in multiple neurons of a functional assembly

Sparse Coding

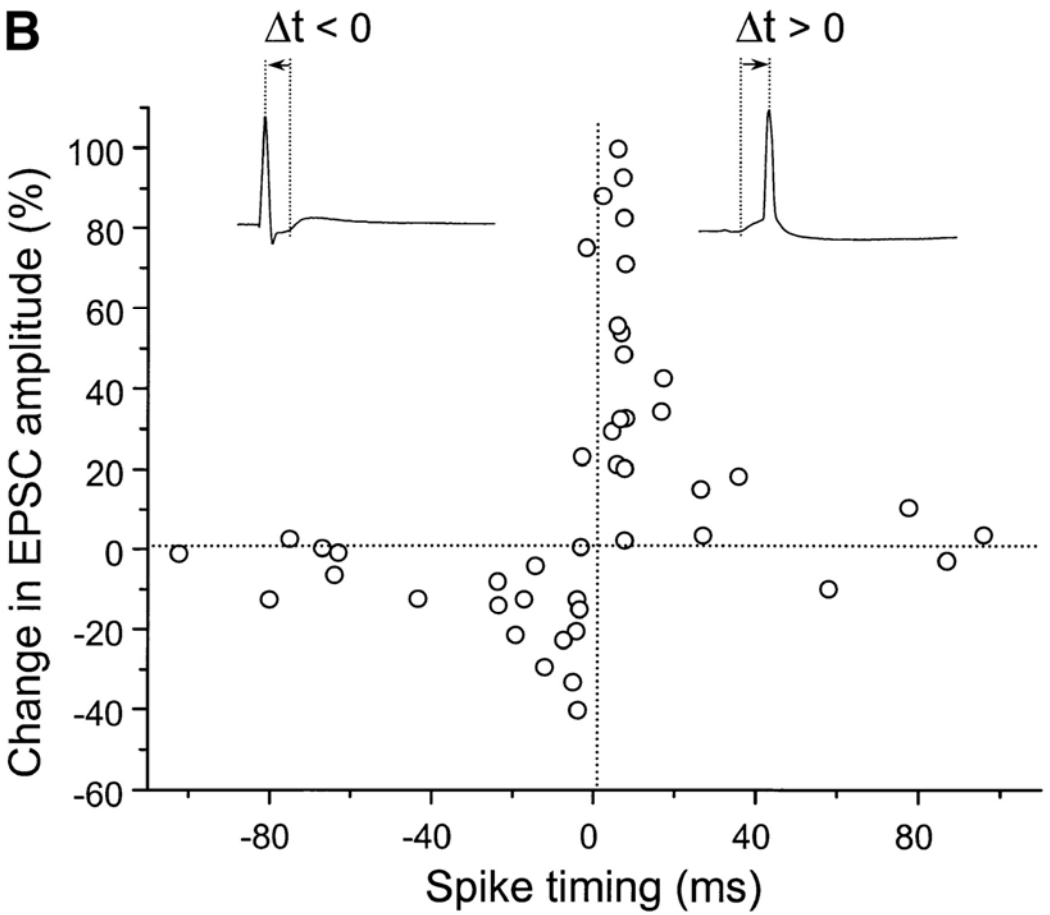


Olshausen and Field, 1996

Receptive field model for

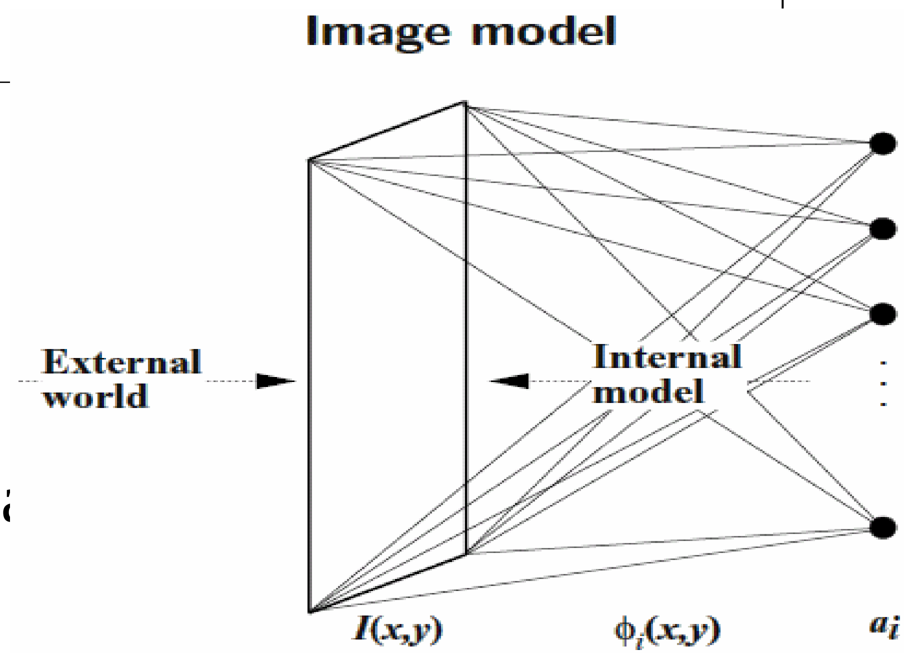
Simple cells in V1

LTP/LTD（非对称学习）



- Figure from [Bi & Poo (1998)]

Sparse Coding

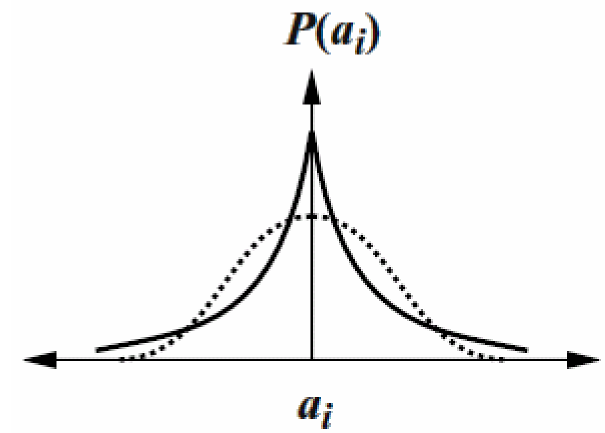


Objective:

To find basis functions $\{\varphi_i\}$, such that natural images can be represented a sparse as possible.

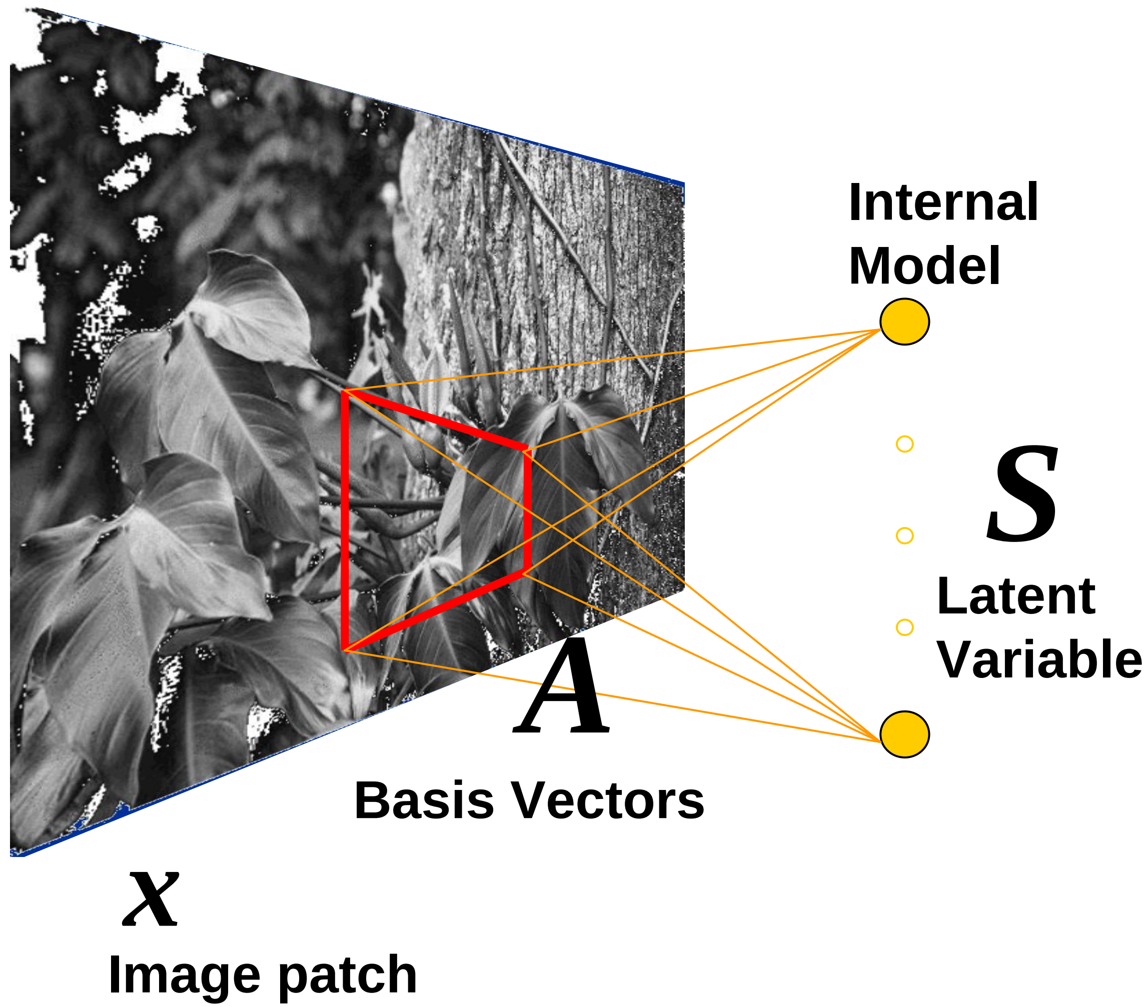
$$P(\mathbf{a}|\theta) = \Pi_i P(a_i|\theta)$$

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



Natural Scene Analysis

External World



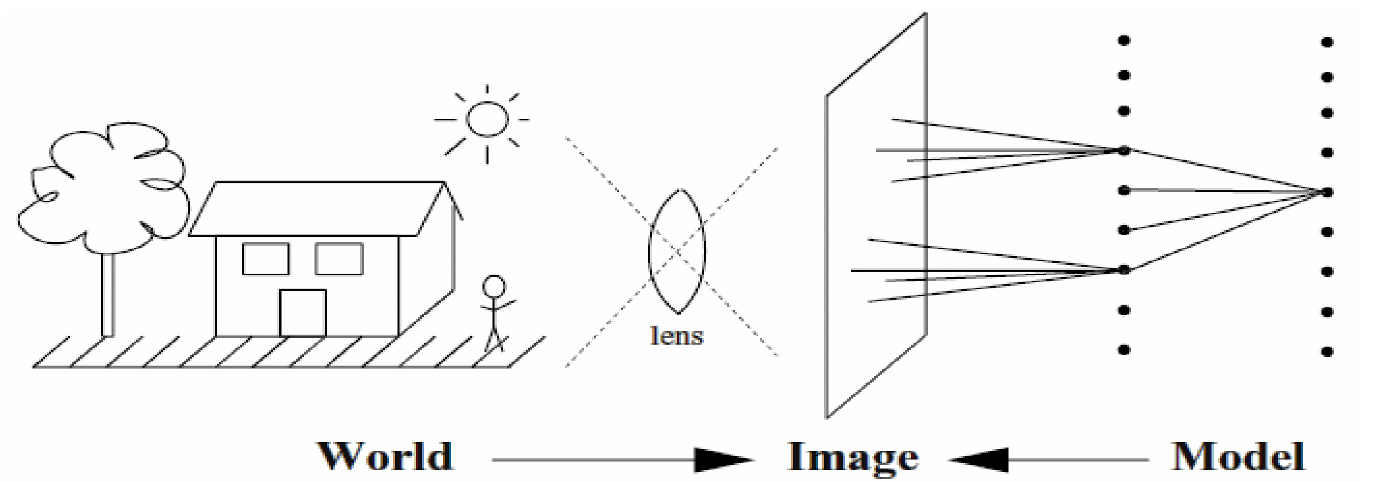
$$x(t) = As(t)$$

$$y(t) = Wx(t)$$

Redundancy Reduction

(Barlow1961; Attneave1954)

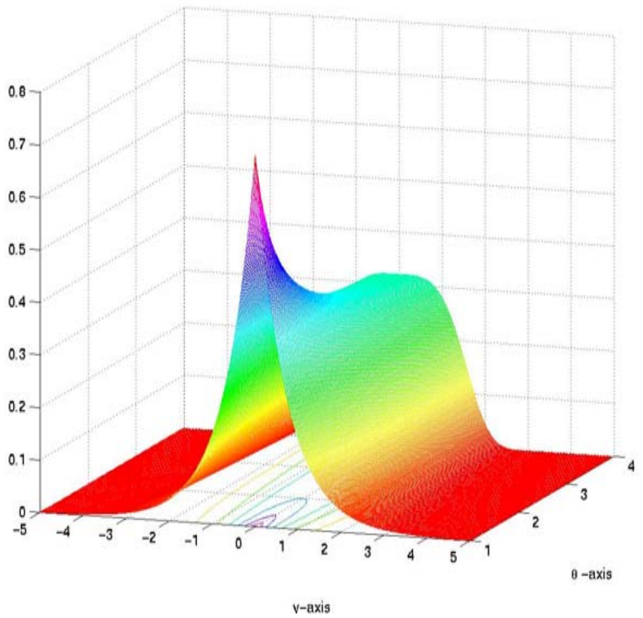
- Natural images are redundant in that there exist statistical dependencies among pixel values in space and time.
- In order to make efficient use of resources, the visual system should reduce redundancy by removing statistical dependencies.



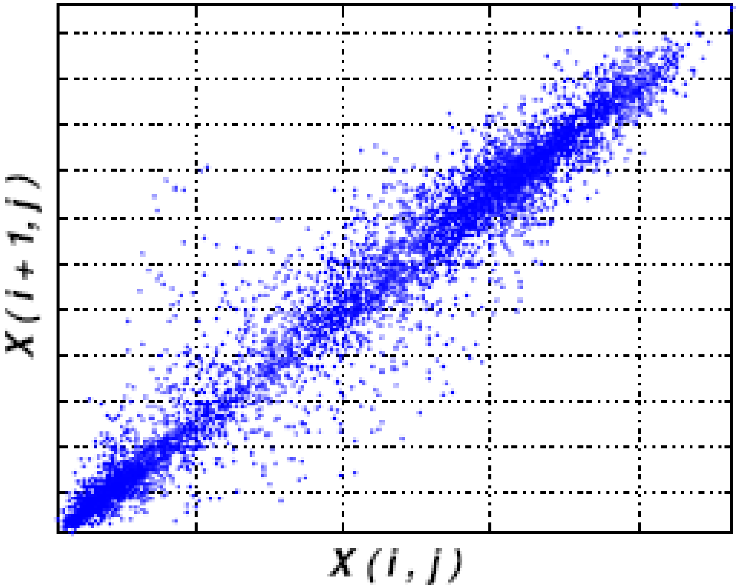
Sparse Representation

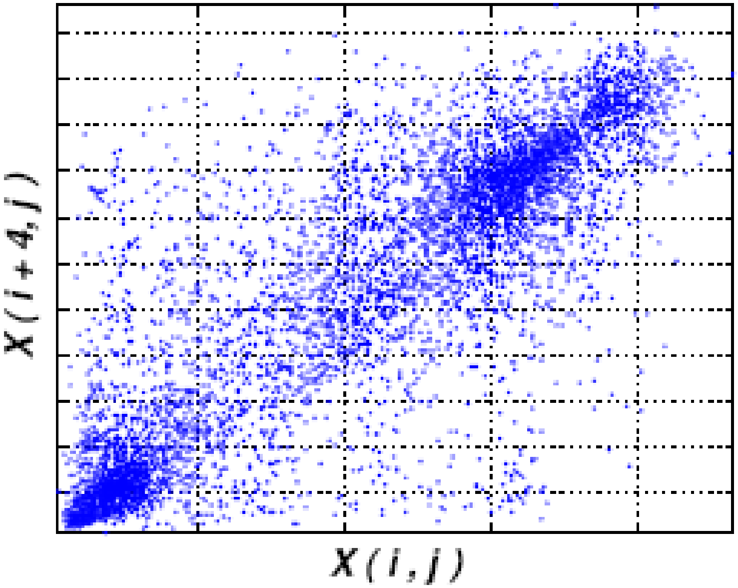
$$p_G(y,\theta)=\frac{1}{2\sigma\Gamma(1/\theta)}\exp\left(-\left|\frac{y}{\sigma}\right|^{\theta}\right)$$

$$\begin{cases} Sparse,\quad when\ \ \theta<2; \\ Gaussian,\quad when\ \ \theta=2; \\ Nonsparse,\quad when\ \ \theta>2. \end{cases}$$



Correlation Between Pixels





The Cost Function

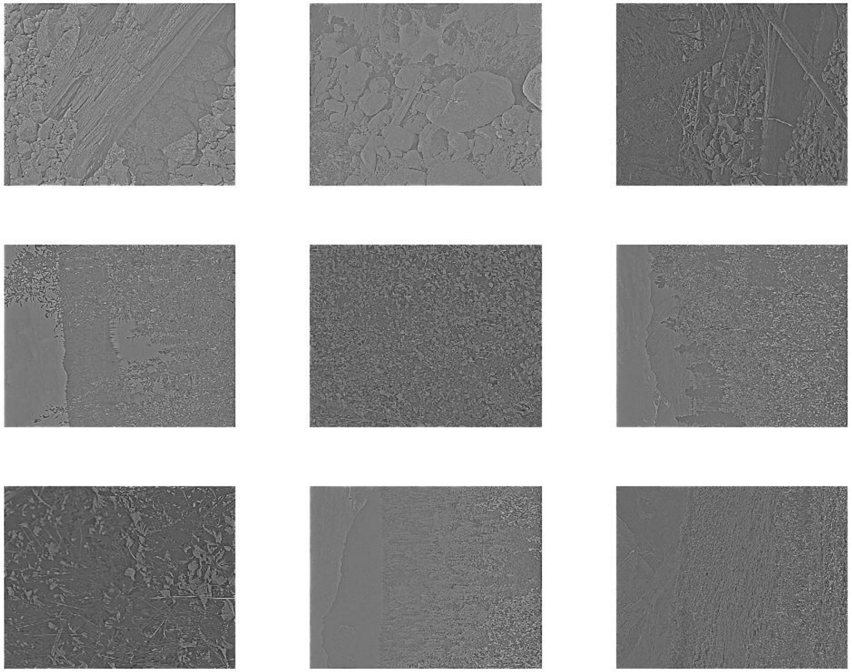
$$E(a,\phi) = \sum_{x,y}[I(x,y) - \sum_i a_i\phi_i(x,y)]^2 + \beta\sum_i S(\frac{a_i}{\sigma_i})$$

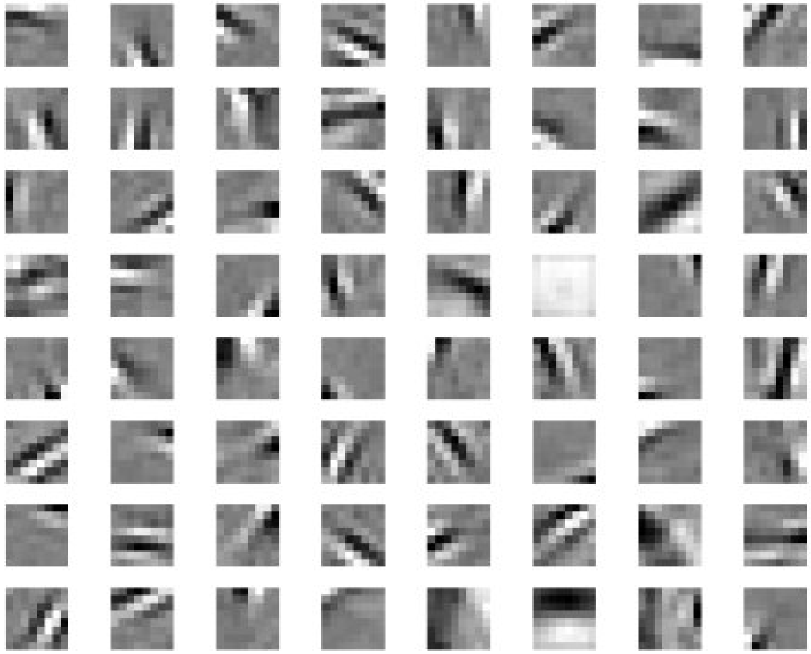
$$S(x): \qquad \log(1+x^2); \quad |\mathrm{x}|^p (0 < p < 2)$$

where I is input; and phi(x,y) is the internal representation

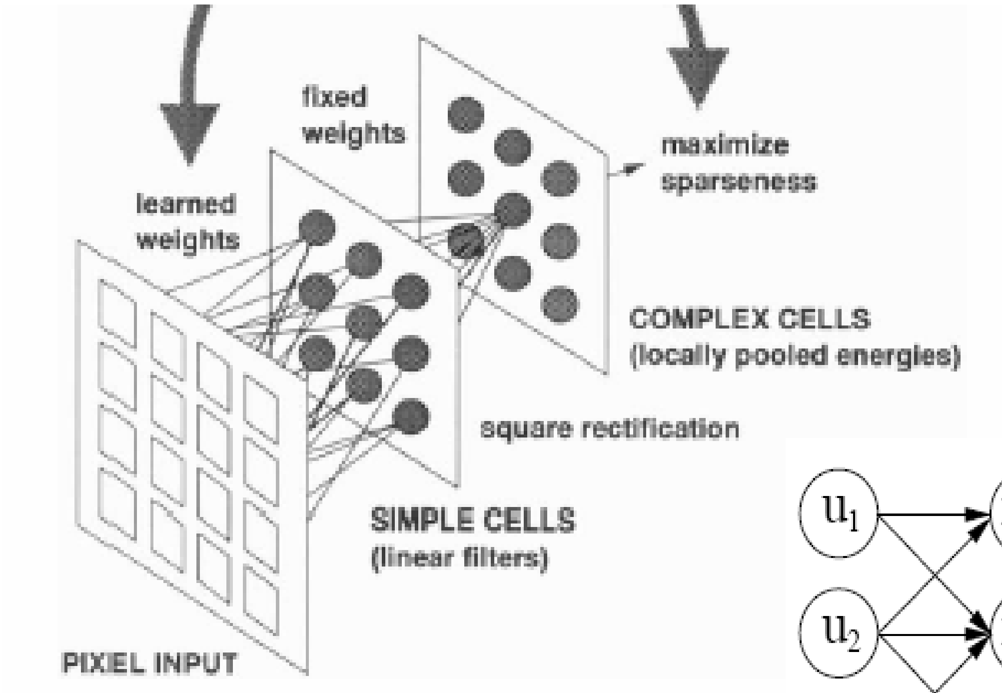
Training with Natural Images

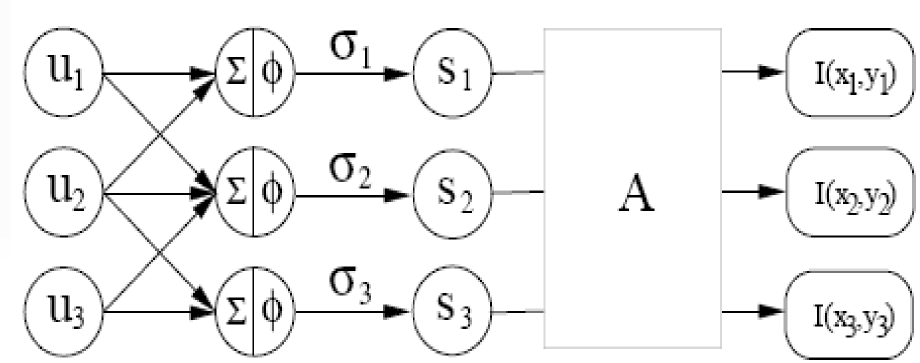
- ❑ Training: 10 images (512x512)
- ❑ 10,000 presentations
- ❑ Batch size: 100
- ❑ Basis Function: 16x16



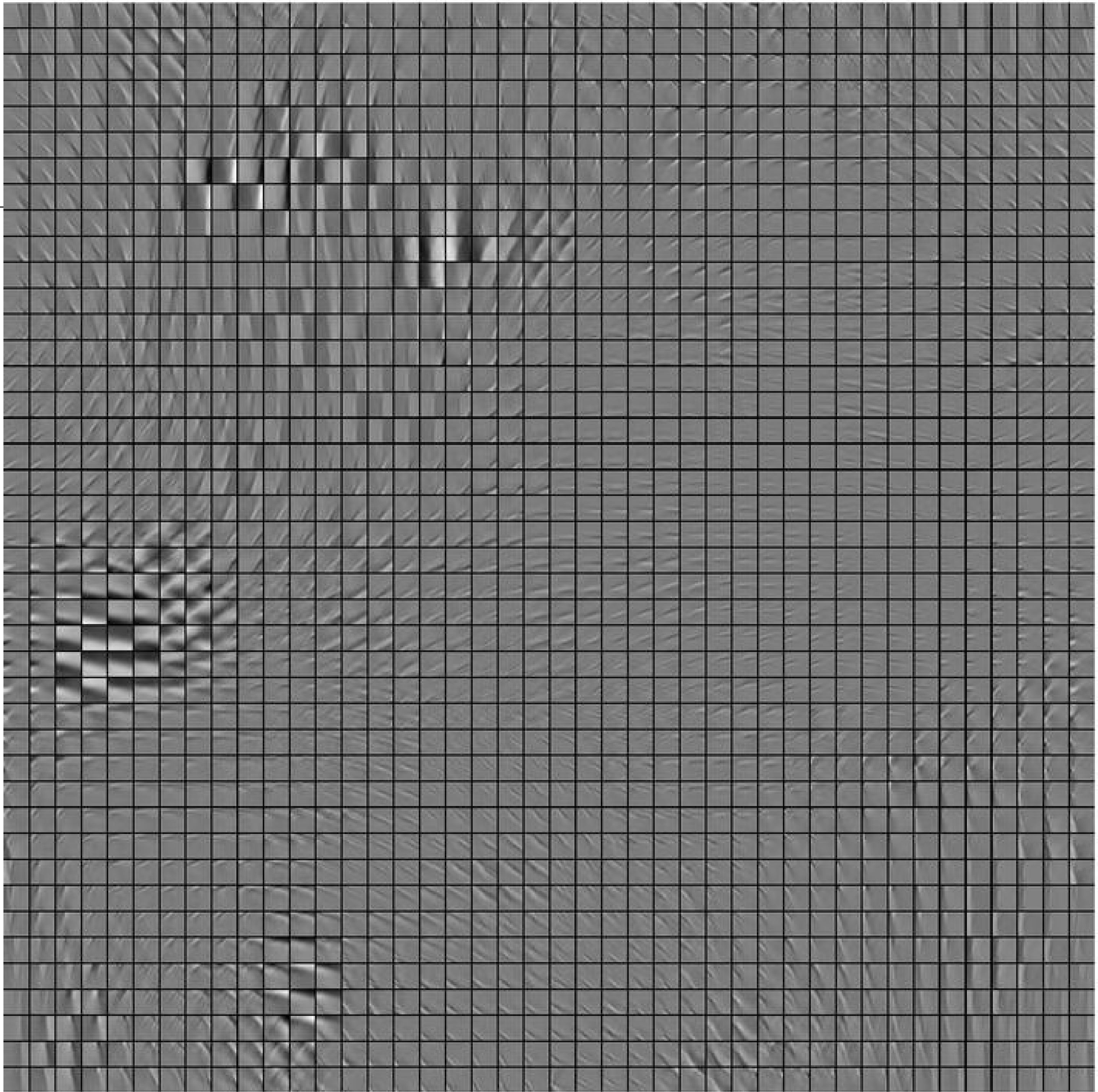


Topographic ICA complex cells





- Hyvarinen and Hoyer，2001



Independent Coding Hypothesis

Each cortical neuron represents a separated signal.

Although many neurons may be involved in coding a particular object, the essential characteristic of the independent coding is that all of the information carried by one neuron is independent of that by any other neurons.

- Retinal ganglion cells; S. Nirenberg, et al, 2001
- Neurons selective for direction; deCharms, 1998
- Neurons for determining a motor task, Georgopous, 1990; Schwartz, 1994

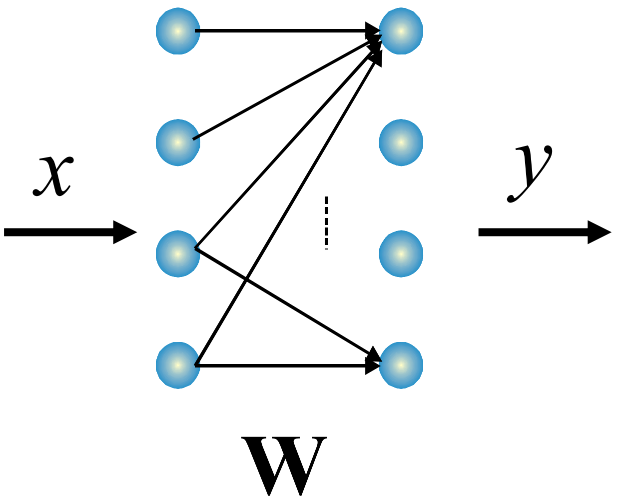
Independent Decomposition

The purpose of Independent Decomposition is to decompose the sensor signals into the components which are maximally mutually independent.

Problem:

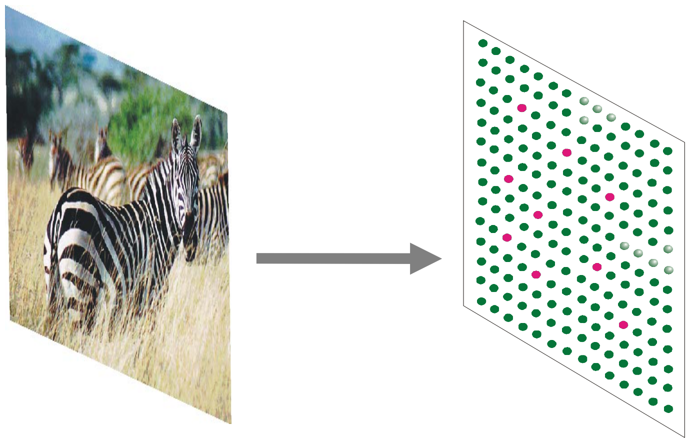
$$\min I(y)$$

where
$$x = (x_1, \cdots, x_n)^T$$
$$y = (y_1, \cdots, y_m)^T$$
$$y = \mathbf{W}x$$

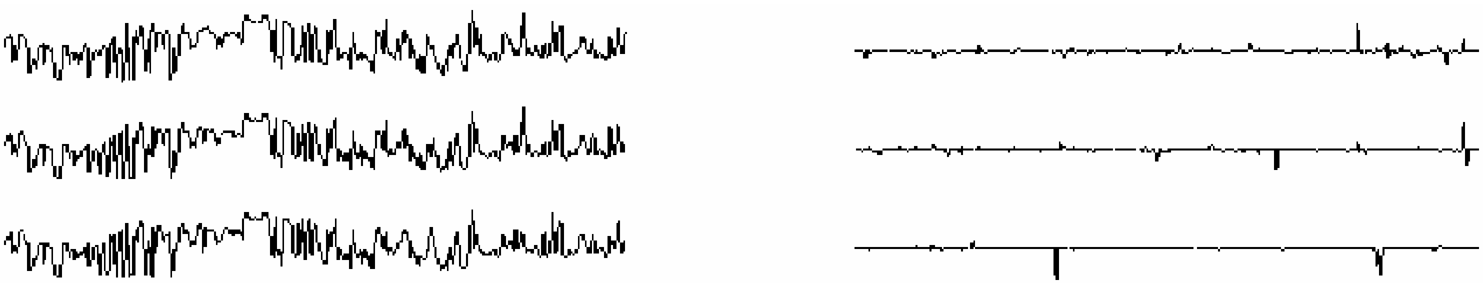


Sparseness I

- Spatial Sparesness
  - Small rate of neurons response to a stimulus



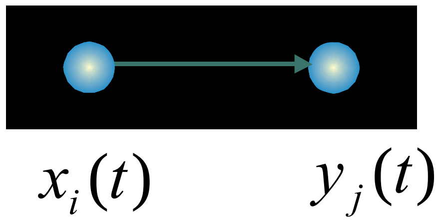
- Temporal Sparseness



Hebbian vs. ICA Learning

- Hebbian Learning rule:

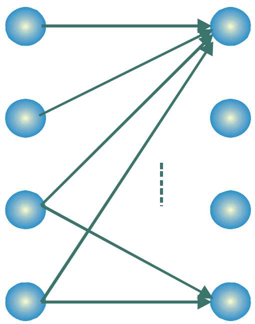
$$\Delta w_{ij}(t) = \eta\, x_i(t)\, y_j(t)$$



- ICA Learning rule:

$$\Delta W = \eta(I - \varphi(y)y^T)W; \quad y = Wx.$$

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$



$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{bmatrix}$$

Sparse Representation for Image

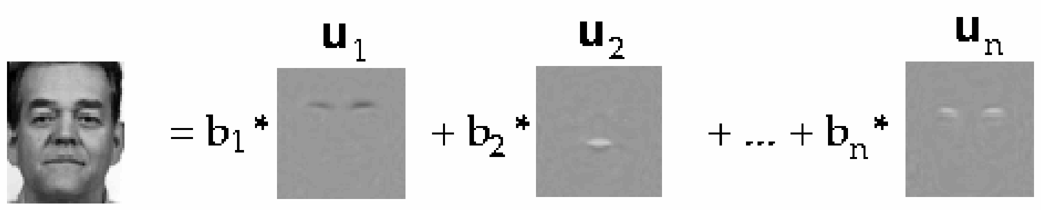
- Image model:
  - a linear superposition of basis functions and adapted these functions so as to maximize the sparsity of the representation while preserving information in the images.
- Probabilistic model
  - Olshausen & Field 1997;
  - Bell, et al 1997,
  - Hateren et al 1998;
  - Lewicki & Olshausen 1999.
  - Hyvärinen & Hoyer (2000) complex cell
  - Vinje & Gallant (2000)
  - Ma, et al 2006 (Overcomplete Sparse Representation)

Independent Image Basis

- Images are modeled as a linear superposition of basis functions $u_i(x,y)$

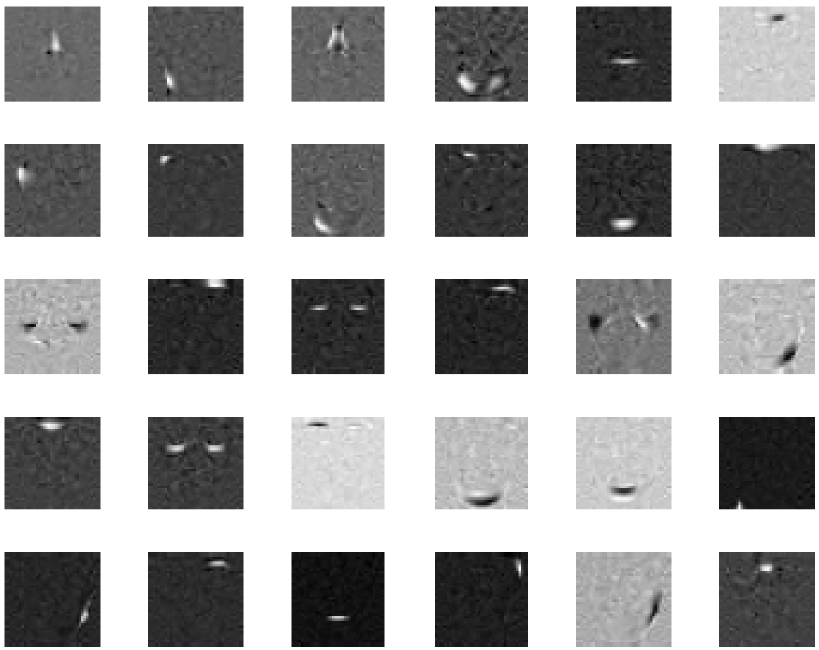
$$I(x,y)=\sum_i b_i u_i(x,y)+v(x,y)$$

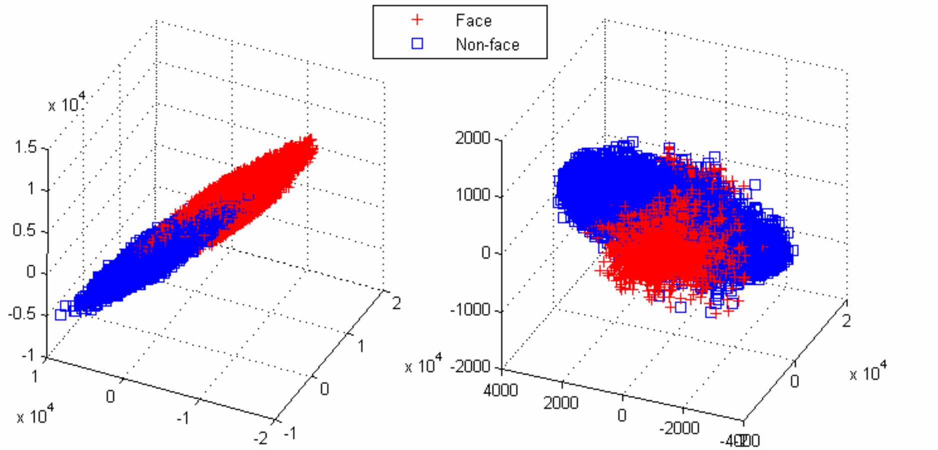
$$\text{ICA Repesentation}=(b_1,b_2,\cdots,b_n).$$



- (Field & Olshausen,1996; Bell & Sejnowski,1997)

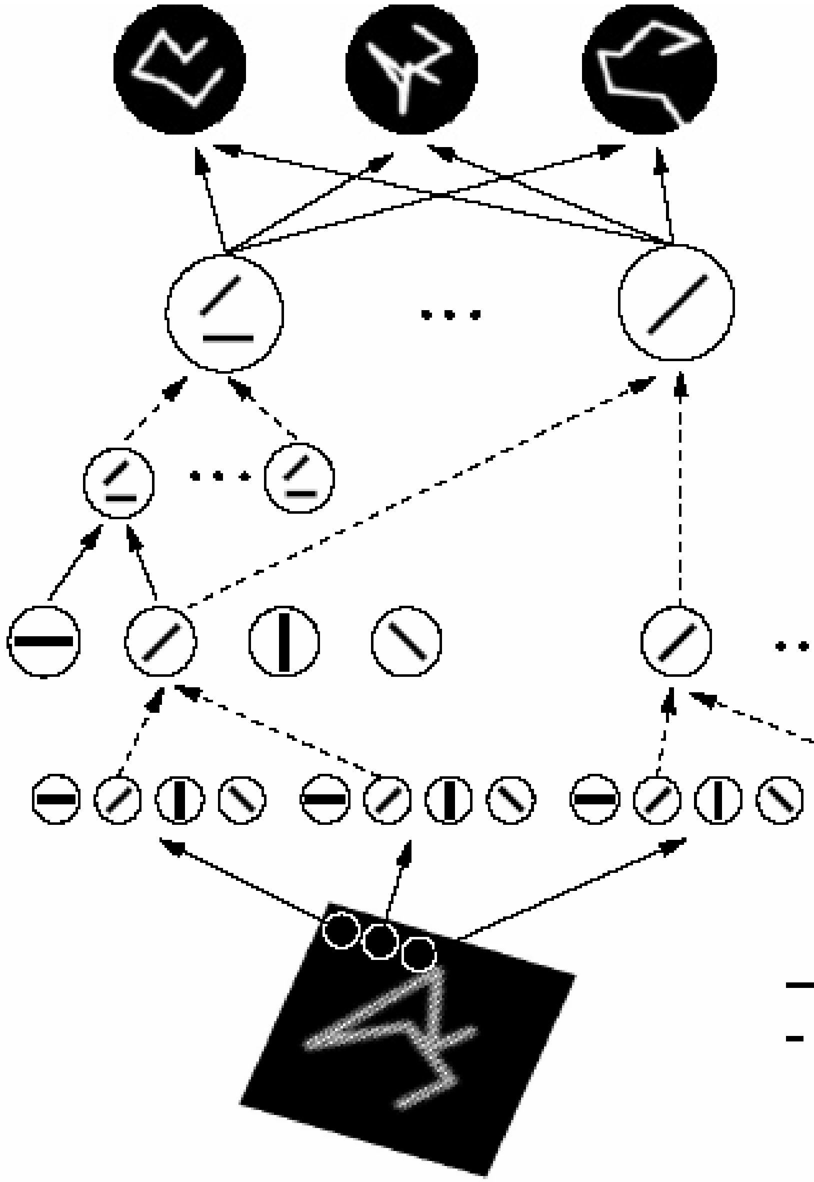
Face Local Features





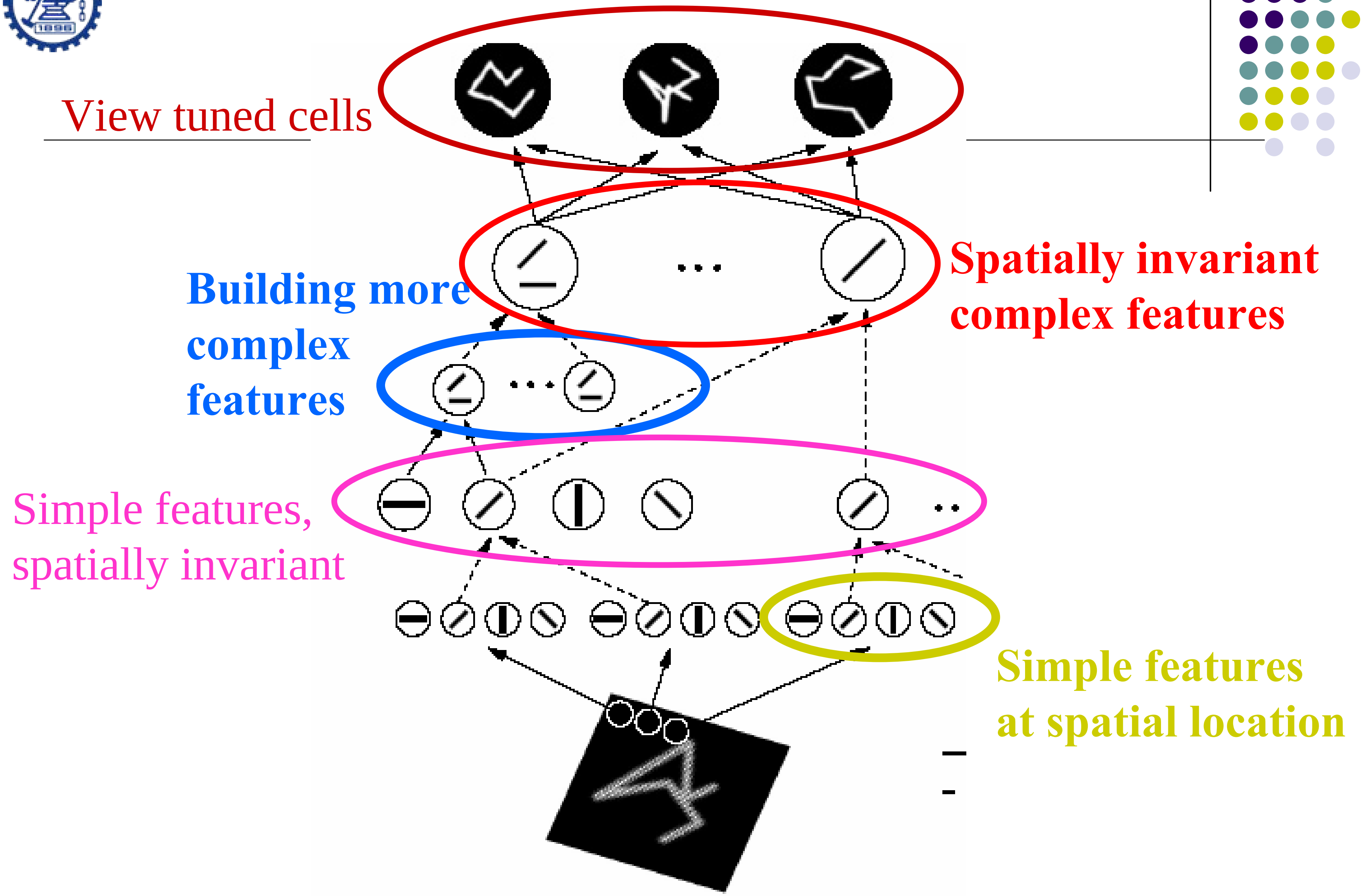
Zhao and Zhang, 2006

Feedforward Hierarchical Architecture



What pooling mechanism to to produce robust feature detectors?

Riesenhuber, etc, 2001


View tuned cells
Building more complex features
Spatially invariant complex features
Simple features, spatially invariant
Simple features at spatial location

Overcomplete Representation

- Statistical Model

$$p(A \mid z) = \frac{p(z \mid A)\, p(A)}{p(z)}$$

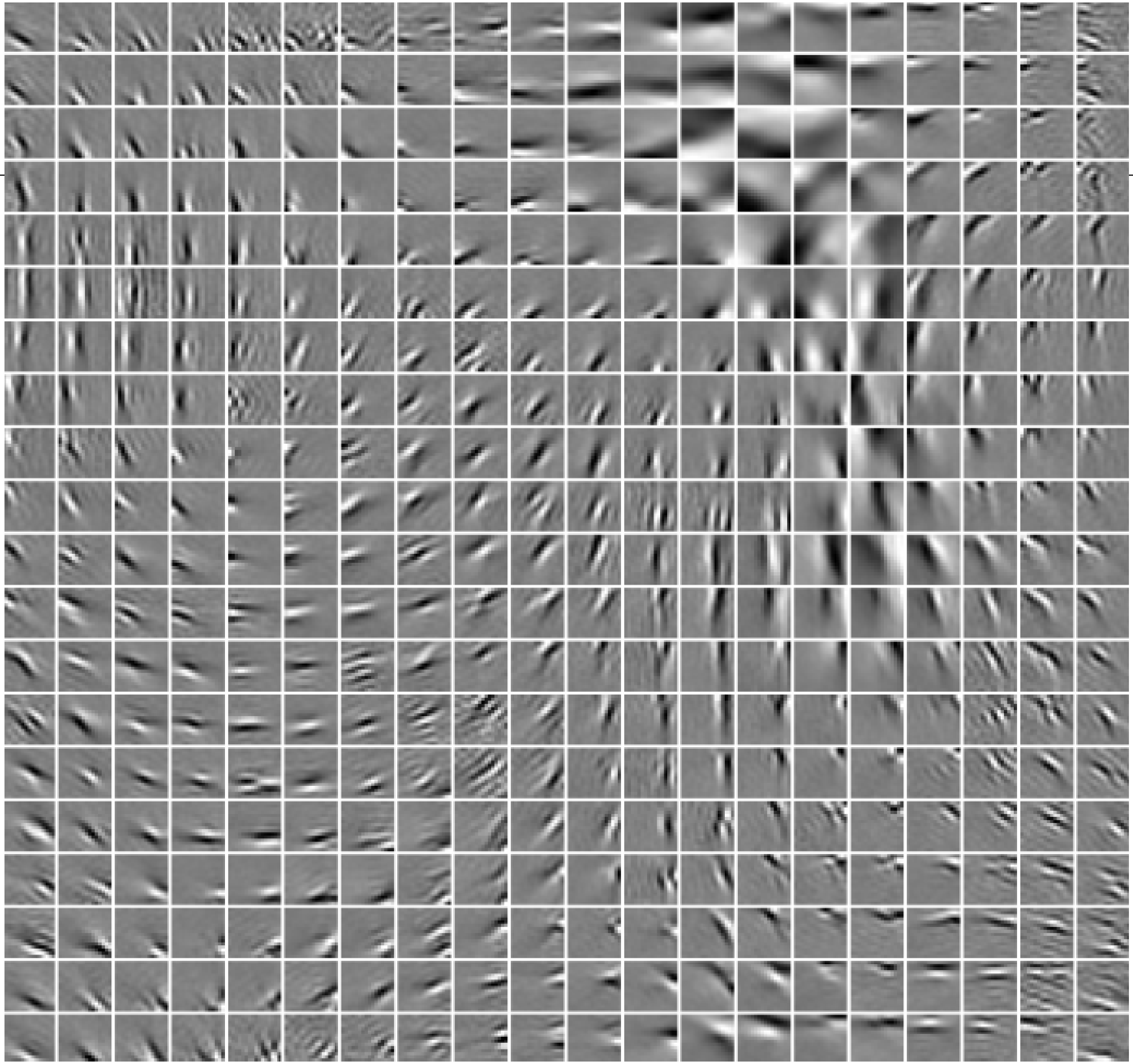
- Intermediate variables

$$y_i = a_i^T z = a_i^T A s = s_i + \sum_i a_i^T a_j s_j$$

- Priors

$$p(A) = c_m \prod_{i<j} (1 - (a_i^T a_j)^2)^{\frac{m-3}{2}}$$

Ma Libo's Poster

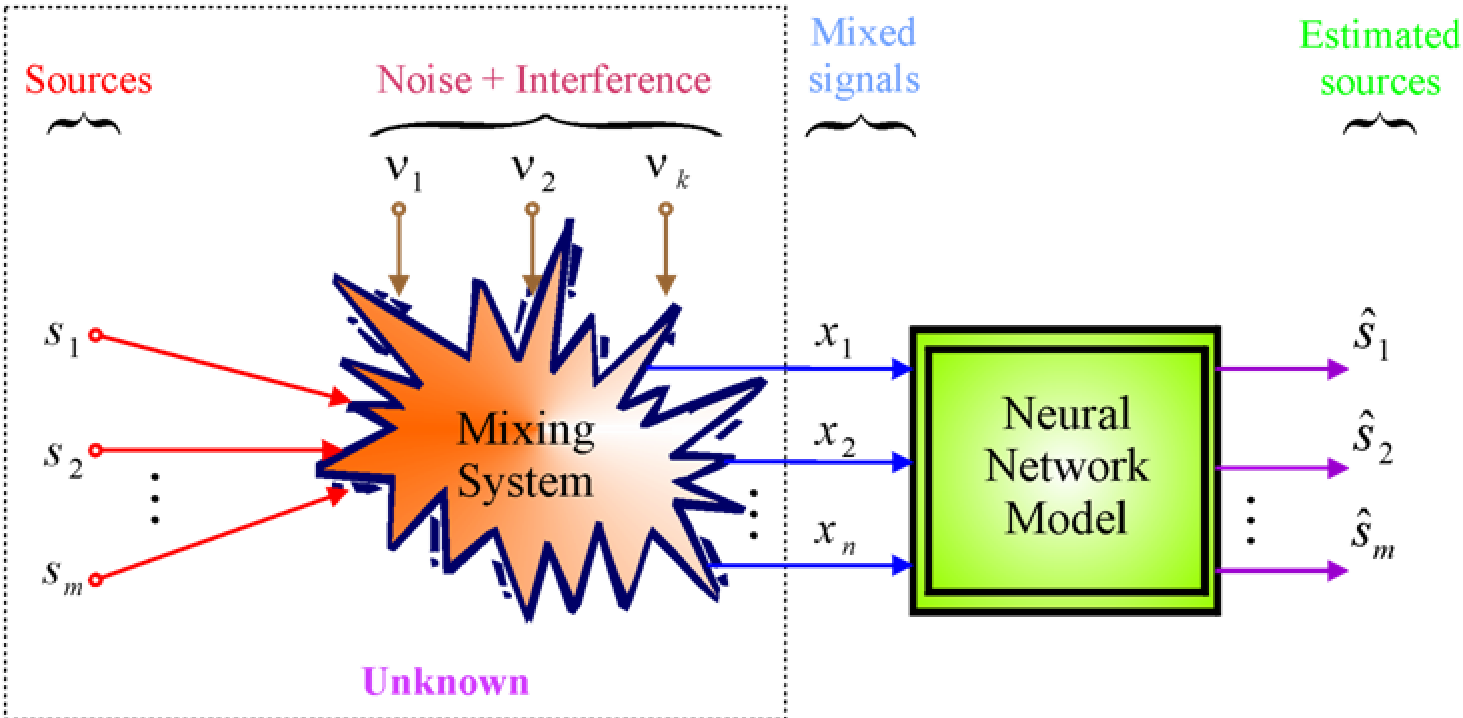


Learning algorithms

- Natural gradient algorithm
- Conjugate gradient algorithm
- Temporal ICA
- Activation adaptation
- Stability Analysis
- Efficiency

ICA Framework

Problem definition: How to recover the original source signals ?



Assumption: Mutual Independency of Sources

Cost Function

The Kullback-Leibler Divergence between two probability density functions $p(y) \quad and \quad q(y)=\prod_{i=1}^{n} p_i(y_i)$

$$D(p,q)=\int p(y)\log\left(\frac{p(y)}{\prod_{i=1}^{n} p_i(y_i)}\right)dy$$

where is the pdf of y. The output signals y are mutually independent if and only if

$$D(p,q)=0$$

Cost Function (II)

The Kullback-Leibler Divergence is employed as a cost function, which can be simplified as follows

$$l(y, W) = -\log\left|\det(W)\right| - \sum_{i=1}^{n} \log q(y_i)$$

where **W** is the demixing model, and $q_i(y_i)$ is an estimation of the true probability density function $p_i(y_i)$ of source signals

(Amari et al 1997; Zhang et al, IEEE SPL1998)

Temporal ICA

- To use the temporal structures of source signals to train the ICA model

$$s(k) \longrightarrow \boxed{H} \xrightarrow{x(k)} \boxed{W} \xrightarrow{y(k)} \boxed{A(z)} \xrightarrow{r(k)}$$

IRA Model: $$r(k) = A(z)Wx(k)$$

$$\Delta W = \mu(I - \sum_{p=0}^{L} A_p \phi(r) y^T (k-p))W \qquad A(z) = \sum_{p=0}^{L} A_p z^{-p}$$

Manifold of Matrices

The manifold of nonsingular $n \times n$ matrices is an open sub-manifold of $\mathcal{M}(n, \mathbf{R})$, the set of $n \times n$ matrices, defined by

$$\mathcal{M}(n) = \{\mathbf{W} \in \mathbf{R}^{n \times n} \| \det(\mathbf{W}) \neq 0\},$$

For example, the blind source separation model

$$\mathbf{r}(k) = \mathbf{A}(z)\mathbf{W}\mathbf{x}(k)$$

Geodesic

A geodesic is the shortest path connecting two elements on a manifold.

$$\mathbf{W}(0) = \mathbf{W}_1, \mathbf{W}'(0) = \mathbf{X}_1.$$

In this case, the geodesic is explicitly expressed by

$$\mathbf{W}(t) = \exp(t\mathbf{X}_1\mathbf{W}_1^{-1})\mathbf{W}_1.$$

Natural Gradient

$$\frac{d}{dt} L(\mathbf{W}(t)) \,|_{t=0} \quad = \quad tr\left(\frac{\partial L(\mathbf{W})}{\partial \mathbf{W}} \mathbf{X}^T \right)$$

$$\nabla L(\mathbf{W}) = \frac{\partial L(\mathbf{W})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W}$$

Conjugate Gradient Direction

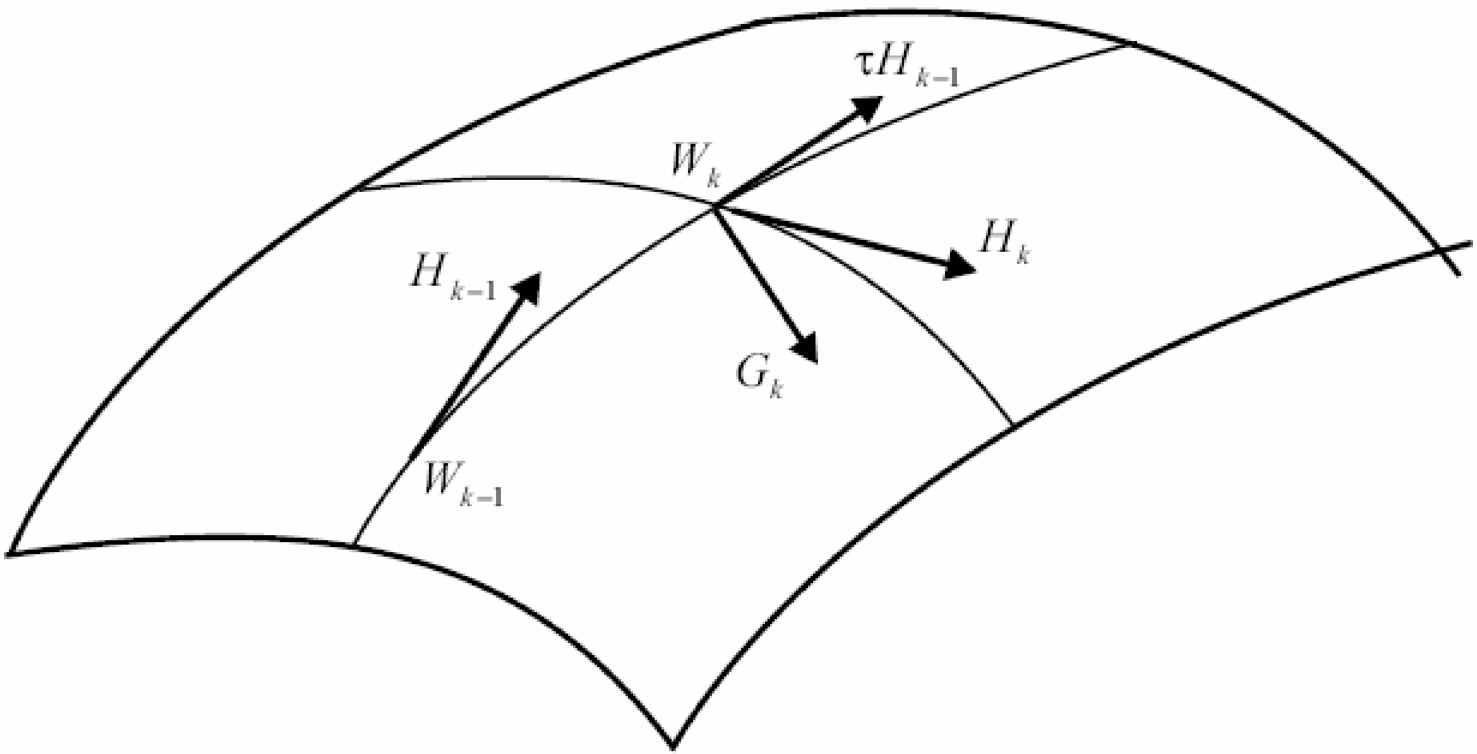
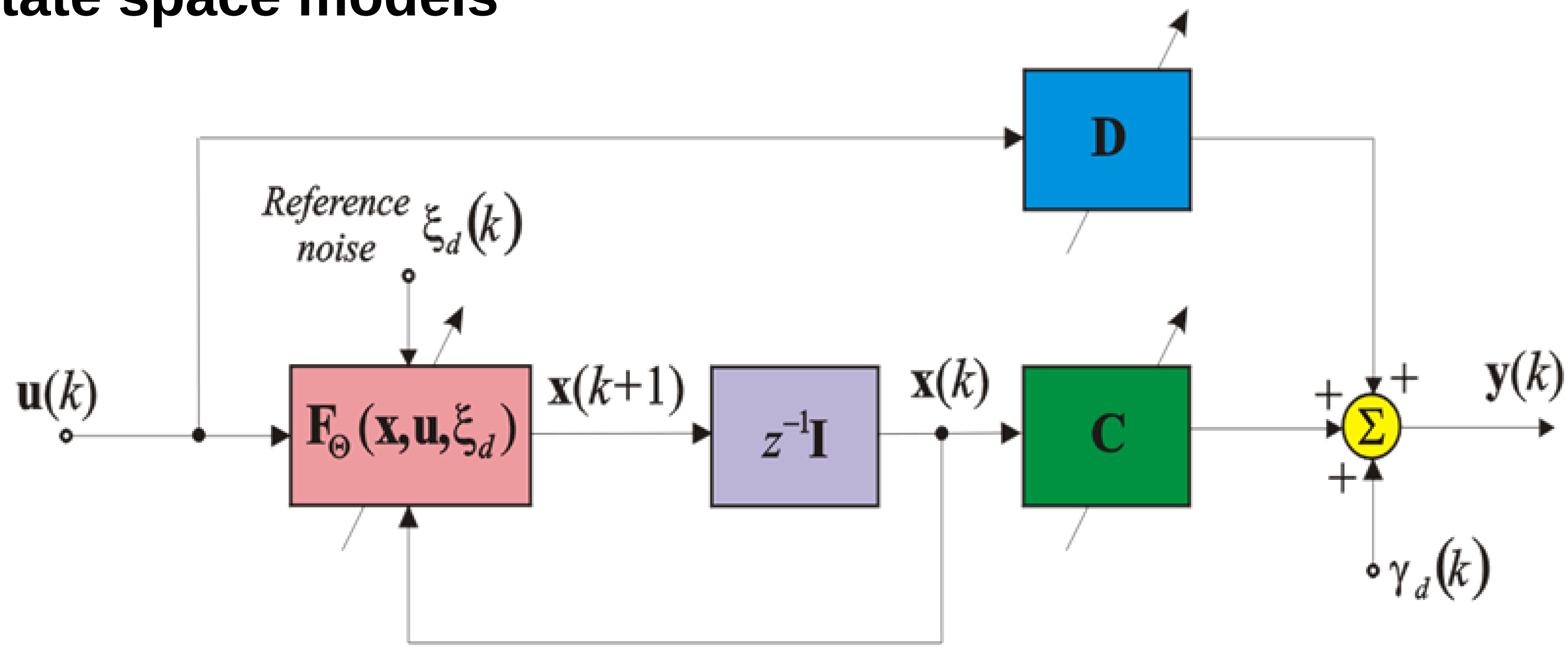


Fig. 2. Illustration of the conjugate gradient direction on the Riemannian manifold

Dynamic ICA

- **The representation model are described by nonlinear state space models**



IEEE SP 2004; SP 2003

Deconvolution Model

$$\mathbf{x}(k+1) = \mathbf{F}_{\mathbf{\Theta}}(\mathbf{x}(k), \mathbf{u}(k), \boldsymbol{\xi}_d(k)),$$

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k) + \boldsymbol{\gamma}_d(k).$$

- $\boldsymbol{u}(\mathrm{k}) = (u_1(k), \ldots, u_m(k))^{\mathrm{T}}$: the vector of m-sensor signals;
- $\boldsymbol{x}(\mathrm{k}) = (x_1(k), \ldots, x_m(k))^{\mathrm{T}}$: the vector of m-states of the deconvolution model;
- $\boldsymbol{y}(\mathrm{k}) = (y_1(k), \ldots, y_m(k))^{\mathrm{T}}$: the output of the demixing model, used to estimate the sources $\boldsymbol{s}(\mathrm{k})$

Θ, C and **D** are the free parameters in the demixing model

Cost Function (cont.)

The Kullback-Leibler Divergence is employed as a cost function, which can be simplified as follows

$$l(y,W) = -\log\left|\det(D)\right| - \sum_{i=1}^{n} \log q(y_i)$$

Where **W=[C, D],** where $q_i(y_i)$ is an estimation of the true probability density function $p_i(y_i)$ of source signals

(Zhang et al, NIPS1998)

Natural Gradient Learning

Natural Gradient

$$\vec{\nabla} l(\mathbf{y}, \mathbf{W}) = \nabla l(\mathbf{y}, \mathbf{W}) \begin{bmatrix} \mathbf{I} + \mathbf{C}^T\mathbf{C} & \mathbf{C}^T\mathbf{D} \\ \mathbf{D}^T\mathbf{C} & \mathbf{D}^T\mathbf{D} \end{bmatrix}$$

Natural Gradient Learning

$$\Delta\mathbf{C}(k) = \eta((I - \varphi(y)y^T)\mathbf{C} - \varphi(y)x^T),$$
$$\Delta\mathbf{D}(k) = \eta(I - \varphi(y)y^T)\mathbf{D}.$$

(Zhang et al, ICA2001,   Zhang et al, JSP2002)

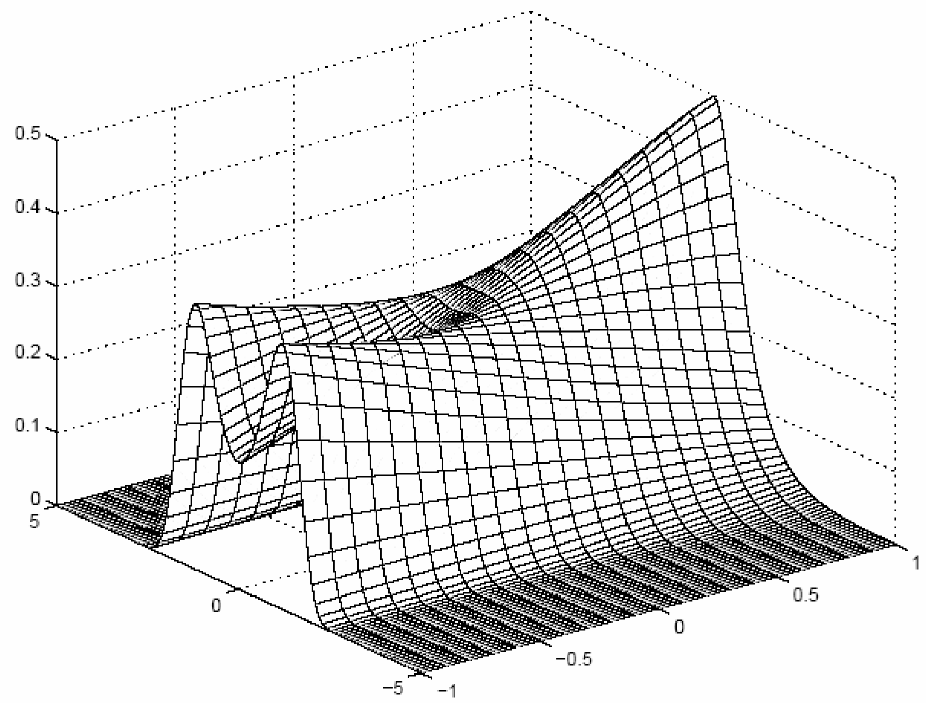
Table 11.1 Family of adaptive learning algorithms for state-space models.

Reference	Model	Algorithm
Zhang and Cichocki (1998) [1352]	linear	$\Delta \mathbf{C} = -\eta\ \mathbf{f}(\mathbf{y})\ ^{T}$ $\Delta \mathbf{D} = \eta\,[\mathbf{I} - \ \mathbf{f}(\mathbf{y})\,\mathbf{x}^{T}\,\mathbf{D}^{T}\ ]\,\mathbf{D}$
Zhang and Cichocki (1998) [1353]	linear	$\Delta \mathbf{C} = \eta\,[(\mathbf{I} - \ \mathbf{f}(\mathbf{y})\,\mathbf{y}^{T}\ )\,\mathbf{C} - \ \mathbf{f}(\mathbf{y})\ ^{T}\ ]$ $\Delta \mathbf{D} = \eta\,[\mathbf{I} - \ \mathbf{f}(\mathbf{y})\,\mathbf{y}^{T}\ ]\,\mathbf{D}$ *with Kalman filter*
Cichocki and Zhang (1998) [1354]	nonlinear	$\Delta \mathbf{C} = \eta\,[(\mathbf{\Lambda} - \ \mathbf{f}(\mathbf{y})\,\mathbf{y}^{T}\ )\,\mathbf{C} - \ \mathbf{f}(\mathbf{y})\ ^{T}\ ]$ $\Delta \mathbf{D} = \eta\,[\mathbf{\Lambda} - \ \mathbf{f}(\mathbf{y})\,\mathbf{y}^{T}\ ]\,\mathbf{D}$
Cichocki and Zhang (1998) [281]	nonlinear	Two-stage approach
Cichocki and Zhang (1999) [282]	nonlinear	$\Delta \mathbf{C} = \eta\,[(\mathbf{I} - \ \mathbf{f}(\mathbf{y})\,\mathbf{y}^{T}\ )\,\mathbf{C} - \ \mathbf{f}(\mathbf{y})\ ^{T}\ \ \mathbf{\Lambda}]$ $\Delta \mathbf{D} = \eta\,[\mathbf{I} - \ \mathbf{f}(\mathbf{y})\,\mathbf{y}^{T}\ ]\,\mathbf{D}$
Salam and Waheed (2001) [1034]	linear	Recurrent network
Salam and Erten (1999) [1033]	nonlinear	Lagrange multiplier approach
Zhang and Cichocki [1358, 1359]	nonholonomic	$\Delta[\mathbf{C}\ \mathbf{D}] = -\eta\,\nabla\mathcal{R}\begin{bmatrix} \mathbf{I}+\mathbf{C}^{T}\mathbf{C} & \mathbf{C}^{T}\mathbf{D} \\ \mathbf{D}^{T}\mathbf{C} & \mathbf{D}^{T}\mathbf{D} \end{bmatrix}$
	NG Algorithm	$\Delta \mathbf{C}(k) = \eta(k)\left[\ \mathbf{\Lambda}^{(k)} - \mathbf{R}^{(k)}_{\mathbf{f}\,\mathbf{y}}\quad \mathbf{C}(k) - \mathbf{R}^{(k)}_{\mathbf{f}}\right]$ $\Delta \mathbf{D}(k) = \eta(k)\left[\mathbf{\Lambda}^{(k)} - \mathbf{R}^{(k)}_{\mathbf{f}\,\mathbf{y}}\right]\mathbf{D}(k)$

(Cichocki and Amari's Book, pp435)

Activation Function Adaptation

$$p_e(y,\theta) = \exp\left(-\theta\frac{|y|^4}{4} - (1-\theta)\log sech(y) + \mathcal{N}(\theta)\right).$$



Activation Function Adaptation (II)

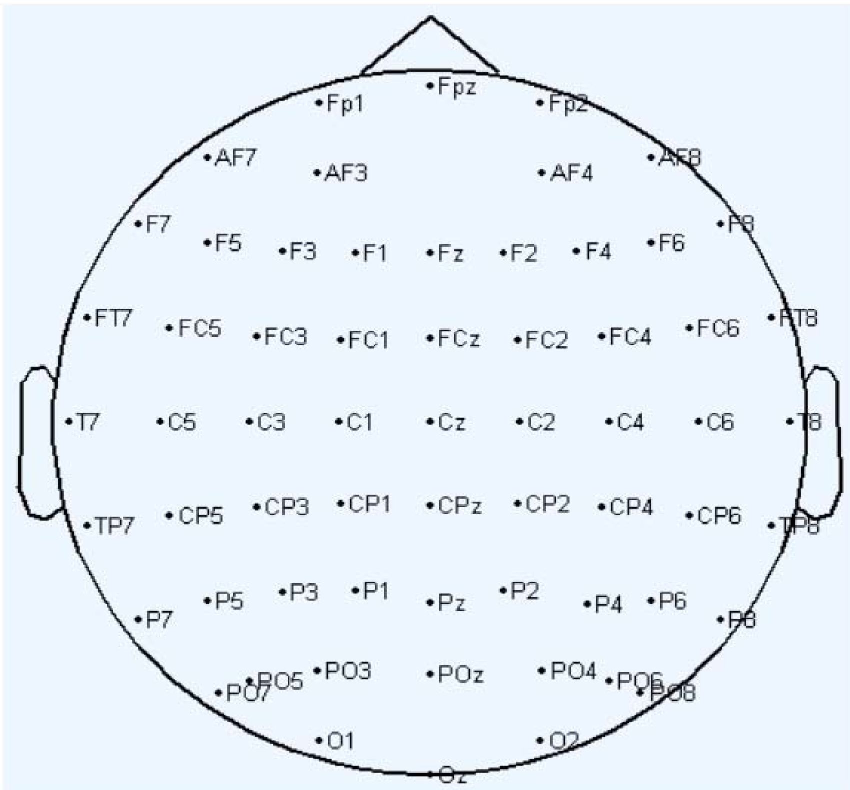
$$\frac{\partial l(\mathbf{y},\boldsymbol{\theta},\mathbf{W})}{\partial \boldsymbol{\theta}_i} = \frac{\partial \psi(y_i,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i} - \mathcal{N}'(\boldsymbol{\theta}_i)$$

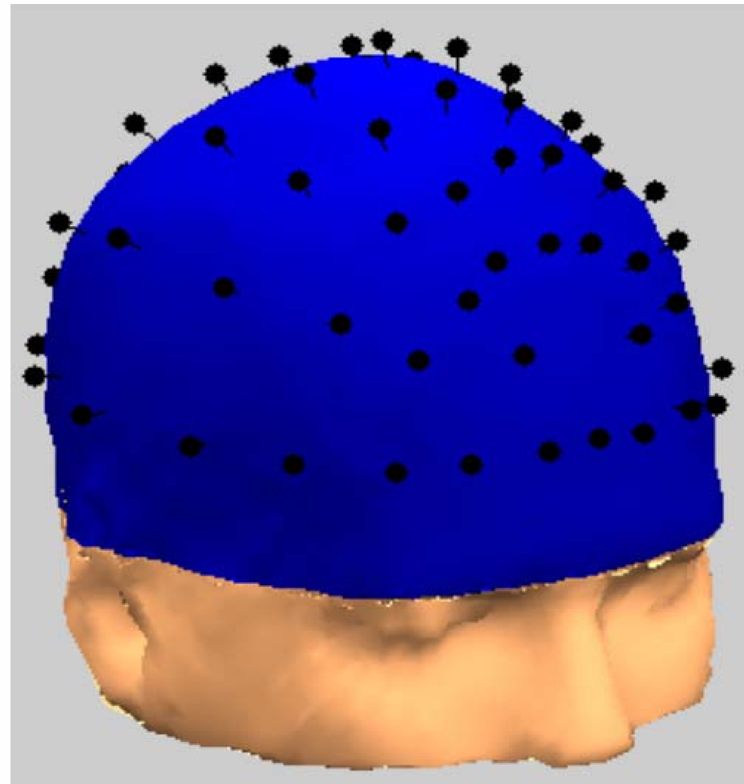
$$\Delta\boldsymbol{\theta}_i = \eta\left(\int_0^{y_i} \frac{\partial \varphi(\tau,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} d\tau - \mathcal{N}'(\boldsymbol{\theta}_i)\right).$$

(Zhang et al, IEEE NN 2004)

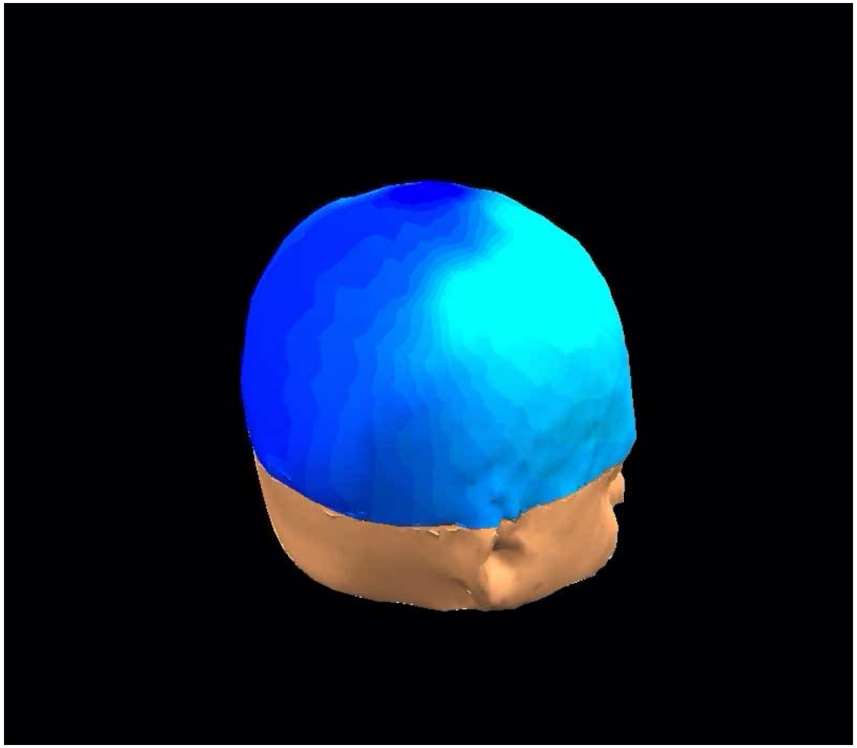
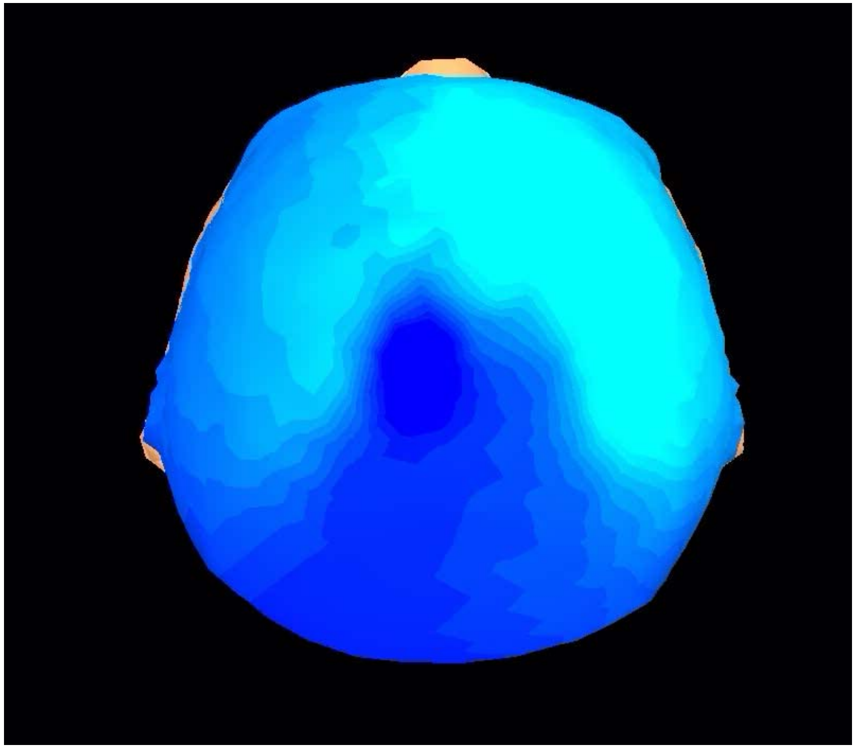
NeuronScan EEG

NeuroScan Electrodes Distribution

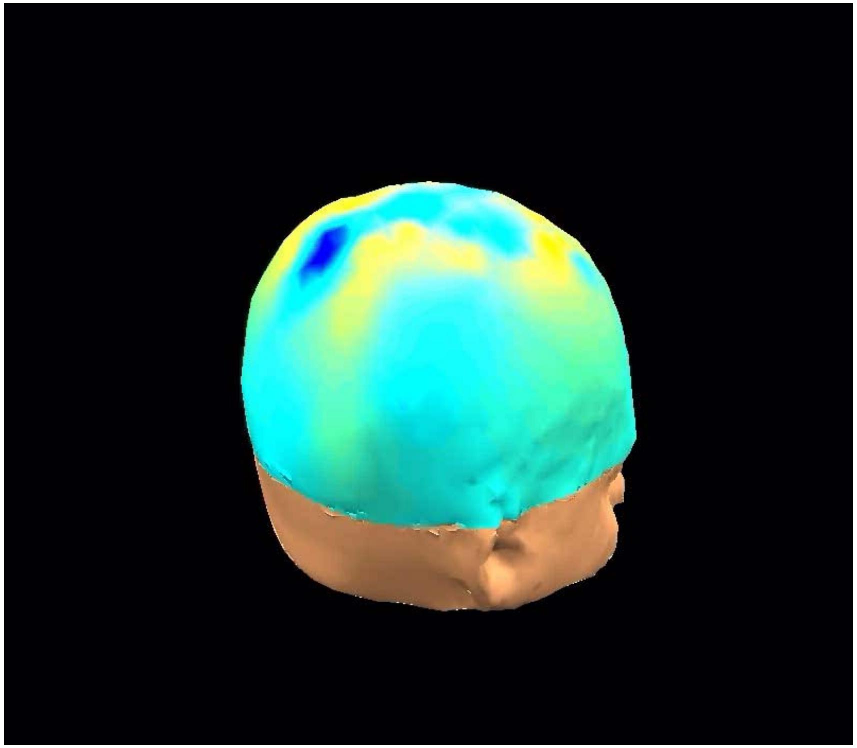


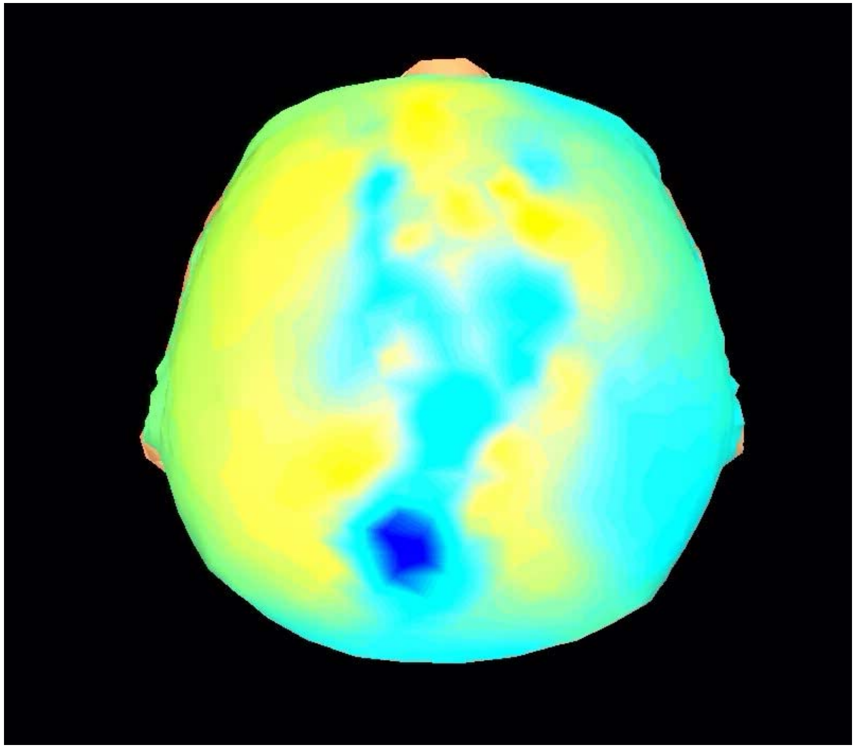


EEG Data Visualization

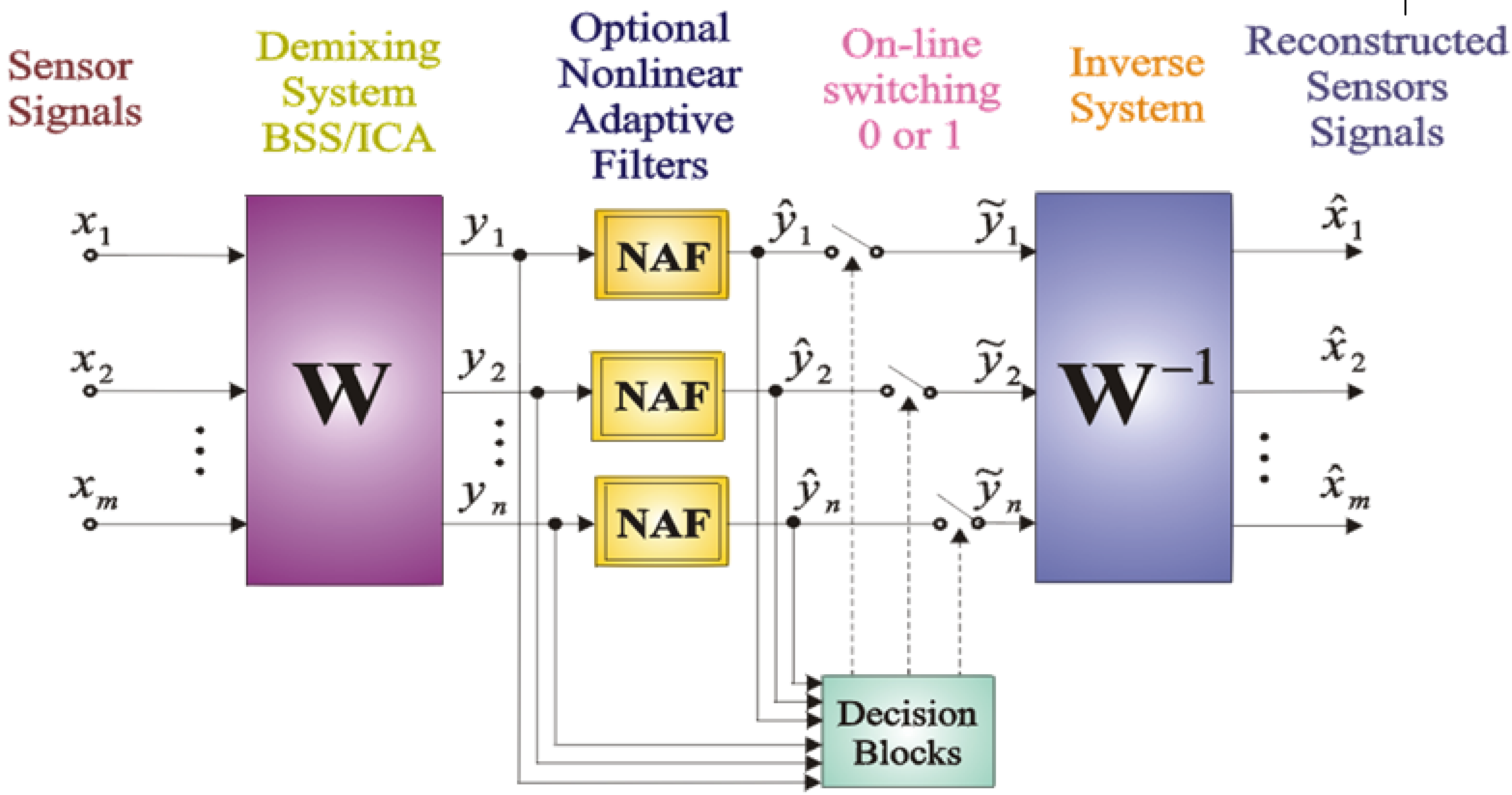
EEG Data Visualization





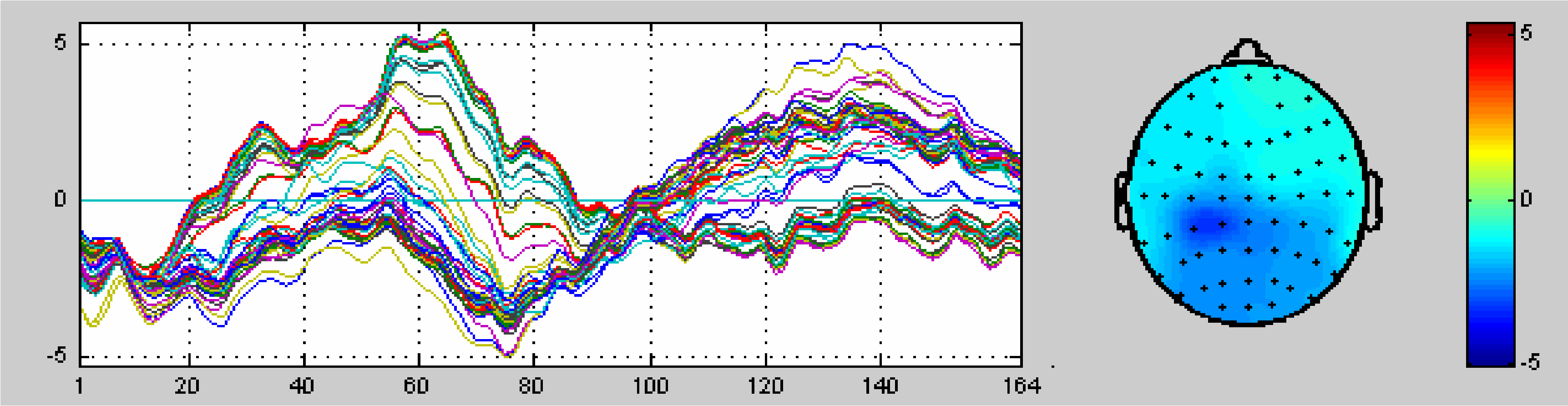
Zhao Qibin, BCI Poster

脑诱发电位（源）估计



Intelligent "Decision Blocks" allow to switch off/on each channel automatically depending on features of separated signals.

VEP Localization



References

- Attneave F. 1954. Some informational aspects of visual perception. Psychol. Rev. 61:183–93 J. Opt. Soc. Am. A 1(8):885–87
- Barlow HB. 1961. Possible principles underlying the transformation of sensory messages. In Sensory Communication, ed. WA Rosenblith, pp. 217–34. Cambridge, MA: MIT Press
- Dan Y, Atick JJ, Reid RC. 1996. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. J. Neurosci. 16:3351–62
- Vinje WE, Gallant JL. 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273–76
- Atick JJ. 1992. Could information theory provide an ecological theory of sensory processing? Netw. Comput. Neural Syst. 3:213–51
- Olshausen BA, Field DJ. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–9
- Bell AJ, Sejnowski TJ. 1997. The “independent components” of natural scenes are edge filters. Vis. Res. 37(23):3327–38
- van Hateren JH, van der Schaaf A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. London Ser. B 265:359–66

References

- Field DJ. 1994. What is the goal of sensory coding? Neural Comput. 6:559–601
- H. Barlow. What is the computational goal of the neocortex? In C. Koch, editor, Large scale neuronal theories of the brain, pages 1–22. MIT Press, Cambridge, MA, 1994.
- P. Comon. Independent component analysis - a new concept? Signal Processing, 36:287– 314, 1994.
- A. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. Neural Computation, 7(6):1129–1159, 1995. 138
- M. McKeown, S. Makeig, G. Brown, T.-P. Jung, S. Kindermann, A. Bell, and T. Sejnowski. Analysis of fmri data by decomposition into independent components. Proc. Nat. Acad. Sci., in press.
- A. Bell and T. Sejnowski. The independent components of natural scenes are edge filters. Vision Research, 37(23):3327–3338, 1997.
- B. Olshausen and D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381:607–609, 1996.

References

- Bin Xia, Liqing Zhang, ”Blind Deconvolution in Non-minimum Phase systems Using Cascade Structure” , to appear EURASIP Journal on Applied Signal Processing (2006)
- Liqing Zhang, Multi-Modal Machine Attention: Sound Localization and Visual-Auditory Signal Synchronization, in “Human Interaction with Machines”, Hommel, G.; Huanye, S. (Eds.), ISBN: 1-4020-4042-3, Springer, pp57-64, 2005
- Li, YQ; Cichocki, A; Zhang, LQ, Blind source estimation of FIR channels for binary sources: a grouping decision approach, SIGNAL PROCESSING, 84 (12): 2245-2263 DEC 2004
- L. Zhang, A. Cichocki and S. Amari, Multichannel Channel Blind Deconvolution of Non-minimum Phase Systems Using Filter Decomposition, IEEE Transactions on Signal Processing, Vol. 52, No5, pp1431-1442, 2004
- L. Zhang, A. Cichocki and S. Amari, Self-Adaptive Blind Source Separation Based on Activation function Adaptation, IEEE Transactions on Neural Networks, Vol.15, No.2, pp233-244, 2004
- L. Zhang, A. Cichocki and S. Amari, Geometrical Structures of FIR Manifold and their Application to Multi-channel blind deconvolution, Journal of VLIS for Signal Processing Systems, 31:31-44, 2002,
- L. Zhang S. Amari and A. Cichocki, Semi-parametric Model and Superefficiency in Blind Deconvolution, Signal Processing, 81:2535-2553, 2001
- L. Zhang and A. Cichocki, Blind Deconvolution of Dynamical Systems: A State Space Approach, Invited paper, Journal of Signal Processing, Vol.4, No.2, pp.111-130, 2000.
- L. Zhang, A. Cichocki and S. Amari, Semiparametric Approach for Blind Deconvolution of Non-minimum Phase Systems, Advances in Neural Information Processing Systems 12, Editors: S.A. Solla, T.K. Leen and K.-R. Muller , MIT Press, pp363-369, 2000.
- L. Zhang, A. Cichocki and S. Amari, Natural Gradient Algorithm to Blind Separation of Over-determined Mixture with Additive Noises, IEEE Signal Processing Letters, Vol. 6, No. 11, pp293-295, 1999.
- L. Zhang and A. Cichocki, Blind Separation of Filtered Source Using State-Space Approach, Advances in Neural Information Processing Systems 11, Editors: M.S. Kearns, S.A. Solla and D.A. Cohn, MIT press, pp648-654, 1999.

Thank you!

zhang-lq@cs.sjtu.edu.cn

上海交通大学感知计算实验室