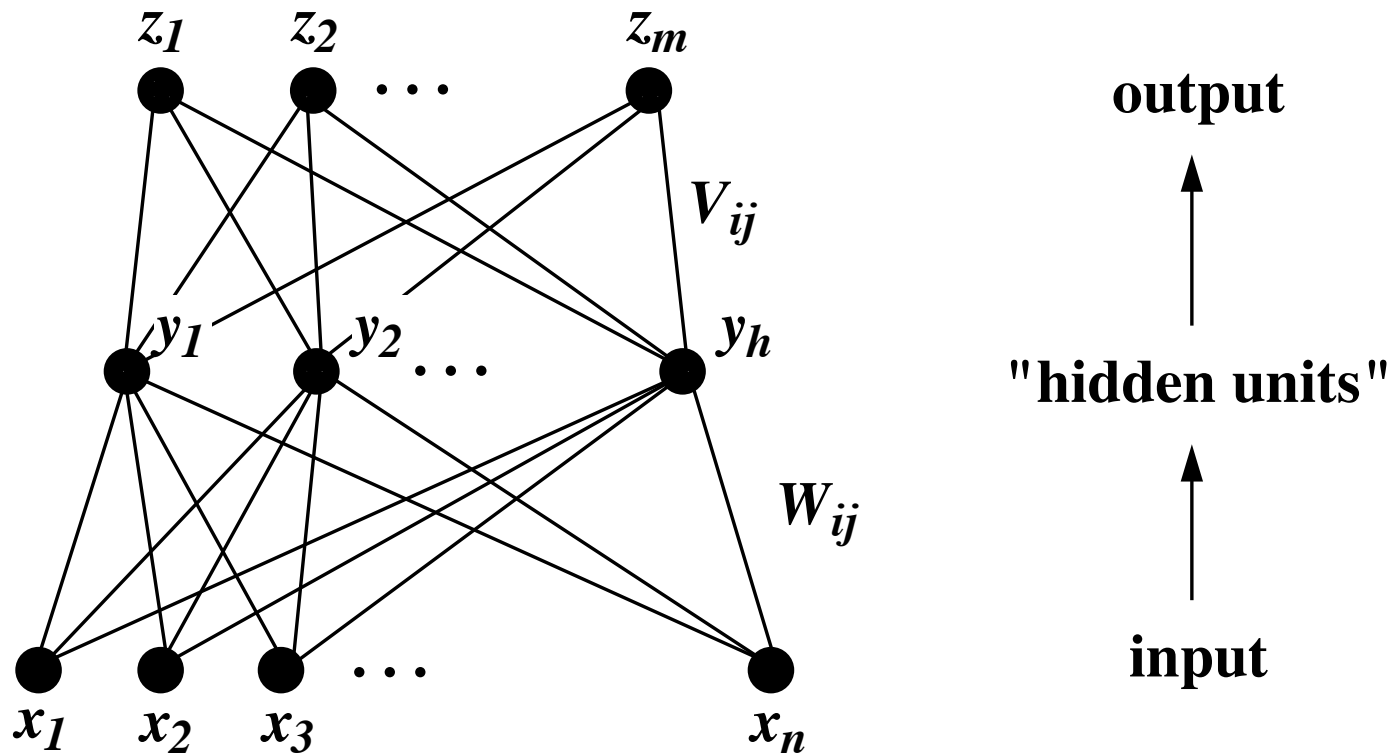


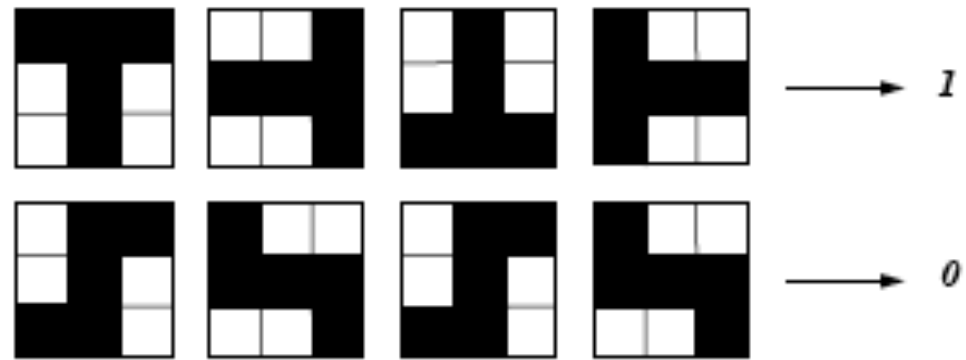
Unsupervised Learning

Two-layer network

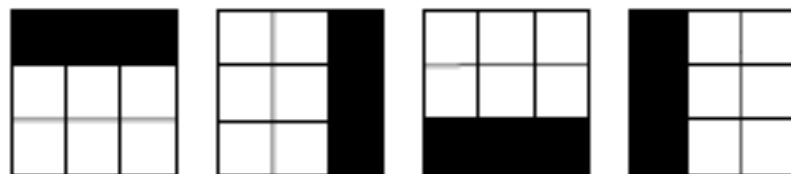


$$z_i = \sigma\left(\sum_j V_{ij} y_j\right)$$

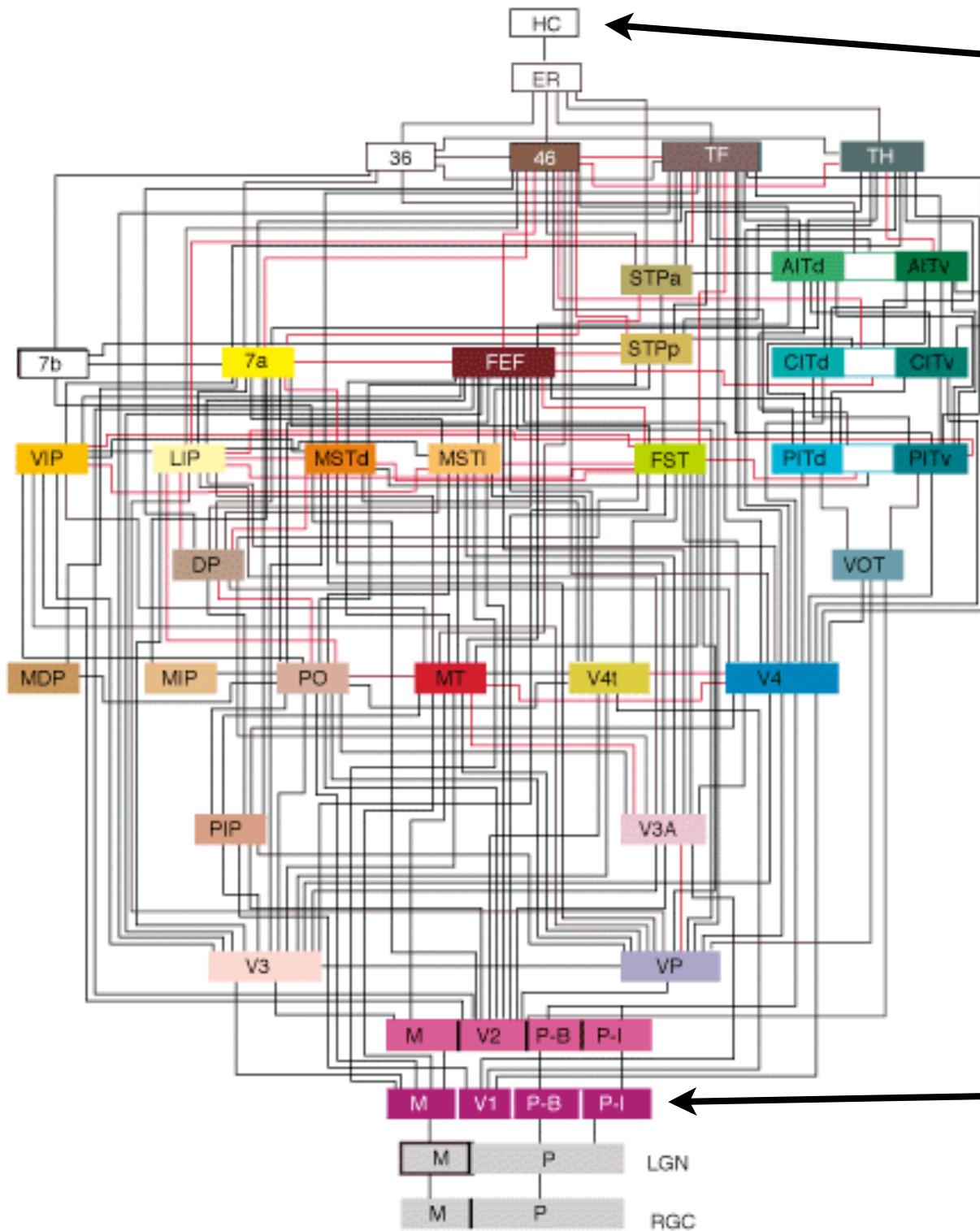
$$y_i = \sigma\left(\sum_j W_{ij} x_j\right)$$



hidden layer weights
(one possible solution)



Hippocampus



·
·
·
·
?
·
·
·
·
·
VI

**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

aliceswasbeginningtogetverytiredofsittingbyhersisteron
thebankandofhavingnothingtodoonceortwiceshehadpeep
edintothebookhersisterwasreadingbutithadnopicturesor
conversationsinitandwhatistheuseofabookthoughtalicesw
ithoutpicturesorconversationssoshewasconsideringinhe
rownmindaswellasshecouldforthehotdaymadeherfeelvery
sleepyandstupidwhetherthepleasureofmakingadaisychai
nwouldbeworththetroubleofgettingupandpickingthedaisi
eswhensuddenlyawhiterabbitwithpinkeyesranclosebyher
therewasnothingsoveryremarkableinthatnordidalicethi
nkitsoverymuchoutofthewaytoheartherabbitsaytoitselfo
hdearohdear

(Neural Computation, 1993)

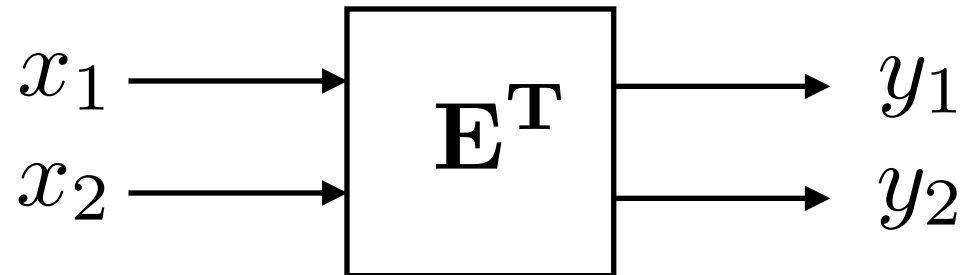
alice was beginning to get very tired of sitting by her sister on the bank and of having nothing to do once or twice she had peeped into the book her sister was reading but it had no pictures or conversations in it and what is the use of a book though t alice without pictures or conversations so she was considering in her own mind as well as she could for the hot day made her feel very sleepy and stupid whether the pleasure of making a daisy chain would be worth the trouble of getting up and picking the daisies when suddenly a white rabbit with pink eyes ran close by her there was nothing so very remarkable in that nor did alice think it so very much out of the way to hear the rabbit say to itself
oh dear oh dear

alice was beginning to get very tired of sitting by her sister on the bank and of having nothing to do once or twice she had peeped into the book her sister was reading but it had no pictures or conversations in it and what is the use of a book though t alice without pictures or conversations so she was considering in her own mind as well as she could for the hot day made her feel very sleepy and stupid whether the pleasure of making a daisy chain would be worth the trouble of getting up and picking the daisies when suddenly a white rabbit with pink eyes ran close by her there was nothing so very remarkable in that nor did alice think it so very much out of the way to hear the rabbit say to itself
oh dear oh dear

Hebbian Learning and PCA

PCA

(Principal Components Analysis)



$$\begin{array}{lll} \langle x_1 \ x_2 \rangle & = & c_{12} \\ & \neq & 0 \end{array} \quad \begin{array}{l} y_1 = \mathbf{e}_1 \cdot \mathbf{x} \\ y_2 = \mathbf{e}_2 \cdot \mathbf{x} \end{array} \quad \begin{array}{ll} \langle y_1 \ y_2 \rangle & = \langle y_1 \rangle \langle y_2 \rangle \\ & = 0 \end{array}$$

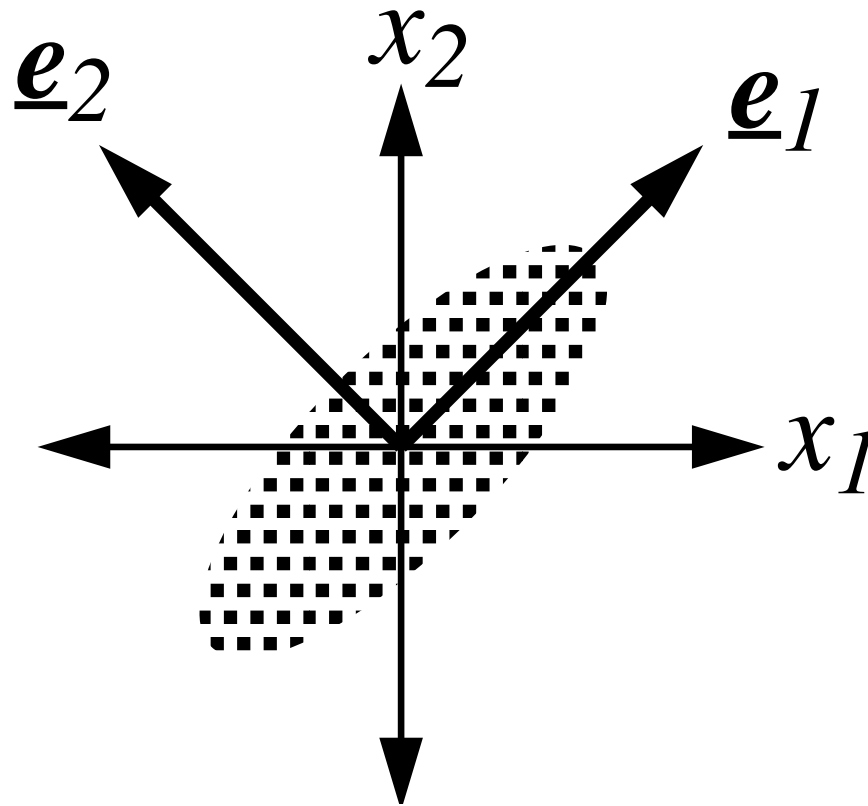
$$\mathbf{E} = \left[\begin{array}{c|c} | & | \\ \mathbf{e}_1 & \mathbf{e}_2 \\ | & | \end{array} \right]$$

$$\mathbf{e}_1 \cdot \mathbf{e}_2 = 0$$

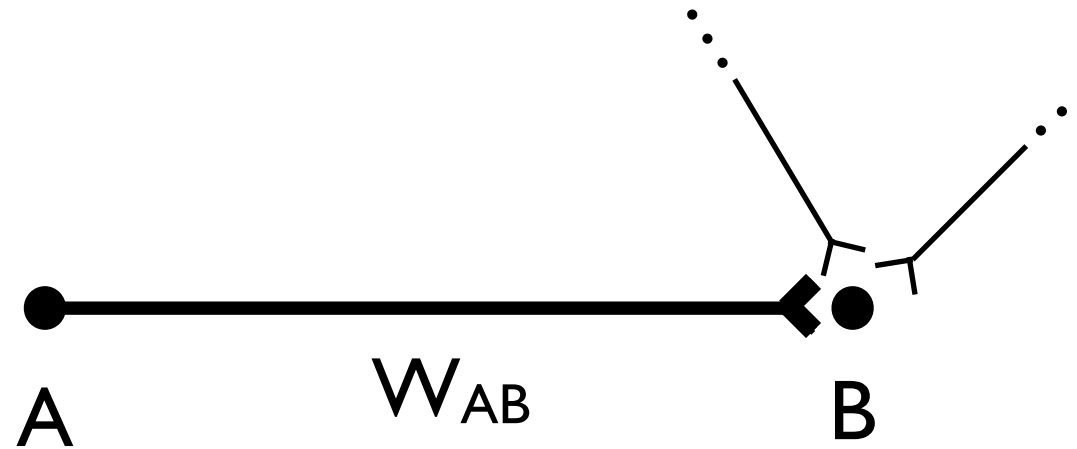
$$|\mathbf{e}_1| = |\mathbf{e}_2| = 1$$

PCA

(Principal Components Analysis)

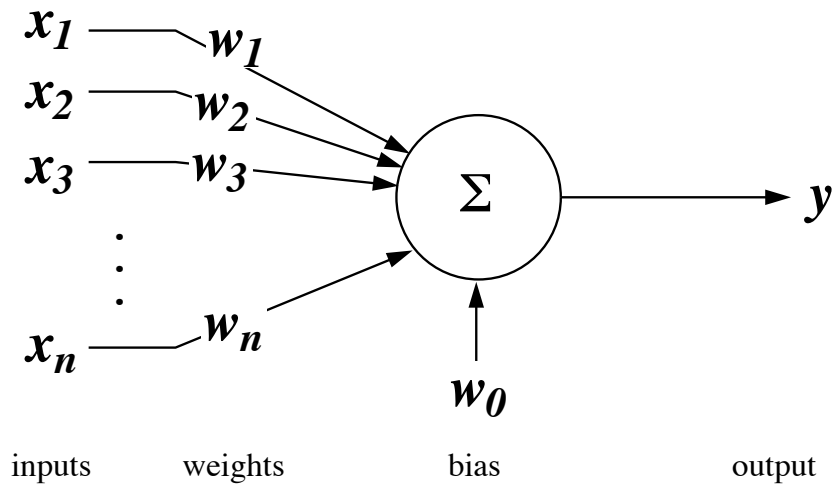


Hebbian learning



$$\Delta W_{AB} \propto \langle AB \rangle$$

Linear Hebbian learning



$$\dot{w}_i \propto \langle y x_i \rangle$$

$$y = \sum_j w_j x_j$$

$$\dot{w}_i \propto \left\langle \sum_j w_j x_j x_i \right\rangle$$

$$= \sum_j w_j \langle x_j x_i \rangle$$

$$\dot{\mathbf{w}} \propto \mathbf{C} \mathbf{w}$$

$$C_{ij} = \langle x_i x_j \rangle$$

$$\dot{\mathbf{w}} \propto \mathbf{C} \mathbf{w}$$

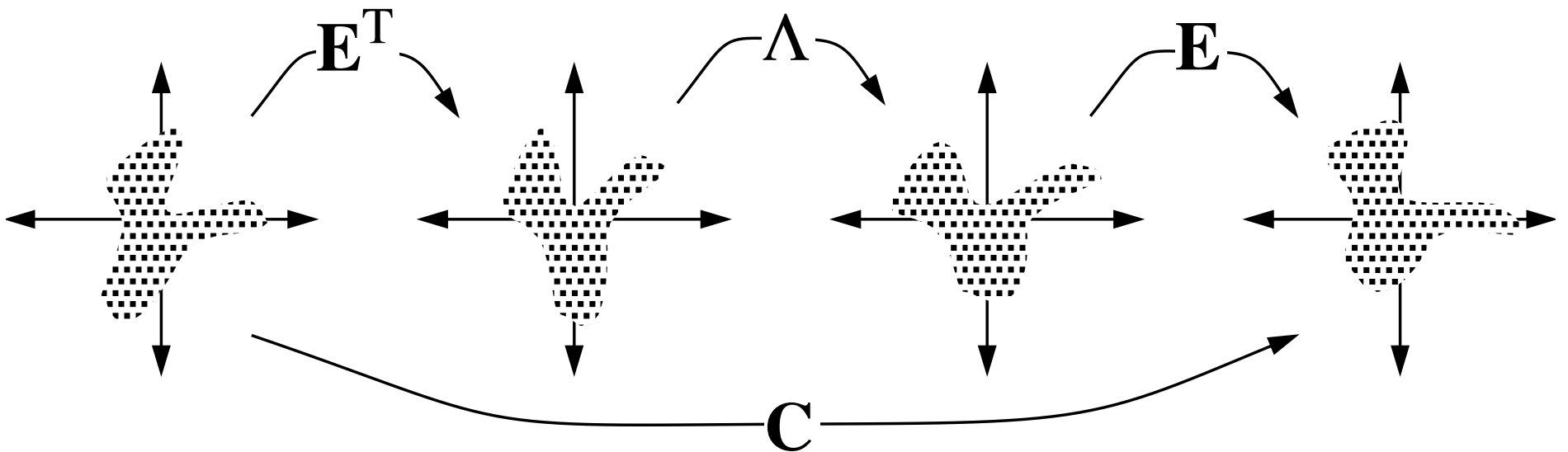
$$\dot{w}_1 \propto C_{11} w_1 + C_{12} w_2$$

$$\dot{w}_2 \propto C_{21} w_1 + C_{22} w_2$$

$$\mathbf{C} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^T$$

$$\mathbf{E} = \begin{bmatrix} | & | & \dots & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \\ | & | & \dots & | \end{bmatrix}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}$$



define: $\mathbf{v} = \mathbf{E}^T \mathbf{w}$

$$\dot{\mathbf{w}} \propto \mathbf{C} \mathbf{w}$$

$$= \mathbf{E} \mathbf{\Lambda} \mathbf{E}^T \mathbf{w}$$

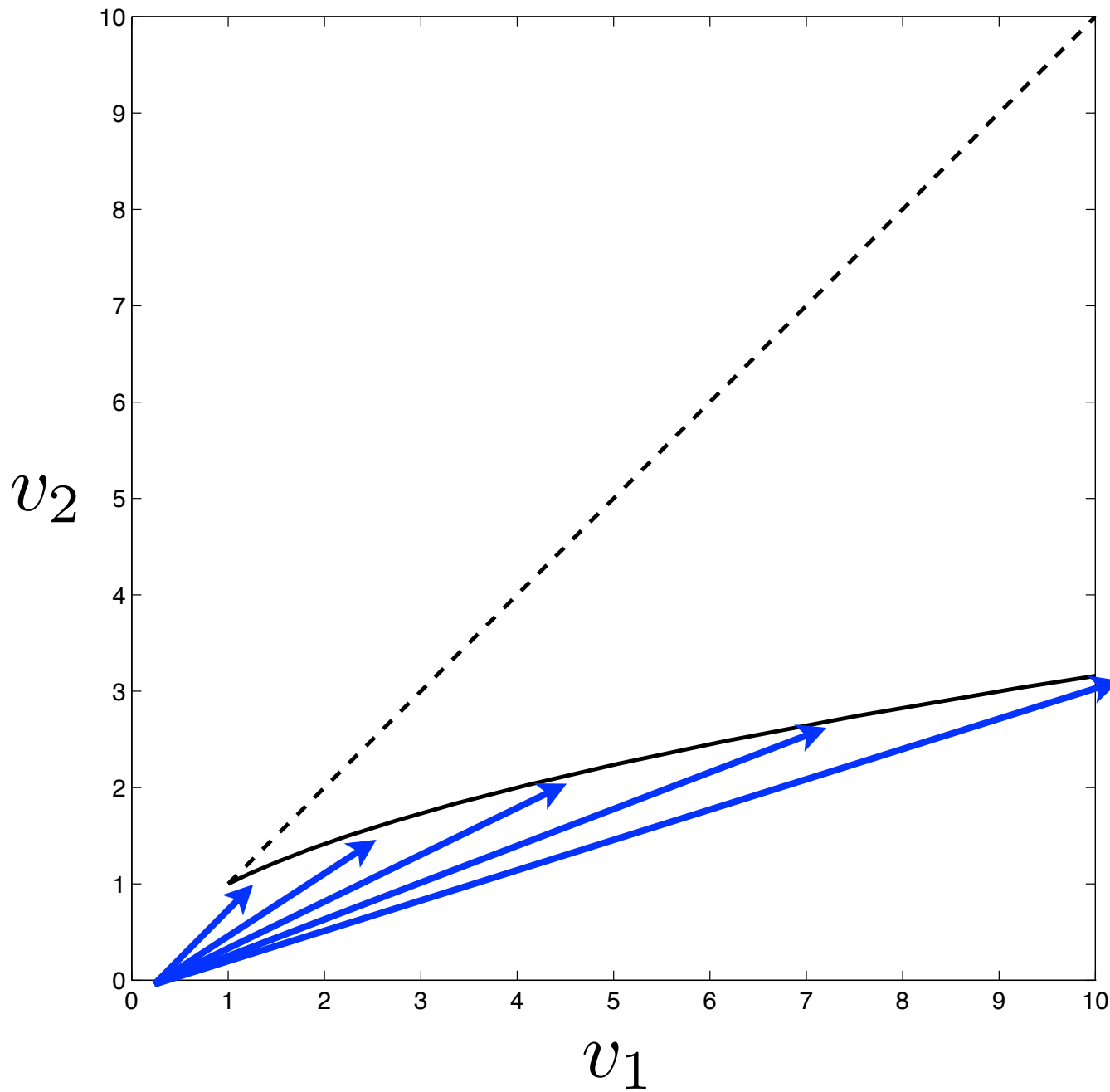
$$= \mathbf{E} \mathbf{\Lambda} \mathbf{v}$$

$$\mathbf{E}^T \dot{\mathbf{w}} \propto \mathbf{E}^T \mathbf{E} \mathbf{\Lambda} \mathbf{v}$$

$$\dot{\mathbf{v}} \propto \mathbf{\Lambda} \mathbf{v}$$

$$\dot{v}_1 \propto \lambda_1 v_1$$

$$\dot{v}_2 \propto \lambda_2 v_2$$



$$v_1(t) = e^{\lambda_1 t} v_1(0)$$
$$v_2(t) = e^{\lambda_2 t} v_2(0)$$

$$\mathbf{w} = \mathbf{E} \mathbf{v}$$

$$\mathbf{w} \rightarrow \alpha \mathbf{e}_1$$

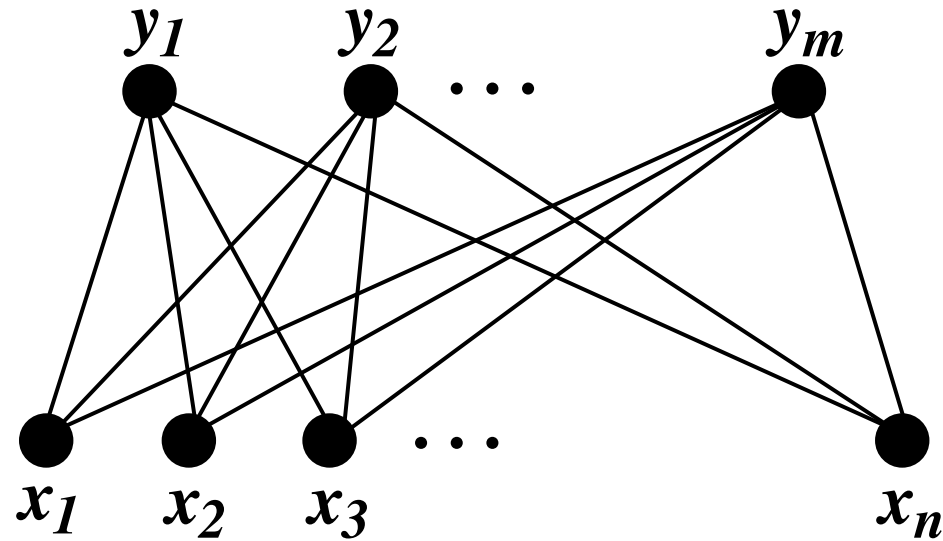
Constraining the growth of the weight vector

Oja's rule

$$\Delta w_i = \eta y (x_i - y w_i)$$

or $\Delta \mathbf{w} = \eta y (\mathbf{x} - y \mathbf{w})$

Multiple output units

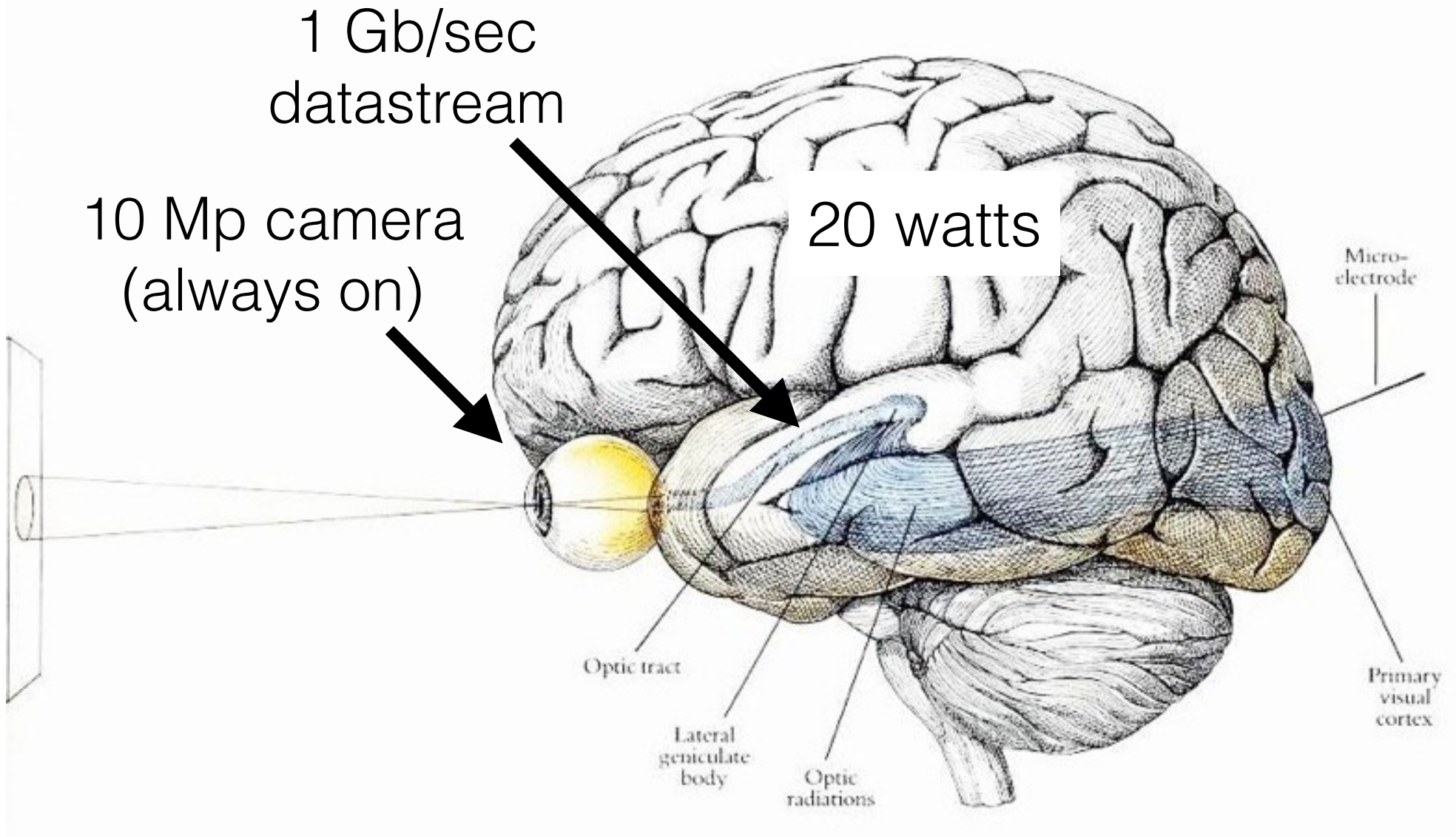


Sanger's rule:
$$\Delta w_{ij} = \eta y_i \left(x_j - \sum_{k=1}^i y_k w_{kj} \right)$$

1 Gb/sec
datastream

10 Mp camera
(always on)

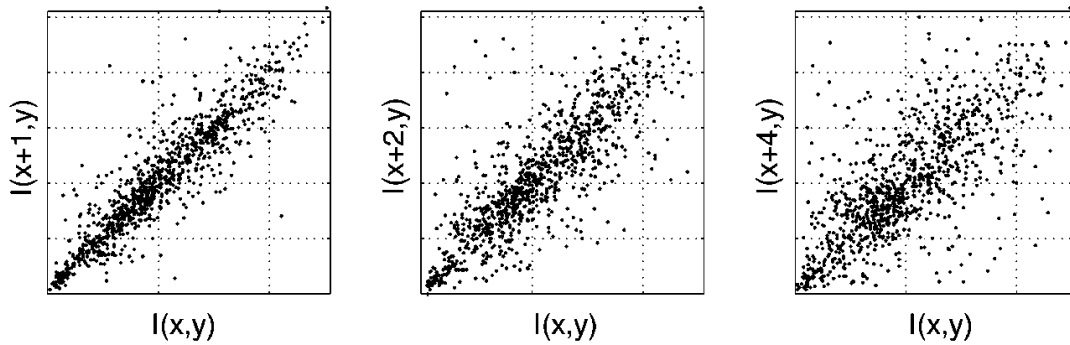
20 watts



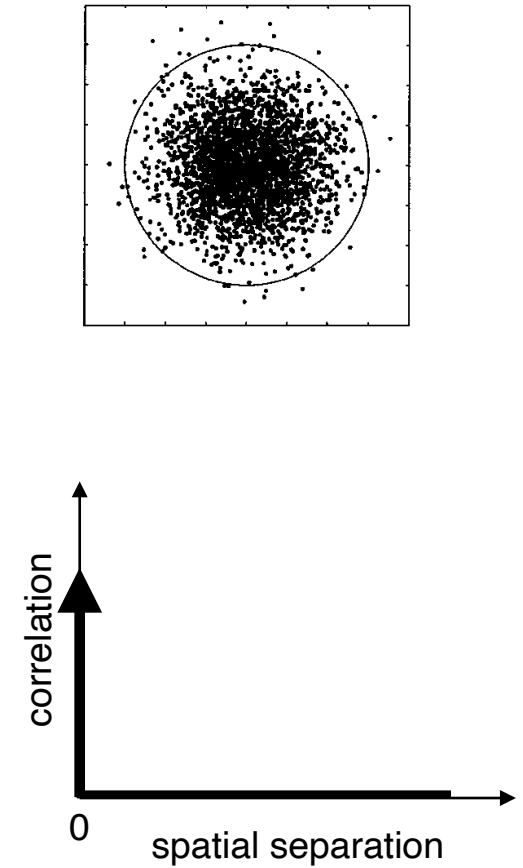
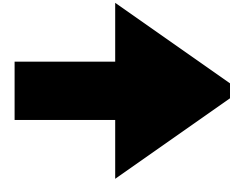
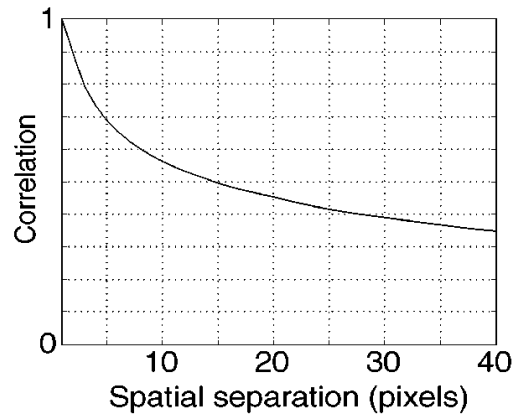
Decorrelation

Joint distribution of pixels
in natural images

a.

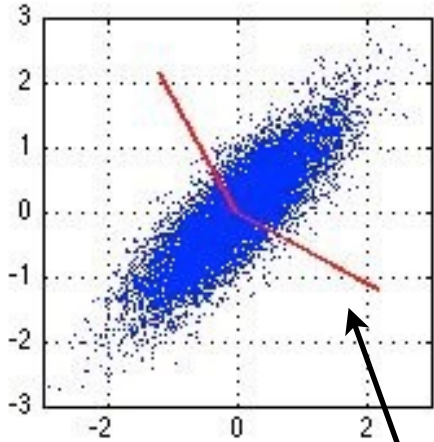


b.

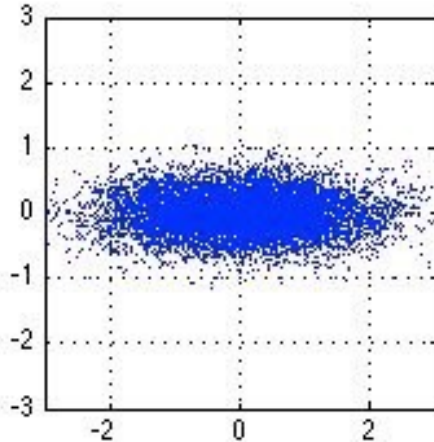


(from Simoncelli & Olshausen 2001)

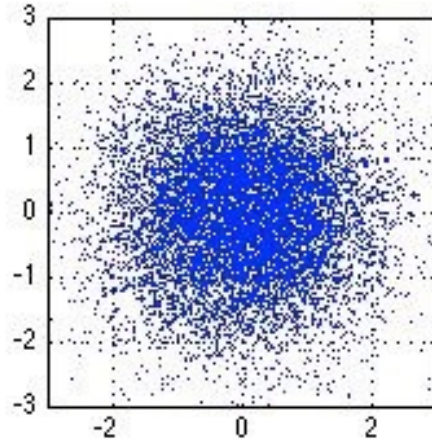
Whitening



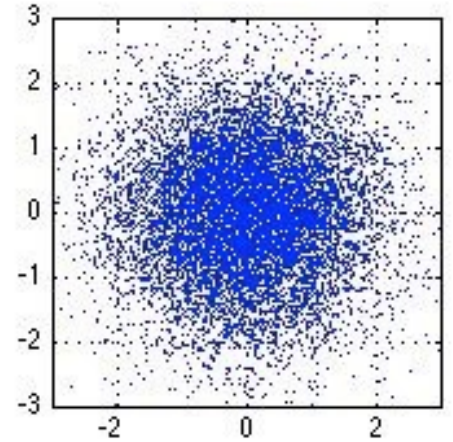
\mathbf{X}



$\mathbf{E}^T \mathbf{X}$



$\mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{E}^T \mathbf{X}$



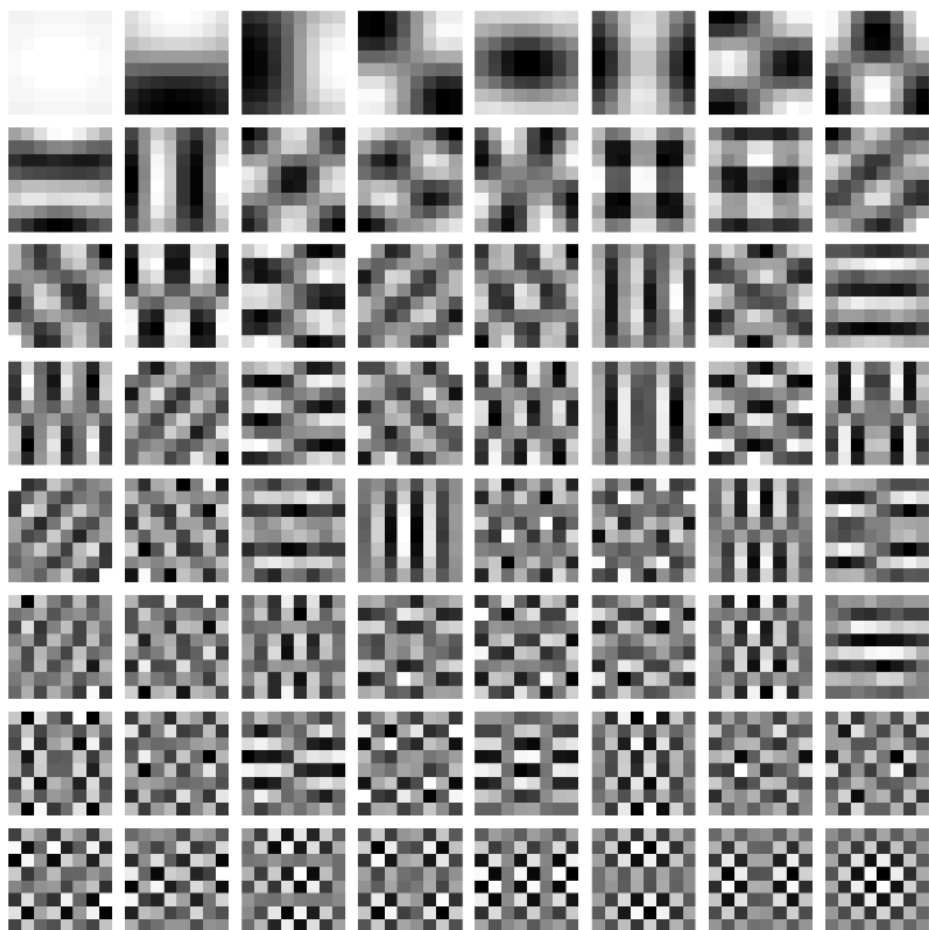
$\mathbf{E} \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{E}^T \mathbf{X}$

$$\mathbf{Y} = \mathbf{W} \mathbf{X}$$

$$\mathbf{W} = \mathbf{E} \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{E}^T$$

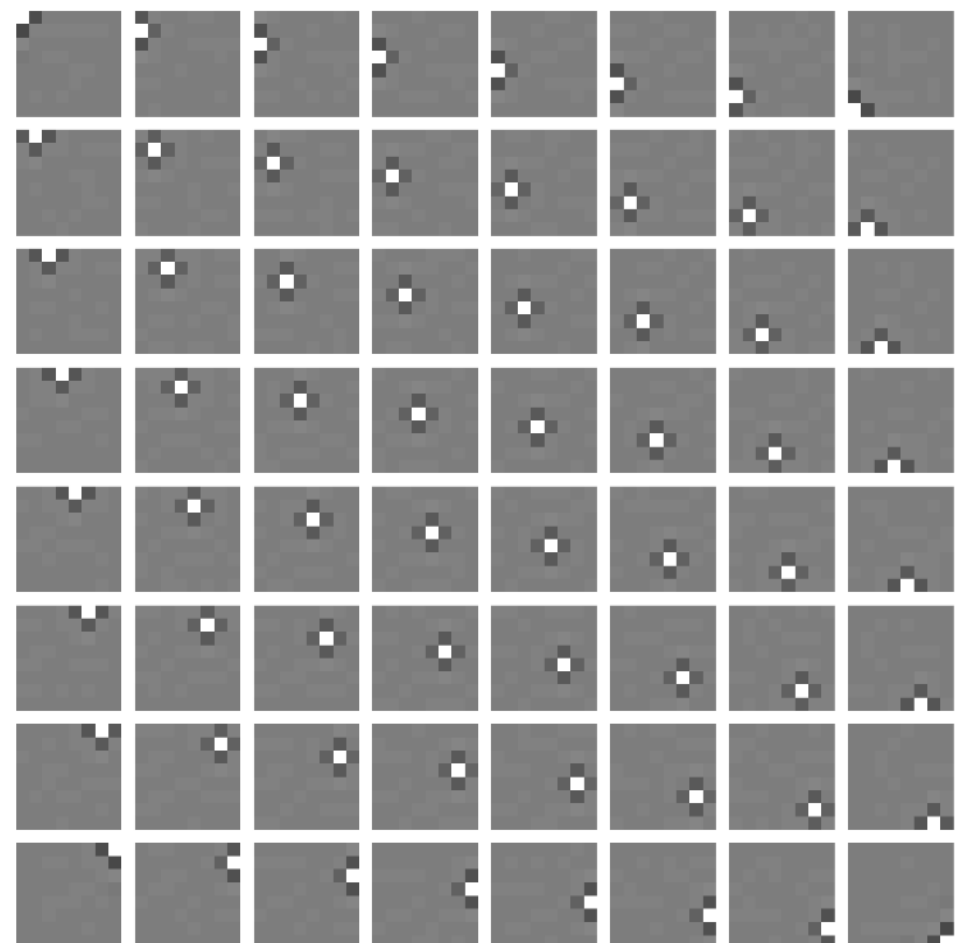
PCA

$$\mathbf{W} = \mathbf{E}^T$$



“Whitening”

$$\mathbf{W} = \mathbf{E}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{E}^T$$

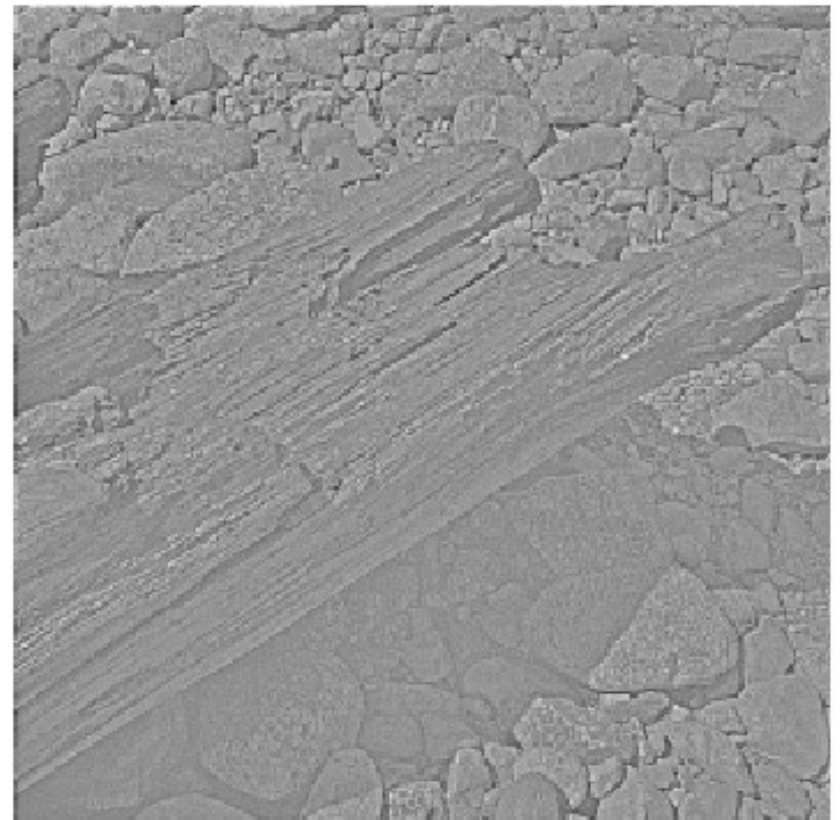


“Whitening”

original image

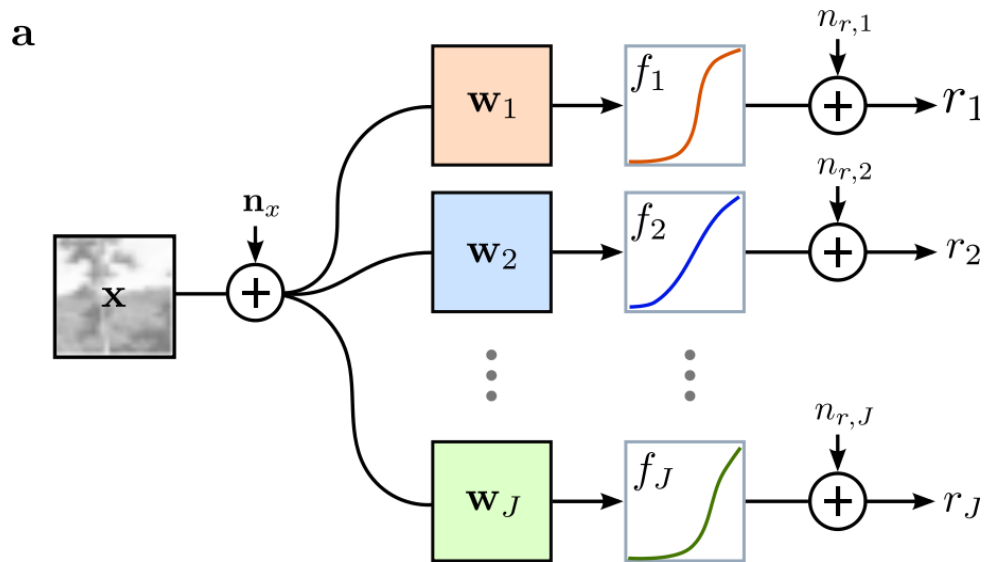


whitened image



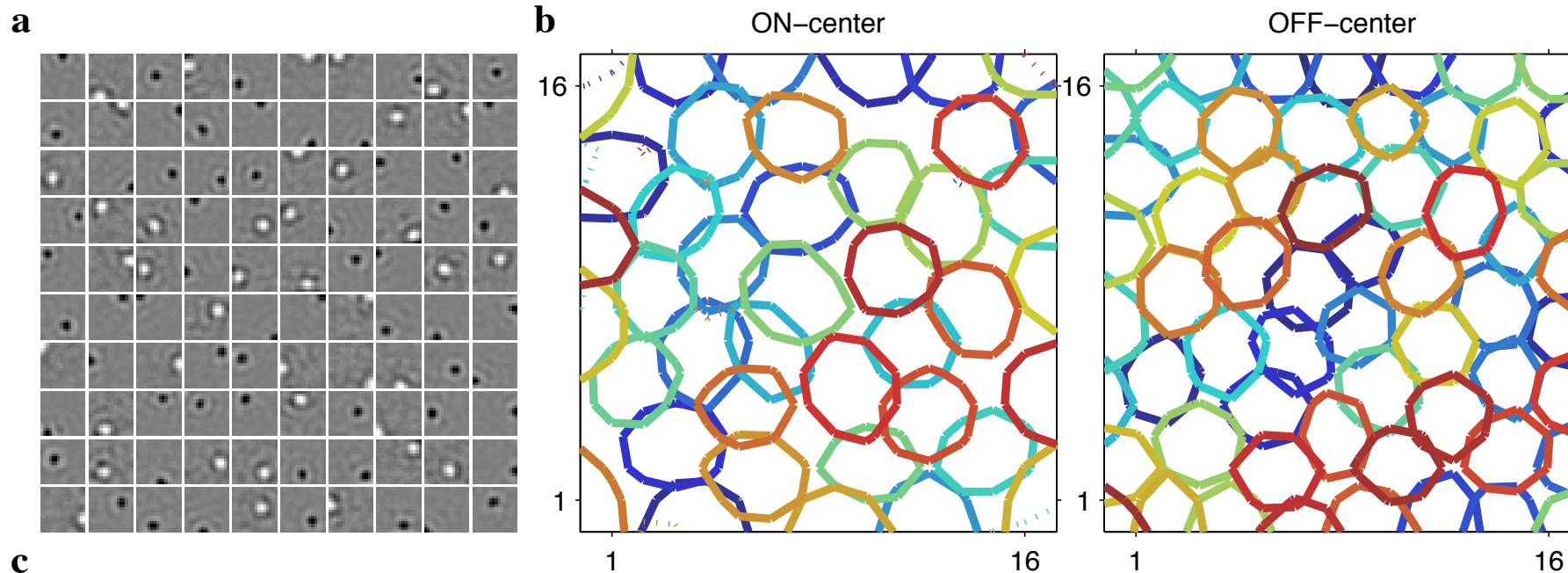
Efficient coding model of retina

(Karklin & Simoncelli 2012)



Objective function:

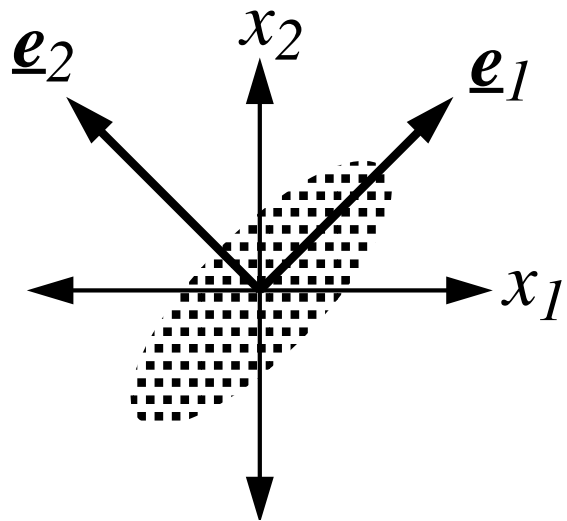
$$I(X; R) - \sum_j \lambda_j \langle r_j \rangle$$



Competitive Learning

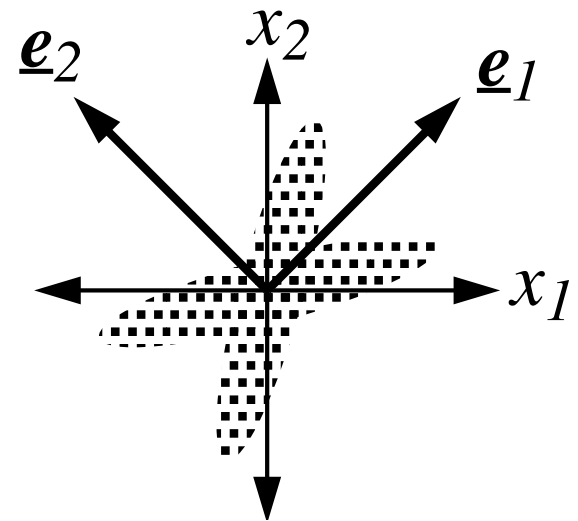
Gaussian

a.



non-Gaussian

b.



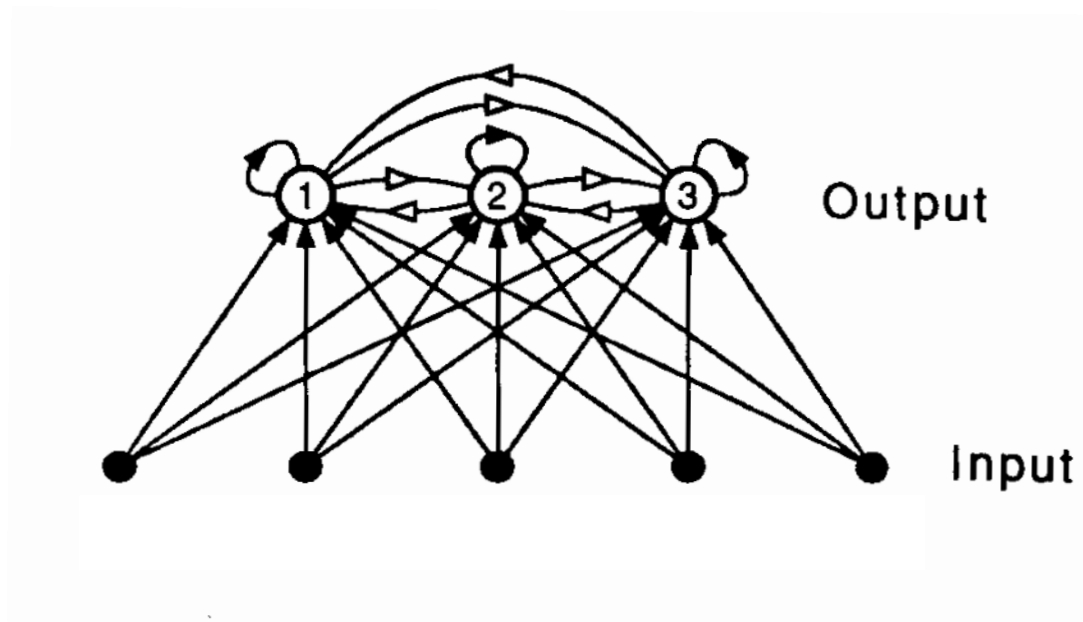
Non-linear Hebbian learning

Non-linear neuron: $y = f\left(\sum_i w_i x_i\right)$

Hebbian learning then yields:

$$\begin{aligned}\Delta w_i &\propto y x_i \\ &= f\left(\sum_j w_j x_j\right) x_i \\ &= k_0 x_i + k_1 \sum_j w_j x_j x_i \\ &\quad + k_2 \sum_{jk} w_j w_k x_k x_j x_i + \dots\end{aligned}$$

Winner-take-all learning



$$u_i = \sum_j w_{ij} x_j$$

Network:

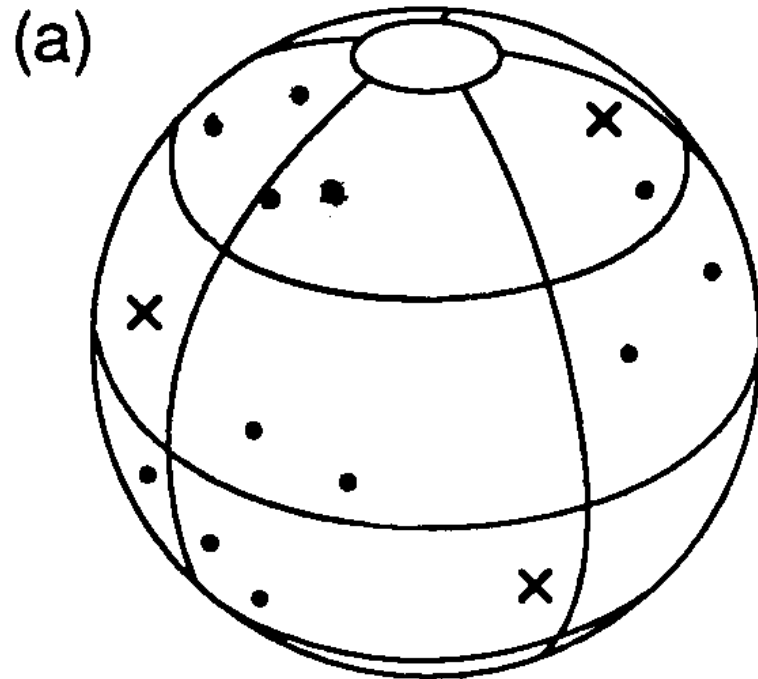
$$y_i = \begin{cases} 1 & u_i > u_j \quad \forall j \neq i \\ 0 & \text{otherwise} \end{cases}$$

Learning rule:

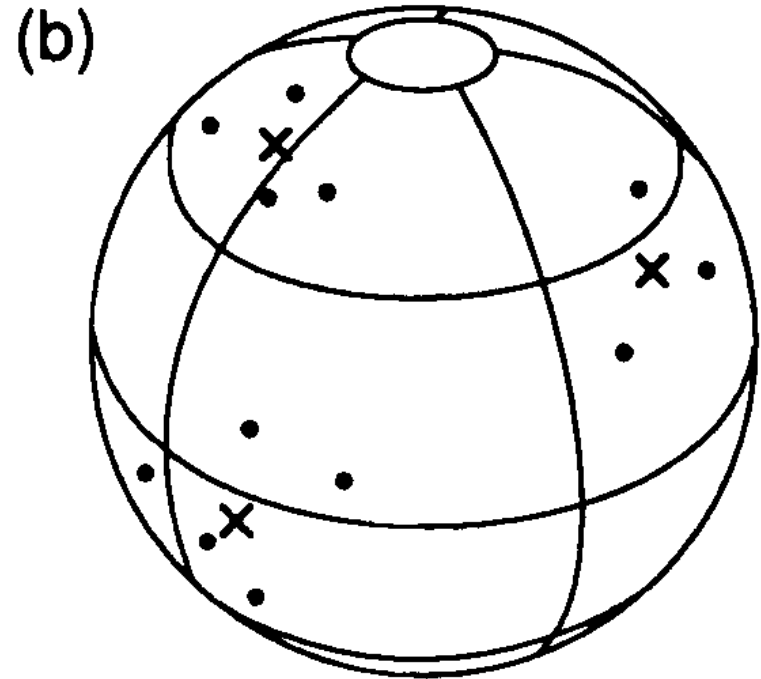
$$\Delta w_{ij} = \eta y_i (x_j - w_{ij})$$

Winner-take-all learning

before learning



after learning



Winner-take-all learning

Energy function:

$$E(\{\mathbf{w}_i\}) = \frac{1}{2} \sum_{i,\mu} M_i^\mu \|\mathbf{x}^{(\mu)} - \mathbf{w}_i\|^2$$

Gradient descent:

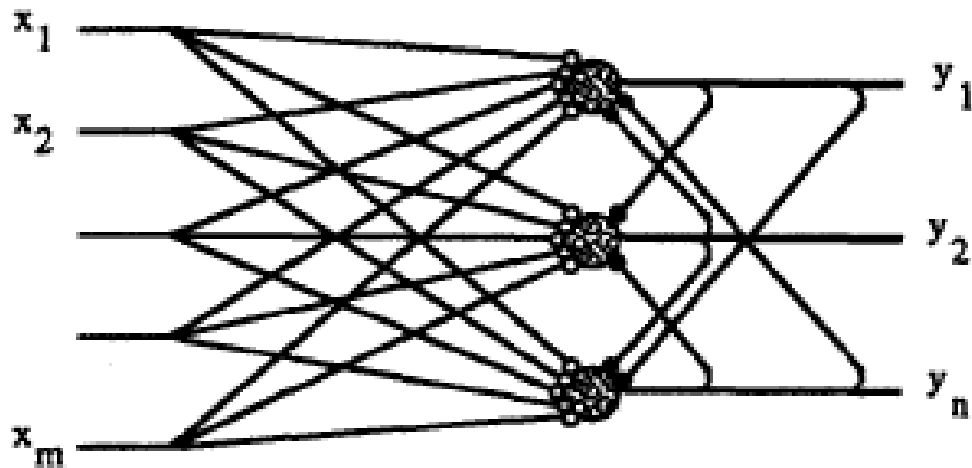
$$\begin{aligned} \Delta \mathbf{w}_i &= -\eta \frac{\partial E}{\partial \mathbf{w}_i} \\ &= \eta \sum_{\mu} M_i^\mu (\mathbf{x}^\mu - \mathbf{w}_i) \end{aligned}$$

Forming sparse representations by local anti-Hebbian learning

P. Földiák

Physiological Laboratory, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

$$\frac{dy_i^*}{dt} = f \left(\sum_{j=1}^m q_{ij} x_j + \sum_{j=1}^n w_{ij} y_j^* - t_i \right) - y_i^*$$



anti-Hebbian rule—

$$\Delta w_{ij} = -\alpha(y_i y_j - p^2)$$

(if $i = j$ or $w_{ij} > 0$ then $w_{ij} := 0$)

Hebbian rule—

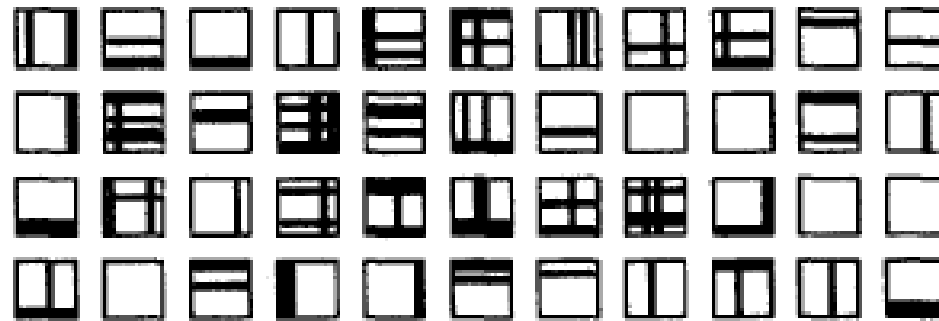
$$\Delta q_{ij} = \beta y_i (x_j - q_{ij})$$

threshold modification—

$$\Delta t_i = \gamma(y_i - p).$$

Learning lines

Input patterns:



Learned weights:

