Lecture 8: Hebbian learning in development

Jochen Braun

Otto-von-Guericke-Universität Magdeburg,
Cognitive Biology Group

Theoretical Neuroscience II, SS 2020

Credits:
Dayan & Abbott (2001) "Theoretical Neuroscience", Chpt. 8
Miller (1994) J. Neurosci. 14(1): 409–41.
Erwin & Miller (1998) J. Neurosci. 18(23): 9870-95.

8. Hebbian learning in development

In visual development, nearby neurons express the same ocular dominance. Recurrent cooperation-competition between output neurons explains this. Combining input correlations $\boldsymbol{Q}$ *with output cooperation-competition* $\boldsymbol{K}$ *produces alternating* **'ocular dominance bands'**.

To illustrate the generality of correlation-driven development, we discuss two classic papers on the **development of simple cells**. *One describes development of dominance by 'on-' or 'off-center' input, the other of combined dominance by 'on' or 'off' and by 'left-' or 'right-eye' input.*

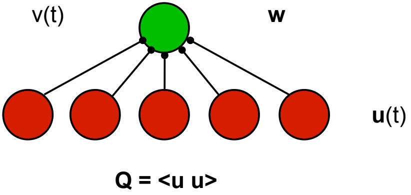
Finally, we return to **supervised Hebbian learning** *(see Lecture 2 and Exercise 4), where input- and output-activity is externally imposed and synaptic weights develop such as to reproduce the imposed input-output correlations. We consider the (intentionally) artificial example of transforming Cartesian to polar coordinates.*

Organization of lecture

- **1 Recurrent connections of target neurons**
- **2 Development of ocular dominance bands**
- **3 Computational models of visual development (two papers)**
- **4 Supervised Hebbian learning (Exercise 4)**

Unsupervised learning with single output

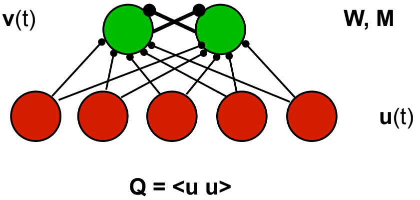
So far, we have considered unsupervised learning with a single (cortical) output neuron v and multiple (subcortical) input neurons $\boldsymbol{u}$:



'Unsupervised' means that output activity is due to the (developing) feedforward synapses. If all inputs are equally large and variable, learning is driven by *correlations* $\boldsymbol{Q}$ of input activities. As we have seen, some outcomes are difficult to achieve in this way (e.g., ocular dominance) .

Unsupervised learning with multiple interacting outputs

We will now consider unsupervised learning with multiple (cortical) output neurons $\boldsymbol{v}$, interacting recurrently through fixed synaptic matrix $\boldsymbol{M}$ (weak competition-cooperation). Plasticity and learning occurs only in the feedforward matrix $\boldsymbol{W}$:



We expect plasticity to depend directly on *correlations* between input and output neurons, but indirectly on both *correlations* between input neurons and *competition-cooperation* between output neurons.

1 Recurrent connections among output neurons

Consider a *linear* network in which N_v output neurons receive recurrent input from each other and feedforward input from N_u input neurons.

$$\tau_r \frac{d\boldsymbol{v}}{dt} = -\boldsymbol{v} + \boldsymbol{W} \cdot \boldsymbol{u} + \boldsymbol{M} \cdot \boldsymbol{v}$$

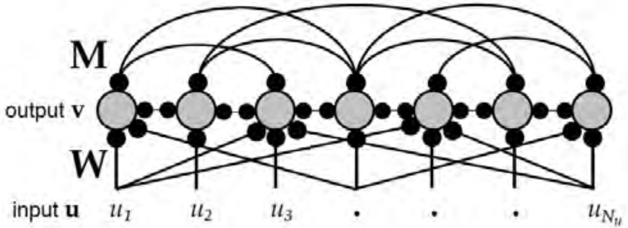


Figure 8.6: A network with multiple output units driven by feedforward synapses with weights **W**, and interconnected by recurrent synapses with weights **M**.

A steady-state solution for

$$\boldsymbol{v}_{ss} = \boldsymbol{W} \cdot \boldsymbol{u} + \boldsymbol{M} \cdot \boldsymbol{v}_{ss}$$

exists if there exists an inverse matrix $\boldsymbol{K} \equiv (\boldsymbol{I} - \boldsymbol{M})^{-1}$ to solve

$$\begin{aligned}
(\boldsymbol{I} - \boldsymbol{M}) \cdot \boldsymbol{v}_{ss} &= \boldsymbol{W} \cdot \boldsymbol{u} \\
\boldsymbol{v}_{ss} &= (\boldsymbol{I} - \boldsymbol{M})^{-1} \cdot \boldsymbol{W} \cdot \boldsymbol{u} \\
\boldsymbol{v}_{ss} &= \boldsymbol{K} \cdot \boldsymbol{W} \cdot \boldsymbol{u}
\end{aligned}$$

where $\boldsymbol{I}$ is the identity matrix. This is the case if the real parts of the eigenvalues of $\boldsymbol{M}$ are smaller than 1 (i.e., if recurrent interactions are comparatively weak).

With fixed recurrent weights $\boldsymbol{M}$ and plastic feedforward weights $\boldsymbol{W}$, the Hebbian learning rule becomes

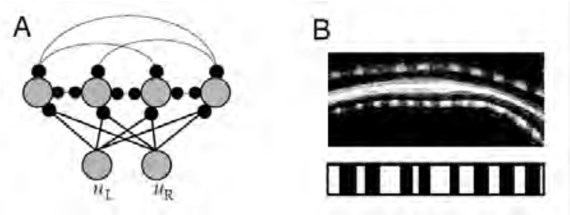
$$\tau_w \frac{d\boldsymbol{W}}{dt} = \langle \boldsymbol{v}\,\boldsymbol{u} \rangle = \boldsymbol{K} \cdot \boldsymbol{W} \cdot \boldsymbol{Q}, \qquad \boldsymbol{Q} \equiv \langle \boldsymbol{u}\,\boldsymbol{u} \rangle, \qquad \boldsymbol{K} \equiv (\boldsymbol{I} - \boldsymbol{M})^{-1}$$

As expected, the development is influenced not only by input correlations $\boldsymbol{Q}$, but also by the recurrent connectivity that shapes $\boldsymbol{K}$.

Role of recurrent connectivity

Without recurrent connectivity, the inputs weights $\boldsymbol{w}_i$ to each output neuron v_i would develop independently, reflecting (in part) the random initial conditions of each neuron.

With recurrent connectivity, a structured cortical map can develop, minimizing differences between neighboring neurons. For example, in the case of ocular dominance, alternating 'bands' of neurons with left- and right-eye dominance can develop.



A: Two input units (left and right eye) and four output units (visual cortex).
B: Ocular dominance columns in area 17 of kitten (top) and schematic illustration of alternating dominance (bottom).

1 Points to note

- Recurrent connectivity in the output layer alters Hebbian learning rules.
- The correlation rule for feedforward weights now reads

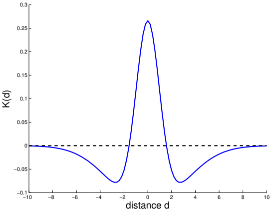
$$\tau_w \frac{d\boldsymbol{W}}{dt} = \langle \boldsymbol{v}\, \boldsymbol{u} \rangle = \boldsymbol{K} \cdot \boldsymbol{W} \cdot \boldsymbol{Q}$$

- Both input correlations $\boldsymbol{Q}$ and output interactions $\boldsymbol{K}$ are important.
- This permits the development of systematic patterns of feedforward connectivity.

2 Development of OD bands

We define recurrent cooperation-competition indirectly in terms of $\boldsymbol{K}$ (rather than directly in terms of $\boldsymbol{M}$) and assume

- translational invariance over cortical locations
- mutual excitation between adjacent units (cooperation)
- mutual inhibition between more distant units (competition)

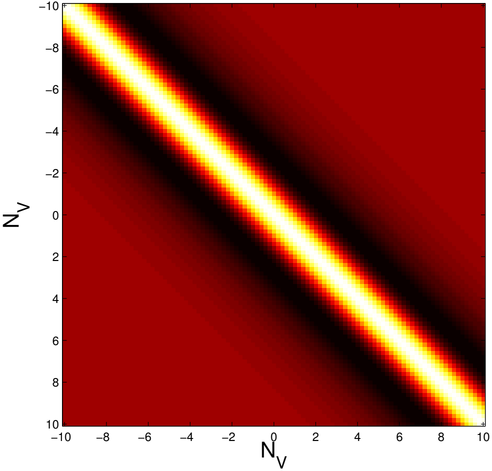


Specifically, we assume a 'difference-of-Gaussian' (DOG) connectivity

$$K(d) = \frac{1}{\sqrt{2\pi}\,\sigma_E} \exp\left(-{d^2}/{2\sigma_E^2}\right) - \frac{1}{\sqrt{2\pi}\,\sigma_I} \exp\left(-{d^2}/{2\sigma_I^2}\right)$$

where d is cortical distance and $\sigma_I \approx 3\,\sigma_E$.

Recurrent cooperation-competition matrix *K*

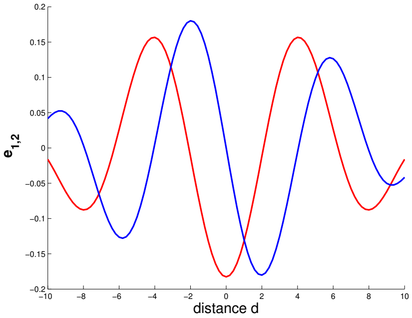


'Heat map' illustration of positive (yellow-white), negative (red-black), and neutral values (red).

Eigenvectors 1 and 2

The principal eigenvectors of the matrix $\boldsymbol{K}$ are pairs of even and odd harmonic functions of cortical distance d

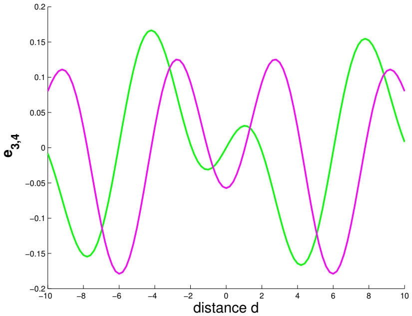
$$\boldsymbol{e}_{1,2} = \cos\left(\frac{2\pi d}{N_v} - \phi\right), \qquad \phi \in \left[0, \frac{\pi}{2}\right], \qquad \lambda_{1,2} = 3.2$$



Eigenvectors 3 and 4

The next eigenvectors of the matrix $\boldsymbol{K}$ are also pairs of even and odd harmonic functions of cortical distance

$$\boldsymbol{e}_\mu = \cos\left(\frac{4\pi d}{N_v} - \phi\right), \qquad \phi \in \left[0, \frac{\pi}{2}\right], \qquad \lambda_{3,4} = 2.8$$

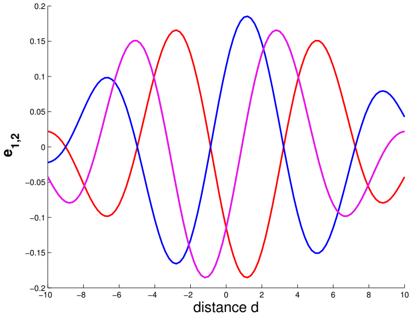


Principal eigenvectors

The two principal eigenvectors are degenerate (have identical eigenvalues) and resemble harmonic functions with alternating bands of positive and negative interactions between output units.

Any linear combination will also be an eigenvector

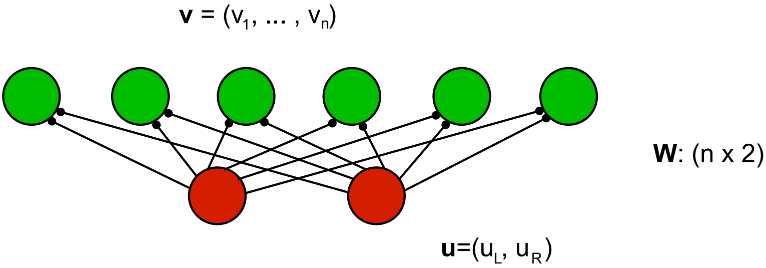
$$\boldsymbol{K} \cdot \boldsymbol{e} = \lambda\, \boldsymbol{e}, \qquad \boldsymbol{e} = \alpha\, \boldsymbol{e}_1 + (1-\alpha)\boldsymbol{e}_2$$



Formal treatment of development

We now formalize the development of OD bands in the primary visual cortex of stereoptic mammals. All output units are stimulated by the same two input units u_R and u_L

$$\boldsymbol{u} = \begin{pmatrix} u_R \\ u_L \end{pmatrix}, \qquad \boldsymbol{v} = \boldsymbol{K} \cdot \boldsymbol{W} \cdot \boldsymbol{u}, \qquad \boldsymbol{K} \equiv (\boldsymbol{I} - \boldsymbol{M})^{-1}$$



We assume a dynamic equation for weights:

$$\tau_w \frac{d\boldsymbol{W}}{dt} = \boldsymbol{K} \cdot \boldsymbol{W} \cdot \boldsymbol{Q} - \frac{1}{2}\boldsymbol{n}^T \cdot \boldsymbol{K} \cdot \boldsymbol{W} \cdot \boldsymbol{Q}, \qquad \boldsymbol{n} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

corresponding to a Hebbian rule with threshold and with subtractive normalization (for simplicity).

The matrix dimensions (rows $\times$ columns) are

$$\boldsymbol{K} : n \times n \qquad \boldsymbol{W} : n \times 2 \qquad \boldsymbol{Q} : 2 \times 2$$

Non-zero mean input

We further assume non-zero mean input with partially anti-correlated statistics

$$\boldsymbol{Q} = \begin{pmatrix} q_S & q_D \\ q_D & q_S \end{pmatrix} = \begin{pmatrix} 1/2 & 1/8 \\ 1/8 & 1/2 \end{pmatrix}$$

This is the statistics for which we were *unable* to obtain OD bands in Lecture 7!

When $\boldsymbol{W}$ is written out in terms of its two columns $\boldsymbol{w}_R$ and $\boldsymbol{w}_L$

$$\boldsymbol{W} = \begin{pmatrix} w_{1R} & w_{1L} \\ \vdots & \vdots \\ w_{nR} & w_{nL} \end{pmatrix} = \begin{pmatrix} \boldsymbol{w}_R & \boldsymbol{w}_L \end{pmatrix}$$

the dynamic equations become

$$\tau_w \frac{d\boldsymbol{w}_R}{dt} = \boldsymbol{K} \cdot (q_S \boldsymbol{w}_R + q_D \boldsymbol{w}_L) - \frac{1}{2} \boldsymbol{K} \cdot (q_S \boldsymbol{w}_R + q_D \boldsymbol{w}_L + q_D \boldsymbol{w}_R + q_S \boldsymbol{w}_L)$$

$$\tau_w \frac{d\boldsymbol{w}_L}{dt} = \boldsymbol{K} \cdot (q_D \boldsymbol{w}_R + q_S \boldsymbol{w}_L) - \frac{1}{2} \boldsymbol{K} \cdot (q_S \boldsymbol{w}_R + q_D \boldsymbol{w}_L + q_D \boldsymbol{w}_R + q_S \boldsymbol{w}_L)$$

Using the same trick as in Exercise 3, we introduce two vectors $\boldsymbol{w}_+$ (sum of weights) and $\boldsymbol{w}_-$ (difference of weights):

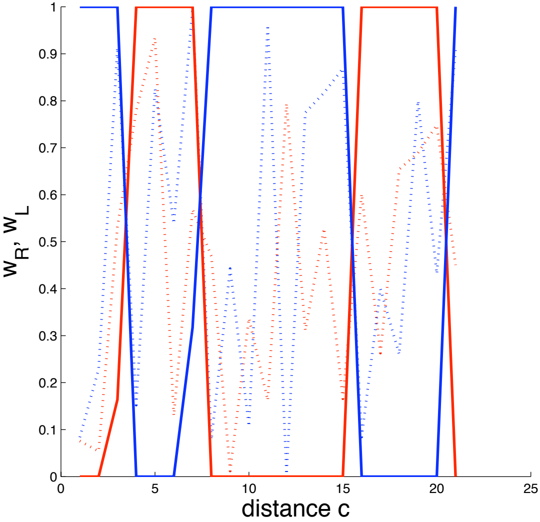
$$\boldsymbol{w}_+ = \boldsymbol{w}_R + \boldsymbol{w}_L \qquad\qquad \boldsymbol{w}_- = \boldsymbol{w}_R - \boldsymbol{w}_L$$

In terms of these aligned weight vectors, the dynamic equations are

$$\begin{aligned} \tau_w \frac{d\boldsymbol{w}_+}{dt} &= 0 \\ \tau_w \frac{d\boldsymbol{w}_-}{dt} &= \boldsymbol{K} \cdot (q_S - q_D)\, \boldsymbol{w}_- \end{aligned}$$

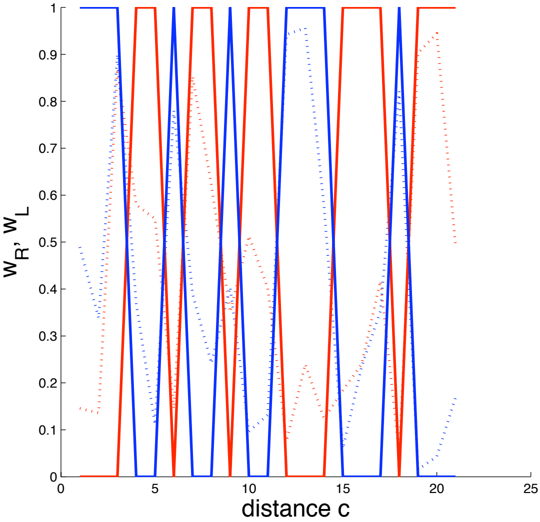
Thus the development of $\boldsymbol{w}_-$ is governed by the principal eigenvectors of $\boldsymbol{K}$!

Accordingly, development results in broad, alternating bands of ocular dominance, smoothing over initial conditions:



Final weights are shown solid, initial weights dotted. Left and right weights are shown red and blue.

Without recurrent connections, development would simply amplify the random initial conditions:



Final weights are shown solid, initial weights dotted. Left and right weights are shown red and blue.

Show animation
Credit: Armin Maddah, Year 2017

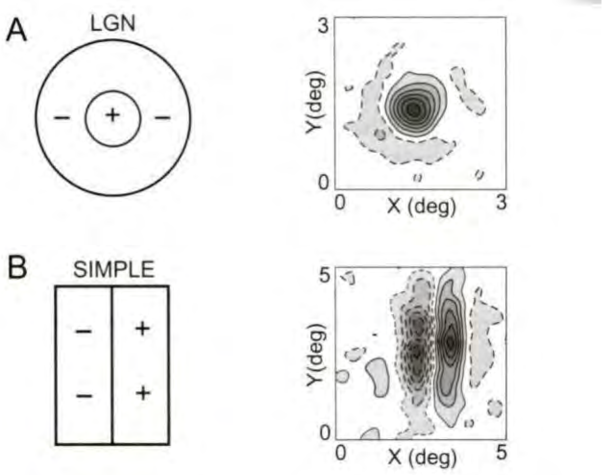
2 Points to note

- For realistic inputs $\boldsymbol{u}$ (with non-zero mean), the correlation matrix is $\boldsymbol{Q}$ is dominated by the 'same' eigenvector.
- This is why, in Lecture 7, we failed to obtain the desired outcome of synaptic weights aligned with the 'different' eigenvector.
- We have now overcome this difficulty by introducing cooperation-competition (difference-of-Gausians) within the output layer.
- Now the development of feedforward weights is driven by a combination of input correlations $\boldsymbol{Q}$ and output interactions $\boldsymbol{K}$.
- This is a plausible model for developing so-called "functional anatomy", such as ocular dominance bands, in sensory cortices of mammals.

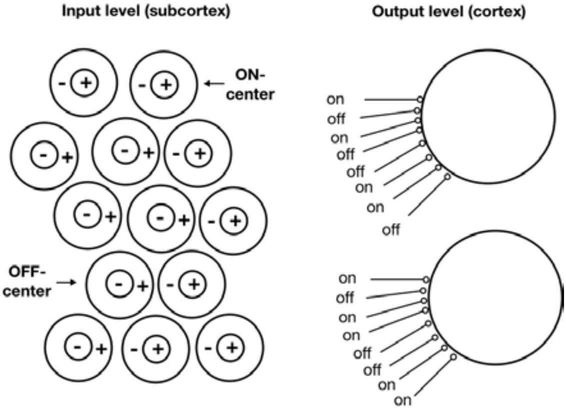
3 Computational models of visual development

Simple cell receptive fields

Cells in visual cortex have receptive fields with alternating 'on-' and 'off-center' inputs. From initially indiscriminate projections, such receptive fields can develop by *unsupervised Hebbian learning*, which selectively strengthens and prunes subcortical projections.

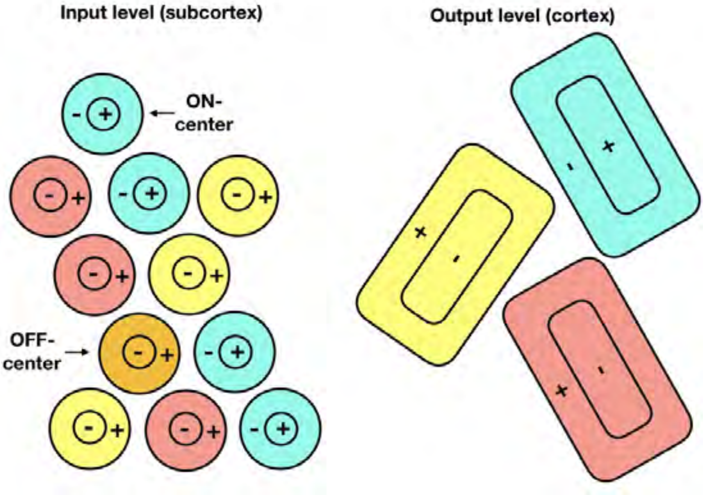


Initially indiscriminate projections



Cortical cells initially receive indiscriminate projections from subcortical cells, both on- and off-center.

Development of orientation selective projections

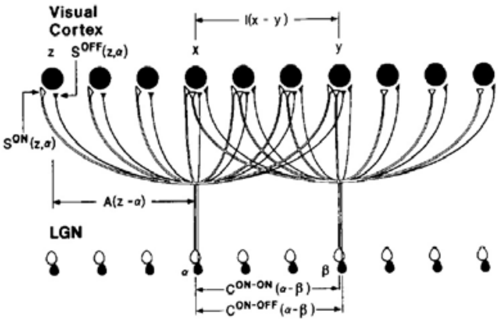


During development, some projections are strengthened and others pruned, selecting correlated activity in natural image responses.

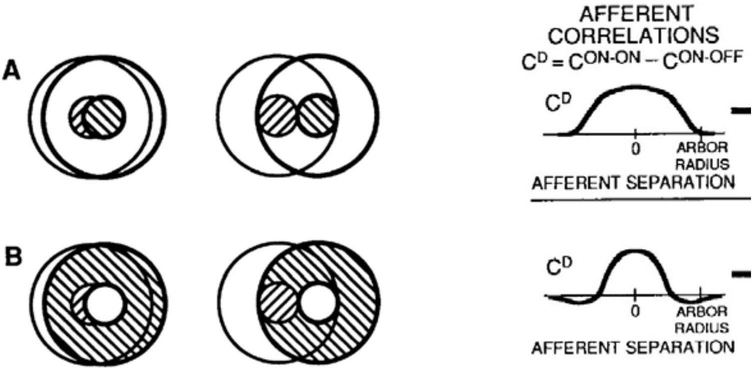
Miller (1994) Orientation-tuned receptive fields

Two initially equivalent input projections (ON- and OFF-center subcortical cells) compete to control a single output layer (cortical simple cells). Study development of orientation selective RFs.

$S^{ON}(z,\alpha), S^{OFF}(z,\alpha)$	**plastic feedforward connectivity**
$A(z-\alpha)$	**fixed arbor attenuation**
$I(x-y)$	**fixed recurrent connection**
$C^{ON,ON}(\alpha-\beta), C^{OFF,OFF}(\alpha-\beta), C^{ON,OFF}(\alpha-\beta)$	**input correlations**

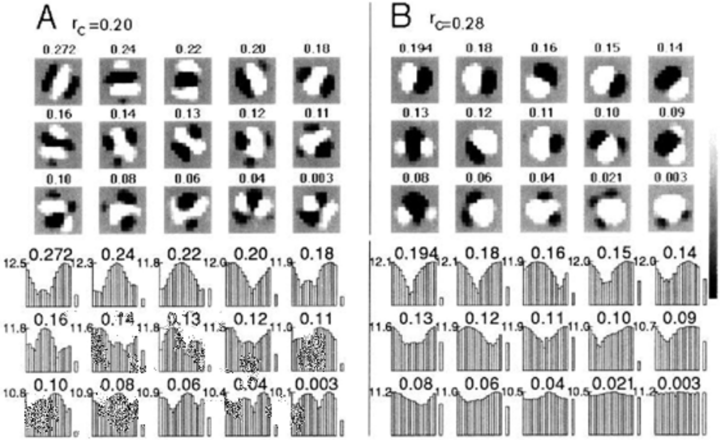


Separation-dependent correlation of ON- and OFF-inputs



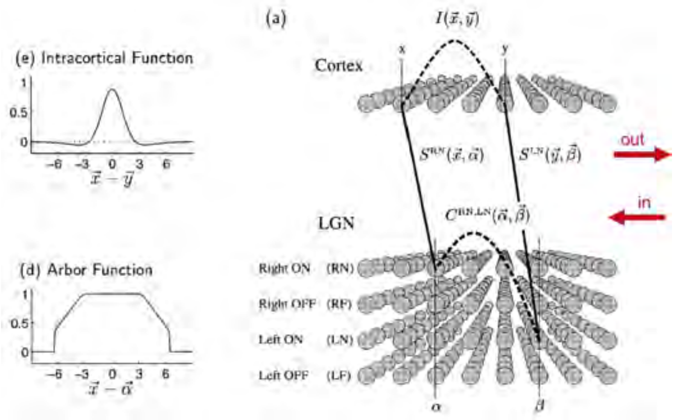
Center-surround receptive fields predict positive or negative correlations in natural image responses, depending on separation of RF centers. Difference C^D between C^{ON-ON} and C^{ON-OFF} may show different dependencies on center-to-center separation.

Development of orientation selectivity



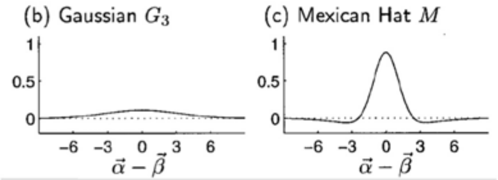
Orientation selective RF develop from differential input correlations (ON-ON vs. ON-OFF) and output recurrent interactions (competition-cooperation). Selectivity reflects natural image statistics!

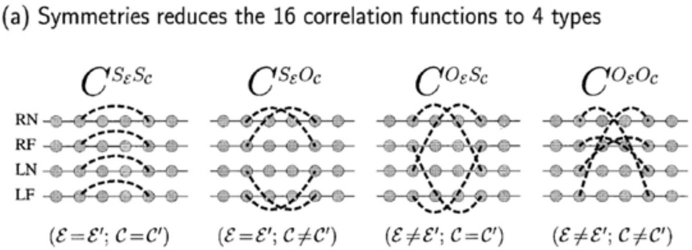
Erwin & Miller (1998) Orientation tuning & ocular dominance



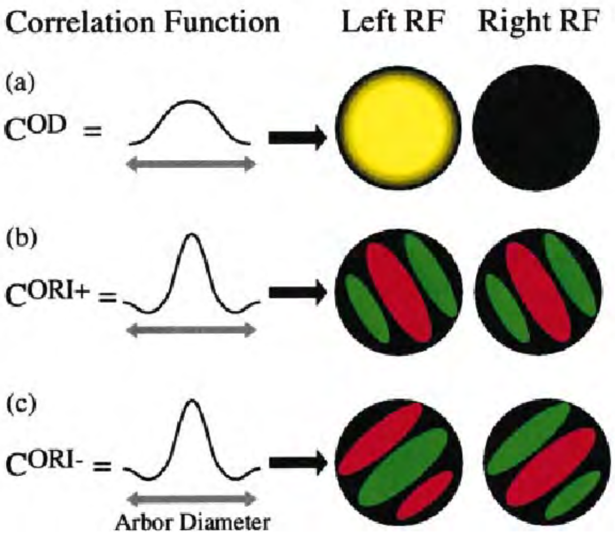
Can correlation-driven plasticity explain the concurrent development of both orientation tuning and ocular dominance?

Assume feedforward arborization, intracortical interactions, and input correlations:

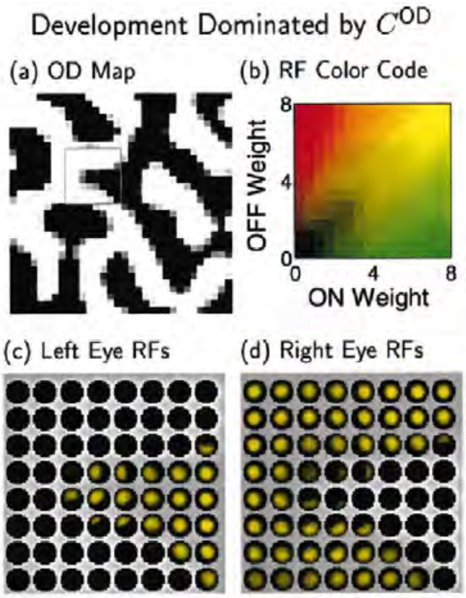




Right, **L**eft, o**N**, o**F**, **S**ame, **O**ther, **E**ye, **C**enter



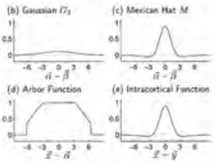
Assumed correlations and the resulting receptive fields.

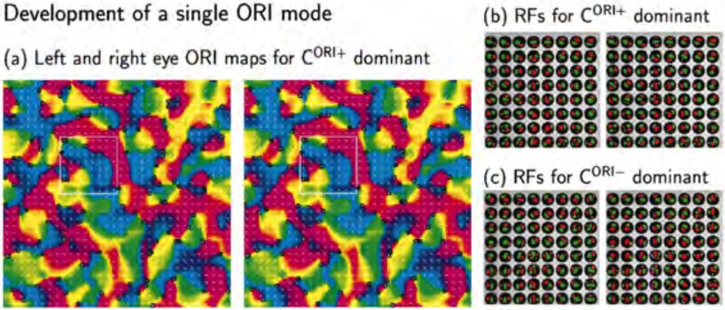


With stronger eye-eye correlations, ocular dominance develops.

Development dominated by ORI

$$C^{ORI+} = r^+ M$$

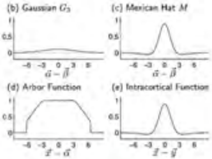


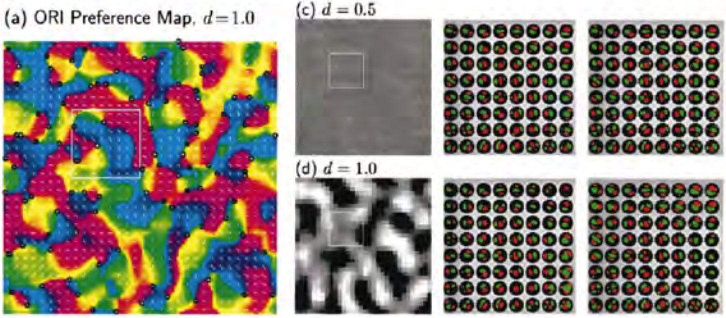


With stronger on-off correlations, orientation tuning develops.

Combined development

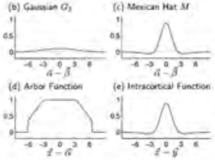
$$C^{OD} = d\, G_3$$

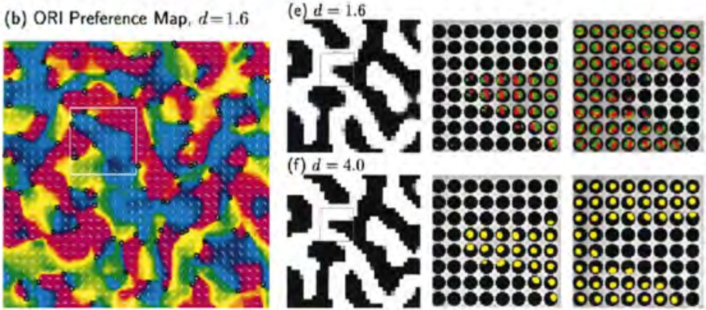




Modulate relative strength of eye-eye and on-off correlations.

$$C^{OD} = d\,G_3$$





Modulate relative strength of eye-eye and on-off correlations.

Two stage development

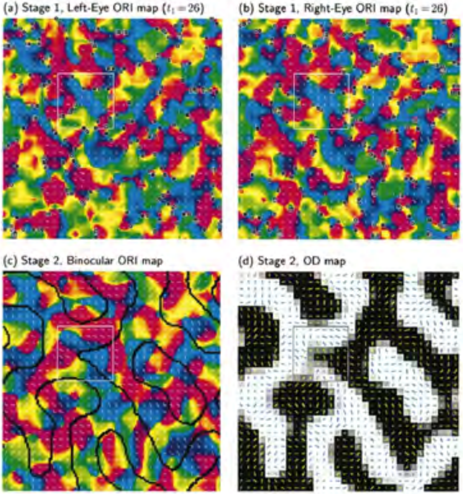
ON vs OFF

$$C^{ORI+} = r^{+} M$$

left vs right

$$C^{OD} = d\, G_3$$

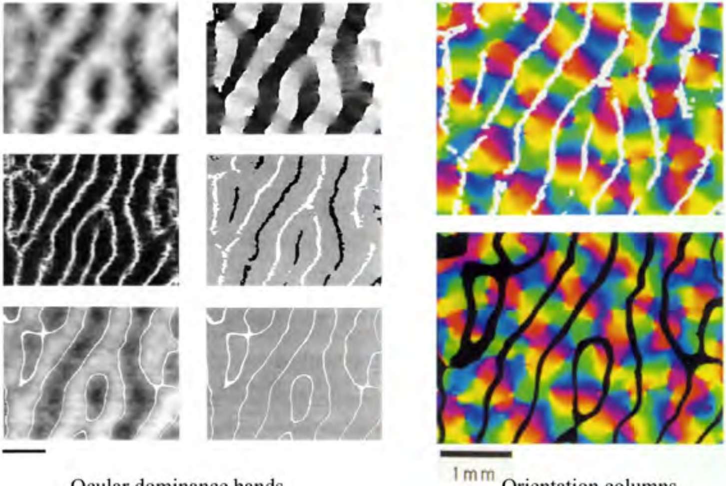
$$C^{ORI+} = r^{+} M$$



Successive phases first with on-off correlations, later with eye-eye correlations.

Actual functional organization of primate visual cortex

Orientation columns and ocular dominance bands (Blasdel, 1992) in macaque. Preferred orientations represented by color.



Ocular dominance bands Orientation columns

Conclusion

The development of *ocular dominance bands* requires

- Same eye inputs more correlated than different eye inputs, especially at small separations ($C^{OD} = dG_3$).
- Holds for aggregate of ON- and OFF-center inputs.

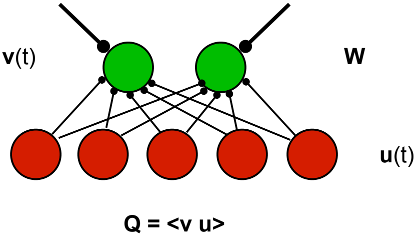
The development of *orientation columns* requires

- Same center inputs more correlated than different center inputs at small separations, and less correlated at larger separations ($C^{ORI} = M$).
- Holds for aggregate of left and right eye inputs.

Combined development is obtained more easily sequentially than simultaneously.

4 Supervised Hebbian learning

In supervised learning, both input activity $\boldsymbol{u}$ and output activity $\boldsymbol{v}$ is imposed externally. Synaptic projections $\boldsymbol{W}$ develop such as to reproduce correlations $\boldsymbol{Q} = \langle \boldsymbol{v}\boldsymbol{u} \rangle$ in imposed activities. In this 'training' phase, synaptic weights do *not affect* output activity.



After training, synaptic projections $\boldsymbol{W}$ map input activity $\boldsymbol{u}$ to output activity $\boldsymbol{v}$ in such a way as to *reproduce* the correlations $\boldsymbol{Q}$ experienced earlier. Only in this 'post-training' phase do projections determine output activity, such that $\boldsymbol{v} = \boldsymbol{W} \cdot \boldsymbol{u}$.

4.1 Learning rule

An (unstable) Hebbian rule for supervised learning of a single output neuron is

$$\tau_w \frac{d\boldsymbol{w}}{dt} = \langle v\, \boldsymbol{u} \rangle$$

where v is single output activity and $\boldsymbol{u}$ multiple input activities imposed externally during training. As neither v nor $\boldsymbol{u}$ depends on $\boldsymbol{w}$ (connections are still being trained!), the unstable solution is

$$\boldsymbol{w}_{(t)} = \boldsymbol{w}_0 + \langle v\, \boldsymbol{u} \rangle \ {}^{t}\!/_{\tau_w}$$

In other words, synaptic weights grow linearly with time, with growth rates proportional to (positive or negative) correlations $\langle v \boldsymbol{u} \rangle$.

Stabilization

This Hebbian rule may be stabilized by a 'relaxation' term, providing for spontaneous 'decay' of synaptic weights

$$\tau_w \frac{d\boldsymbol{w}}{dt} = -\boldsymbol{w} + \langle v\, \boldsymbol{u} \rangle$$

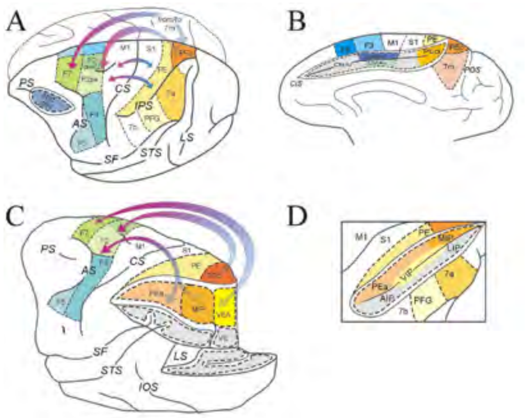
with the steady-state solution

$$\boldsymbol{w}_{ss} = \langle v\, \boldsymbol{u} \rangle$$

In other words, synaptic weights develop such as to become proportional to input- output correlations $\langle v\, \boldsymbol{u} \rangle$. After training, when synaptic projections become effective, the effect will tend to *reproduce* the correlations experienced earlier during training.

Next we use this mechanism to 'train' arbitrary input-output transformations, or 'functional mappings'.

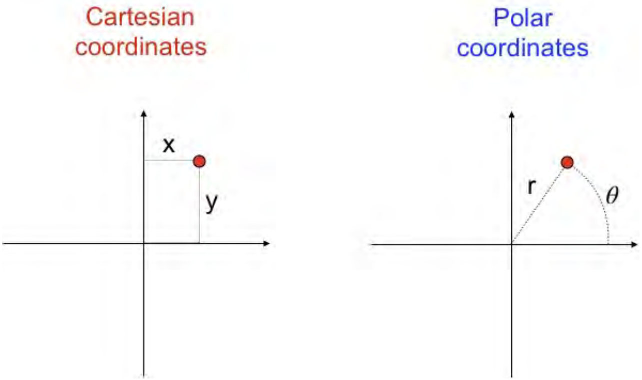
A biological example is the transformation from eye to hand coordinates, between posterior parietal cortex and premotor areas in frontal cortex:



To demonstrate generality (and also for simplicity), we consider an artificial example: transforming Cartesian to polar coordinates.

Cartesian and polar coordinates

Recall that position in a plane may be specified either by Cartesian coordinates (x, y), or by polar coordinates (r, θ) with radius r and angle θ:

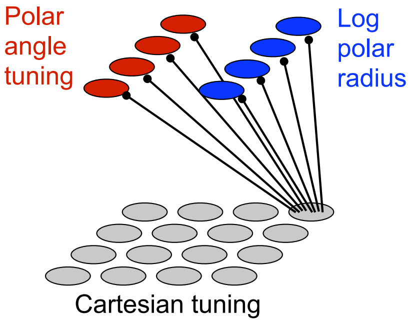


$$\theta = \arctan \frac{y}{x}, \qquad r = \sqrt{x^2 + y^2}$$

Cartesian $\rightarrow$ polar with neural populations

We assume a two-dimensional input population with index i, tuned in Cartesian coordinates (x, y), projecting to two one-dimensional output populations with indices j and k, tuned in polar angle θ and log polar radius $\ln r$, respectively:

$$\theta = \arctan \frac{y}{x}, \qquad \ln r = \ln \sqrt{x^2 + y^2}$$



Training phase: noisy responses i, j, and k

For many (hundreds or thousands) random stimulus positions (x, y, r, θ), we compute input responses R_i^c and output responses R_j^θ and R_k^r.

- Input responses show Gaussian tuning – preferred position (x_i, y_i) and width σ_{xy} – with noise η_i:

$$R_i^c \quad = \quad \mathcal{N}\left(|x_i - x|, |y_i - y|, \sigma_{xy}\right) + \eta_i$$

- Output responses show Gaussian – preferred position θ_j or $\ln r_k$, and width σ_θ or σ_r – with noise η_j or η_k:

$$\begin{aligned} R_j^\theta &= G\left(|\theta - \theta_j|, \sigma_\theta\right) + \eta_j \\ R_k^r &= G\left(|\ln r - \ln r_k|, \sigma_r\right) + \eta_k \end{aligned}$$

where $\mathcal{N}(\mu, \sigma)$ and $\mathcal{N}(\mu_x, \mu_y, \sigma)$ are normal distributions.

Training phase: correlations between noisy responses

- Next, we compute the covariance of noisy responses for every pair of pre- and postsynaptic neurons $i \to j$ and for $i \to k$: such as to collect the resulting values into covariance matrices C_{ji} and C_{ki}:

$$\begin{aligned} C_{ji} &= \langle R_j^\theta \, R_i^c \rangle - \langle R_j^\theta \rangle \, \langle R_i^c \rangle \\ C_{ki} &= \langle R_k^r \, R_i^c \rangle - \langle R_k^r \rangle \, \langle R_i^c \rangle \end{aligned}$$

- Finally, we set feedforward projection matrices W_{ji} and W_{ki} equal to the covariance matrices C_{ji} and C_{ki}:

$$\begin{aligned} W_{ji} &= C_{ji} \\ W_{ki} &= C_{ki} \end{aligned}$$

- We have thereby 'trained' synaptic weights to *reproduce* the input-output correlations experienced during training.

Testing phase: noisy responses i, j, and k

For random stimulus positions (x, y, r, θ), we compute input responses R_i^c as before. However, unlike before, we compute responses R_j^θ and R_k^r as non-linear functions of total synaptic input.

- Input responses (as before):

$$R_i^c = f\left(|x_i - x|, |y_i - y|\right) + \eta_i$$

- Output responses (new) as feedforward many-to-many projections, analogous to $\boldsymbol{v} = F(\boldsymbol{W} \cdot \boldsymbol{u})$:

$$R_j^\theta = \left[\sum_i W_{ji} \cdot R_i^c\right]_+$$

$$R_k^r = \left[\sum_i W_{ki} \cdot R_i^c\right]_+$$

with rectification $[]_+$ as activation function $F()$.

Testing phase: decoding noisy responses

To verify whether responses of output population correctly encode stimulus position, we compare the correct polar coordinates coordinates $(\ln r_{true}, \theta_{true})$ of a stimulus (x, y)

$$\theta_{true} = \arctan \frac{y}{x} \qquad \ln r_{true} = \ln \sqrt{x^2 + y^2}$$

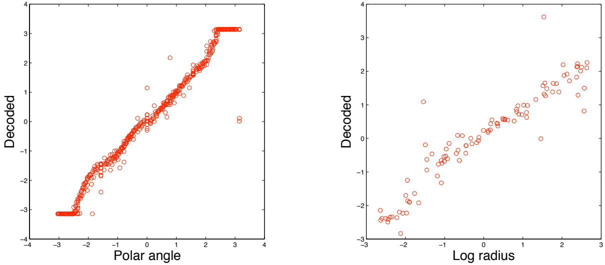
with the polar-coordinates decoded from noisy output responses (by response-weighted averaging):

$$\theta_{decoded} = \frac{\sum_j R_j^\theta \, \theta_j}{\sum_j R_j^\theta} \qquad \ln r_{decoded} = \frac{\sum_k R_k^\theta \ln r_k}{\sum_k R_k^\theta}$$

Recall response-weighted averaging from ML decoding: average over preferred positions weighed by associated responses.

Testing phase: results

Comparison of correct values (horizontal axes) and decoded values (vertical axes)



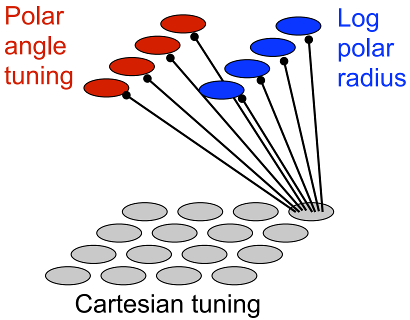
confirmes that the desired ‘functional mapping’ was obtained by supervised Hebbian learning.

4 Points to note

- We found that an arbitrary 'functional mapping' may be accomplished by supervised Hebbian learning.
- As an artificial example, we transformed neural responses from Cartesian to polar coordinates.
- In a training phase, both input and output responses were imposed directly by (random) external stimulation, and covariances established for all input-output pairs. Feedforward projection strengths were set equal to covariance values.
- In testing phase, only input response were due to stimulation. Output responses were determined by feedforward projection strengths.
- To validate results, noisy output responses were decoded by response-weighted averaging.

ctd

- Supervised Hebbian learning can realize arbitrary coordinate transforms (eye-to-head, head-to-hand, eye-to-hand, etc.)
- Supervised learning requires a training period, during which the correlations to-be-learned occur spontaneously.



Next: Reinforcement learning