

# Lecture 6 Hebbian plasticity

Jochen Braun

April 19, 2006

### Goals:

Basic principles and problems of Hebbian learning

### Credits:

Dayan & Abbott, Chapter 8

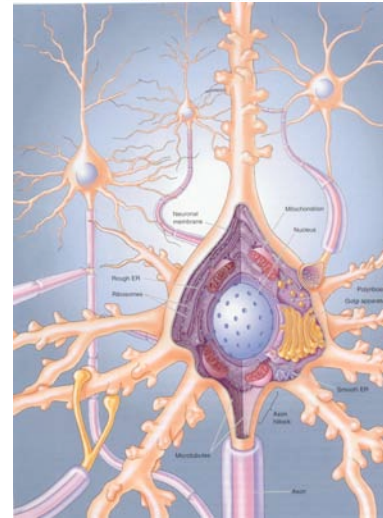
Obermayer & Blasdel (1993) Geometry of orientation and ocular dominance columns in monkey striate cortex. *J Neurosci* 13: 4114-4129.

Miller (1994) A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. *J Neurosci* 14: 409-441.

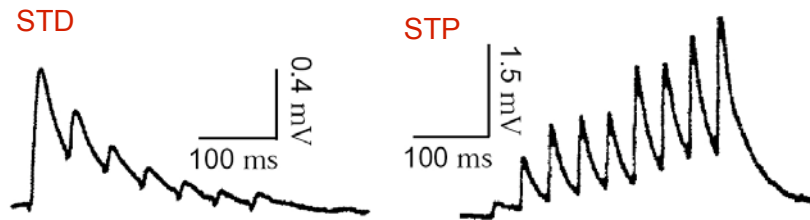
Erwin & Miller (1998) Correlation-based development of ocularly matched orientation and ocular dominance maps: determination of required input activities. *J Neurosci* 18: 9870-9895.

## 1 Cellular basis of learning

Synaptic weights are not fixed but changed by neuronal activity. This "plasticity" is thought to be the basis of learning and memory.

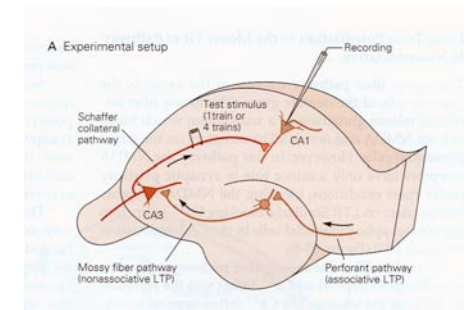
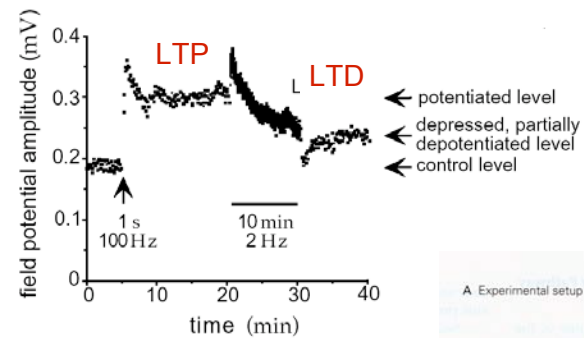


Activity  $\leftrightarrow$  Plasticity

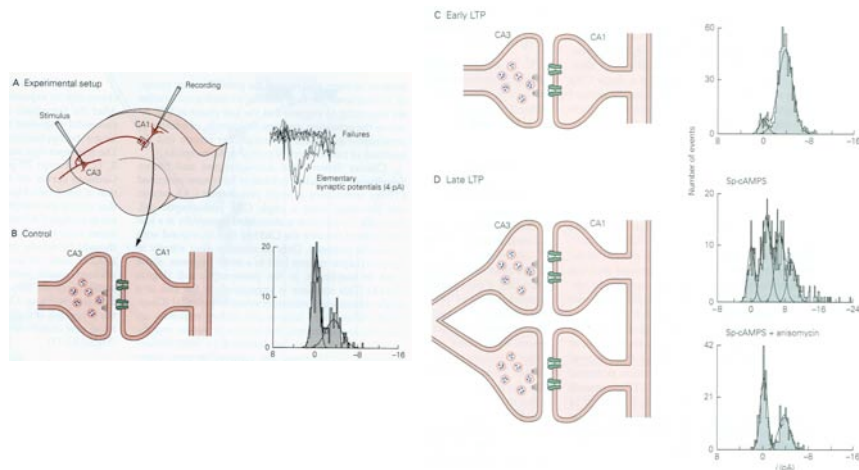


Short-term depression and short-term potentiation reflect recent activity (last few 100 ms).

Long-term potentiation and long-term depression reflect activity long ago (from 10 min to a lifetime).

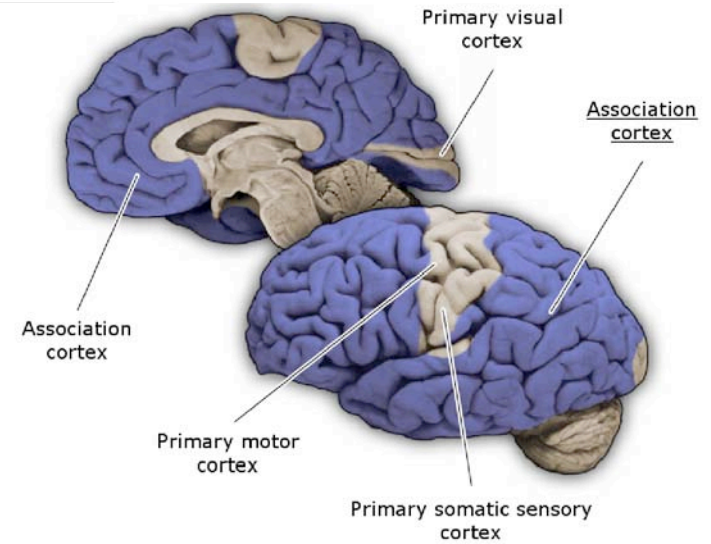


Early LTP increases the probability of release at existing sites, late LTP adds additional sites.

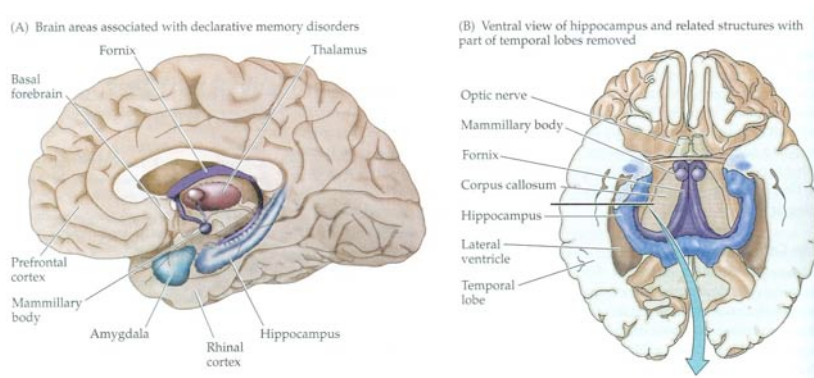


## 2 Brain circuits of learning

Learning occurs all over the brain, especially in association cortices.



The hippocampal formation is needed to implant **declarative memories** (things that you know that you know).

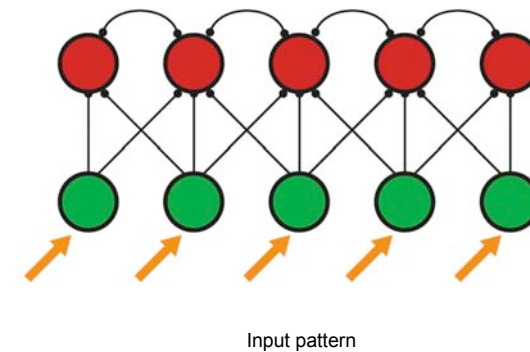


In Patient H.M., the hippocampal formation on both sides was removed surgically (to treat epilepsy). H.M. woke up with severe amnesia. While he remembered his earlier life, he was unable to recall subsequent events in daily life. His perception, abstract thinking, or reasoning was unaffected. Similarly, he could learn new skills such as mirror writing or puzzle solving. As soon as his attention turns to something new, all memory of recent activities is lost.

## 3 Formalizing learning

### 3.1 Unsupervised learning

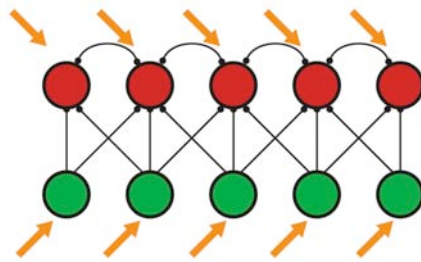
The network responds to training stimuli according to its internal connectivity and dynamics. Any learning constitutes self-organisation in the face of the stimulation imposed.



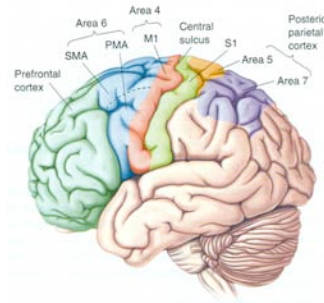
### 3.2 Supervised learning

Input activity is imposed by a stimulus and output activity by a "teacher" network. Feed-forward connections between parietal and premotor cortex are an example (Lecture 2).

Desired output pattern from "teacher" network



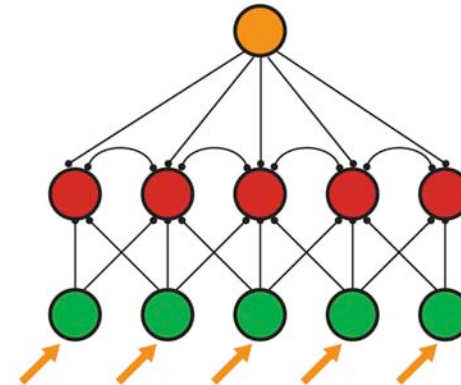
Input pattern



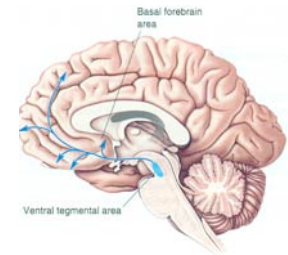
### 3.3 Reinforcement learning

A reinforcement signal (reward if positive, punishment if negative) is propagated by a "teacher" unit within the network. Dopamine neurons in the ventral tegmental area may constitute such a "teacher" unit.

Reward/punishment



Input pattern



## 4 Hebbian and non-Hebbian plasticity

### 4.1 Donald Hebb



Working at the Yerkes Primate Lab in Florida, Donald O Hebb published his ground-breaking book "The Organization of Behaviour: A Neuropsychological Theory" in 1949. Combining data about brain and behaviour into a single theory, he proposed fundamental principles for the interaction between neuronal activity and synaptic plasticity, which became known as **Hebbian learning**:

*When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased*

The modern paraphrase goes

*Neurons that fire together wire together*

Increases in synaptic strength must be compensated by decreases elsewhere. Thus, failure of A to trigger B should weaken the connection. In both cases, Hebbian plasticity *compares* pre- and post-synaptic firing.



### 4.2 Non-Hebbian plasticity

In neuronal adaptation, activity modifies excitability and response properties of neurons. In *non-Hebbian*, pre- or post-synaptic activity alone modifies synaptic weights (e.g., STD, STP).

To avoid uncontrolled growth of synaptic strength, we need to assume constraints on the individual or collective weight of synapses:

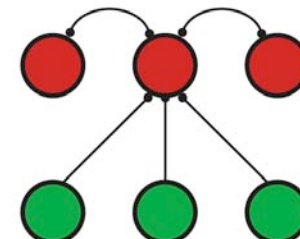
#### Synaptic saturation

$$0 \leq w \leq w_{max}$$



#### Synaptic competition

$$\sum_i w_i \leq W_{max}$$



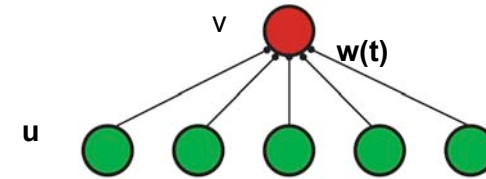
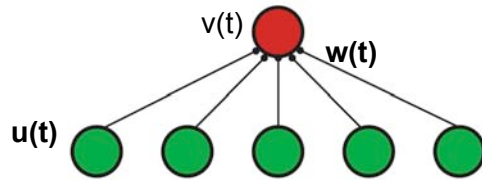
## 5 Synaptic plasticity rules

Consider a single postsynaptic neuron driven by  $N_u$  presynaptic inputs with activities  $u_b$  for  $b = 1, 2, \dots, N_u$ . We use a linear version of the firing rate model, with  $w_b$  representing the synaptic weight of the connection from presynaptic neuron  $b$ , which can be positive (excitation) or negative (inhibition).

$$\tau_v \frac{dv}{dt} = -v + \sum_{b=1}^{N_u} w_b u_b = -v + \mathbf{w} \cdot \mathbf{u}$$

We now need to specify how the  $w_b$  change as a joint function of presynaptic activity  $u_b$  and postsynaptic activity  $v$ . Unfortunately, this produces a coupled system of differential equations.

$$\tau_w \frac{dw}{dt} = f(v, \mathbf{u})$$



To simplify the situation, we focus on the dynamics of synaptic weights and assume that neuronal activity has time to reach equilibrium:

$$v = \sum_{b=1}^{N_u} w_b u_b = \mathbf{w} \cdot \mathbf{u}$$

$$\tau_w \frac{dw}{dt} = f(v, \mathbf{u})$$

### 5.1 Basic and averaged Hebb rule

The simplest implementation of Hebb's rule is

$$\tau_w \frac{dw}{dt} = v \mathbf{u} \quad \tau_w \frac{dw_b}{dt} = v u_b \quad b = 1, 2, \dots, N_u$$

To avoid overly rapid changes in  $\mathbf{w}$ , we can average over a set of input patterns (denoted by  $\langle \rangle$ ).

$$\begin{aligned} \tau_w \frac{d\mathbf{w}}{dt} &= \langle v \mathbf{u} \rangle = & \tau_w \frac{dw_k}{dt} &= \langle v u_k \rangle = \\ &= \langle \mathbf{w} \cdot \mathbf{u} \mathbf{u} \rangle = & &= \left\langle \sum_{b'=1}^{N_u} w_{b'} u_{b'} u_k \right\rangle = \\ &= \langle \mathbf{u} \mathbf{u} \rangle \cdot \mathbf{w} & &= \sum_{b'=1}^{N_u} \langle u_b u_{b'} \rangle w_{b'} \\ \mathbf{Q} &= \langle \mathbf{u} \mathbf{u} \rangle & Q_{bb'} &= \langle u_b u_{b'} \rangle \end{aligned}$$

where  $\mathbf{Q}$  is the autocorrelation matrix of the vector  $\mathbf{w}$ .

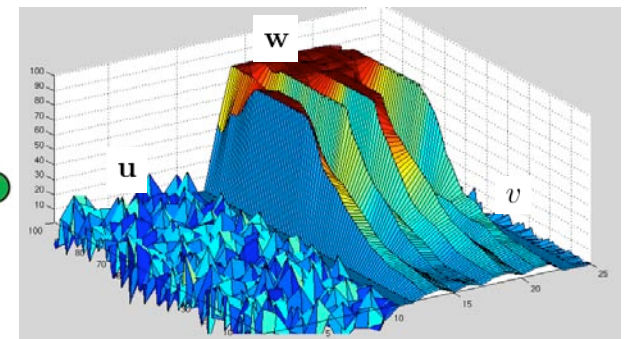
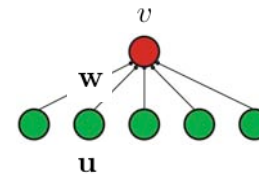
$$\begin{array}{c} \mathbf{v} \cdot \mathbf{u} \\ \mathbf{w} \cdot \mathbf{u} \end{array} = \begin{array}{c} \mathbf{u} \cdot \mathbf{u} \\ \mathbf{u} \cdot \mathbf{w} \end{array} = \mathbf{Q} \cdot \mathbf{w}$$

We already know that Hebb's rule is not stable and that synaptic weights grow beyond all bounds. To demonstrate this formally, consider

$$\begin{aligned} |\mathbf{w}|^2 &= \mathbf{w} \cdot \mathbf{w} = \sum_b w_b^2 \\ \tau_w \frac{d|\mathbf{w}|^2}{dt} &= 2 \tau_w \mathbf{w} \cdot \frac{d\mathbf{w}}{dt} = 2 \mathbf{w} \cdot v \mathbf{u} = 2 \mathbf{w} \cdot \mathbf{u} v = 2 v^2 > 0 \end{aligned}$$

A discrete version of the basic Hebb rule is

$$\mathbf{w} \rightarrow \mathbf{w} + \epsilon \mathbf{Q} \cdot \mathbf{w}$$



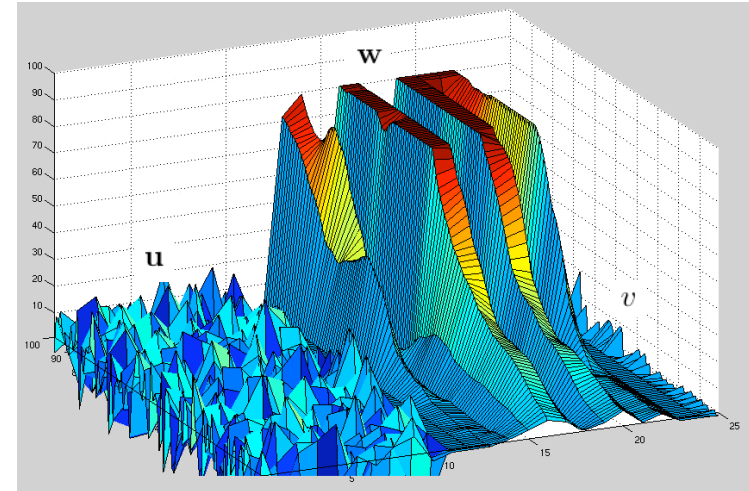
## 5.2 Covariance rule

The Hebb rule can be modified to ensure that synaptic strength may both increase and decrease, depending on pre- and postsynaptic activity. Specifically, we need to make sure that synaptic strength decreases when high presynaptic activity is associated with low postsynaptic activity. This can be achieved either by a postsynaptic threshold  $\theta_v$  or by a presynaptic threshold vector  $\theta_u$ :

$$\begin{aligned}\tau_w \frac{d\mathbf{w}}{dt} &= (v - \theta_v) \mathbf{u} & \theta_v &= \langle v \rangle \\ \tau_w \frac{d\mathbf{w}}{dt} &= v (\mathbf{u} - \theta_u) & \theta_u &= \langle \mathbf{u} \rangle\end{aligned}$$

The thresholds determine the level of activity above which LTD turns into LTP. A convenient choice of threshold is the average value of the corresponding variable. Both choices produce the same average rule, where  $\mathbf{C}$  is the covariance matrix of the input:

$$\begin{aligned}\tau_w \frac{d\mathbf{w}}{dt} &= \langle (\mathbf{w} \cdot \mathbf{u} - \langle \mathbf{u} \rangle) \mathbf{u} \rangle = \langle \mathbf{w} \cdot \mathbf{u} \mathbf{u} \rangle - \langle \mathbf{w} \cdot \mathbf{u} \rangle \langle \mathbf{u} \rangle = \mathbf{C} \cdot \mathbf{w} & \mathbf{C} &= \langle \mathbf{u} \mathbf{u} \rangle - \langle \mathbf{u} \rangle^2 \\ \tau_w \frac{d\mathbf{w}}{dt} &= \langle \mathbf{w} \cdot \mathbf{u} (\mathbf{u} - \langle \mathbf{u} \rangle) \rangle = \langle \mathbf{w} \cdot \mathbf{u} \mathbf{u} \rangle - \langle \mathbf{w} \cdot \mathbf{u} \rangle \langle \mathbf{u} \rangle = \mathbf{C} \cdot \mathbf{w}\end{aligned}$$



Although now synaptic weights both increase and decrease, they remain unstable due to positive feedback. For both rules, the time derivative of the sum of all squared weights becomes

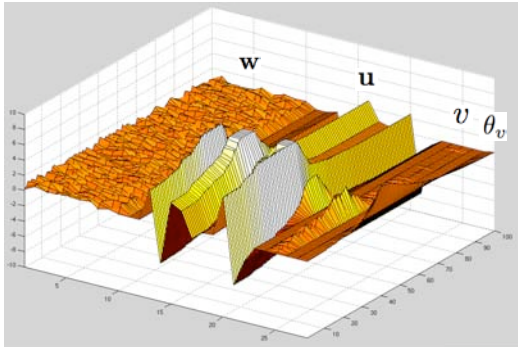
$$\frac{d|\mathbf{w}|^2}{dt} = 2v(v - \langle v \rangle) \quad \Rightarrow \quad \langle \frac{d|\mathbf{w}|^2}{dt} \rangle = 2(\langle v^2 \rangle - \langle v \rangle^2) > 0 \quad \text{if } v \neq \text{const}$$

## 5.3 BCM rule

Bienenstock, Cooper, and Munro (1982) suggested an alternative rule ("BCM rule") that changes synaptic weights depending on current pre- and postsynaptic activity, consistent with biological evidence. In contrast to the covariance rule, the BCM rule uses a sliding threshold which adapts to the current output activity:

$$\tau_w \frac{d\mathbf{w}}{dt} = v \mathbf{u} (v - \theta_v) \quad \tau_\theta \frac{d\theta_v}{dt} = v^2 - \theta_v$$

The threshold  $\theta_v$  must grow more rapidly than  $v$  with increasing output activity becomes large. This implements synaptic competition, because raising a global threshold weakens all synapses other than the strongest one.



## 5.4 Synaptic normalization

Instead of introducing synaptic normalization indirectly via the BCM rule, it is also possible to introduce an explicit constraint, for example on the total sum of synaptic weights (if they are all positive) or on the total sum of squared weights. The constraint can either be imposed rigidly (forcing each neuron to satisfy the constraint on every time step) or dynamically (requiring it to be satisfied only asymptotically). Such weight normalizations can drastically alter the outcome of a training procedure. Two examples of synaptic normalizations are the "subtractive normalization" and the "Oja rule":

$$\begin{aligned}\tau_w \frac{d\mathbf{w}}{dt} &= v \mathbf{u} - \frac{v(\mathbf{n} \cdot \mathbf{u}) \mathbf{n}}{N_u} & \text{subtractive normalization} \\ \tau_w \frac{d\mathbf{n} \cdot \mathbf{w}}{dt} &= v \mathbf{n} \cdot \mathbf{u} \left(1 - \frac{\mathbf{n} \cdot \mathbf{n}}{N_u}\right) = 0\end{aligned}$$

This rule has to be used with saturation constraints, to prevent weights from becoming negative (floor) or all weights but one from becoming zero (ceiling). This procedure is highly competitive, as small weights are reduced disproportionately.

$$\tau_w \frac{d\mathbf{w}}{dt} = v \mathbf{u} - \alpha v^2 \mathbf{w} \quad \text{Oja rule}$$

$$\tau_w \frac{d|\mathbf{w}|^2}{dt} = 2v^2 (1 - \alpha |\mathbf{w}|^2)$$

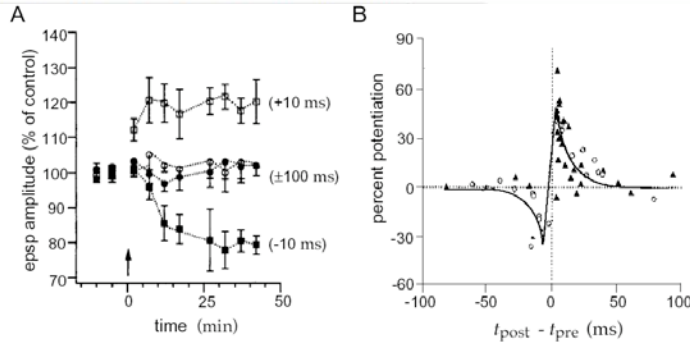
This rule uses only local information (pre- and post-synaptic activities and local synaptic weight) but is not supported by direct evidence. The value of  $|\mathbf{w}|^2$  will relax over time to a value  $1/\alpha$ , preventing the weights from growing without bound and thereby introducing competition between weights.

## 5.5 Spike-time dependence

Biological experiments demonstrate the critical importance of the relative timing of pre- and post-synaptic action potentials in synaptic changes. If pre- and postsynaptic spikes are paired repeatedly, the synaptic strength changes only when both fall in a window of  $\pm 50$  ms. Presynaptic spikes preceding postsynaptic ones produce LTP, the reverse order produces LTD. The maximal change occurs when the time difference is only a few milliseconds.

A) Paired stimulation in cortical slice. Amplitude of postsynaptic EPSP, as percentage of unpaired stimulation control. 50 to 75 pairings produce LTP/LTD of some 20%.

B) Paired stimulation in *Xenopus* tadpole retinotectal synapse. Percent LTP/LTD from multiple pairs of action potentials as a function of relative spike timing.



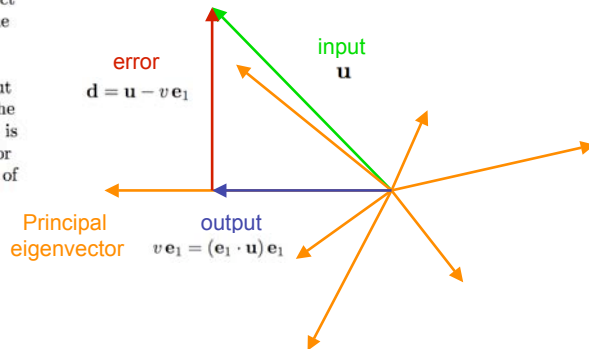
And the solution is:

$$\mathbf{w}(t) = \sum_{\mu=1}^{N_u} (\mathbf{w}(0) \cdot \mathbf{e}_{\mu}) \exp\left(\frac{\lambda_{\mu} t}{\tau_w}\right) \mathbf{e}_{\mu}$$

As all eigenvalues  $\lambda_{\mu}$  are positive, all of the exponential factors grow, albeit not equally rapidly. If the largest eigenvalue  $\mu = 1$  is unique and the initial weights not unfavourable ( $\mathbf{w}(0) \cdot \mathbf{e}_1 \neq 0$ ), the corresponding eigenvector dominates the development so that the response of the trained system to an arbitrary input vector is approximately

$$\mathbf{w} \propto \mathbf{e}_1 \cdot \mathbf{u} \quad \mathbf{w}(\infty) \propto \mathbf{e}_1$$

In short, the weights will reflect the principal eigenvector of the input covariance. Thus, what Hebbian synapses do learn (without supervision) are input correlations. After learning, the output of the trained network is the dot product of input vector and weight vector (projection of input onto eigenvector).



## 6 Linear analysis of single post-synaptic neuron

If we ignore all constraints, we can analyse the basic Hebb rule with standard techniques for solving differential equations. In particular, we can obtain an explicit solution for  $\mathbf{w}(t)$  in terms of the eigenvectors of the covariance matrix  $\mathbf{Q}$  of the input. As covariance matrices are symmetric, all eigenvalues  $\lambda_{\mu}$  are real and non-negative.

$$\tau_w \frac{d\mathbf{w}}{dt} = v \mathbf{u} = \mathbf{Q} \cdot \mathbf{w} \quad \text{where } \mathbf{Q} = \langle \mathbf{u} \mathbf{u} \rangle$$

$$\mathbf{Q} \cdot \mathbf{e}_{\mu} = \lambda_{\mu} \mathbf{e}_{\mu} \quad \mu = 1, 2, \dots, N_u \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{N_u} \geq 0$$

Using the eigenvectors as a basis, we can write the  $N_u$  dimensional vector  $\mathbf{w}(t)$  as

$$\mathbf{w}(t) = \sum_{\mu=1}^{N_u} c_{\mu}(t) \mathbf{e}_{\mu}$$

Substituting this expansion, we can solve the differential equation

$$\tau_w \sum_{\mu=1}^{N_u} \frac{dc_{\mu}(t)}{dt} \mathbf{e}_{\mu} = \sum_{\mu=1}^{N_u} c_{\mu}(t) \mathbf{Q} \cdot \mathbf{e}_{\mu} = \sum_{\mu=1}^{N_u} c_{\mu}(t) \lambda_{\mu} \mathbf{e}_{\mu}$$

$$\Rightarrow \tau_w \frac{dc_{\mu}(t)}{dt} = c_{\mu}(t) \lambda_{\mu} \quad \Rightarrow \quad c_{\mu}(t) = c_{\mu}(0) \exp\left(\frac{\lambda_{\mu} t}{\tau_w}\right)$$

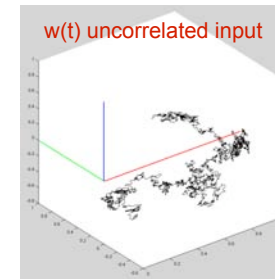
### 6.1 Hebbian rule with weight normalization

What about the exponential growth of weights  $w$ ? We can normalize weights without losing learning abilities, that is, the ability to extract covariances in the input. Consider a Hebbian rule with presynaptic threshold and weight normalization:

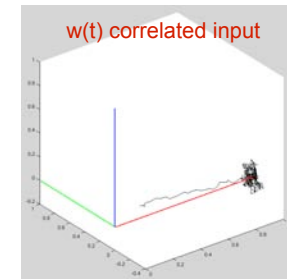
$$v_t = \mathbf{w}_t \cdot \mathbf{u}_t \quad \tau_w \frac{d\mathbf{w}}{dt} = v_t (\mathbf{u}_t - \langle \mathbf{u} \rangle) \quad \langle \mathbf{u} \rangle = \frac{1}{t} \sum_1^t \mathbf{u}_t \quad |\mathbf{w}| = \sqrt{\mathbf{w} \cdot \mathbf{w}} = 1$$

A discrete version of this rule is

$$\mathbf{w}_{t+1} = \frac{\mathbf{x}}{\sqrt{\mathbf{x} \cdot \mathbf{x}}} \quad \mathbf{x} = \mathbf{w}_t + \epsilon v_t (\mathbf{u}_t - \langle \mathbf{u} \rangle) \quad \epsilon = \frac{\Delta t}{\tau_w}$$

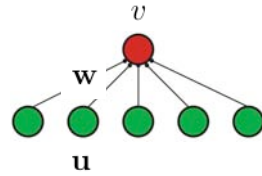


$$\lambda_{1,2,3} = 1.0706 \quad 1.0631 \quad 1.0332$$



$$\lambda_{1,2,3} = 6.7563 \quad 1.0347 \quad 0.7292$$

## 6.2 Principal component projection



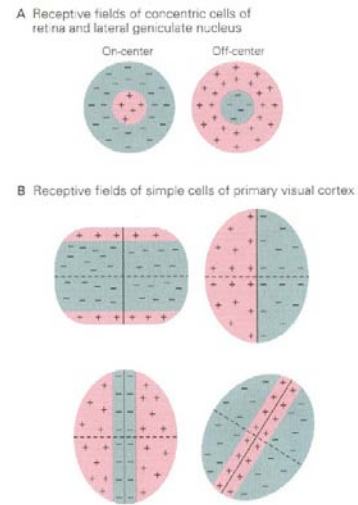
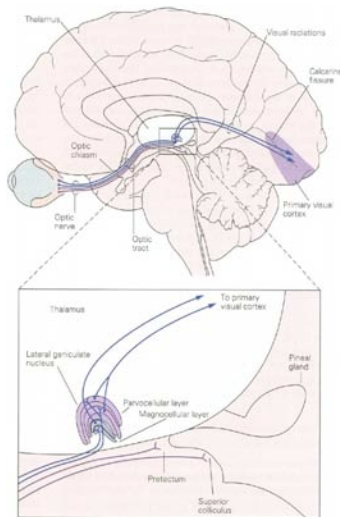
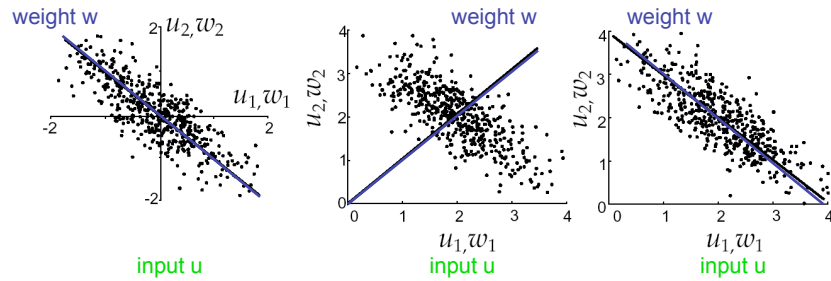
### Correlation rule

$$v = w \cdot u$$

$$\tau_w \frac{dw}{dt} = Q \cdot u, Q = \langle u u \rangle$$

### Covariance rule

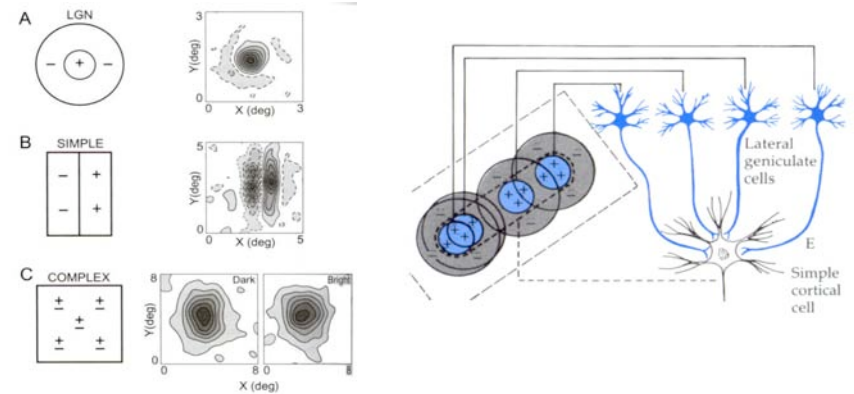
$$\tau_w \frac{dw}{dt} = C \cdot u, C = \langle u u \rangle - \langle u \rangle^2$$



## 7 Biological examples of unsupervised learning

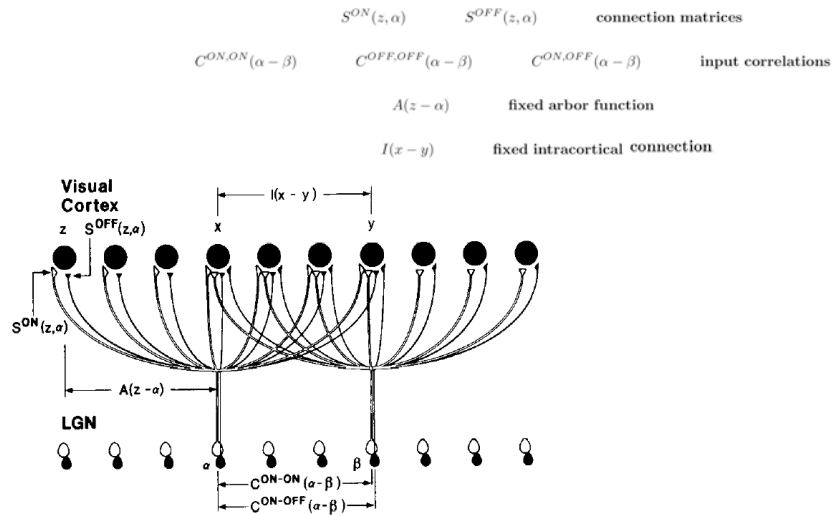
### 7.1 Simple cell receptive fields

Hebbian learning can explain the orientation selectivity of neurons in visual area V1. Orientation selectivity requires a receptive fields with alternating stripes of ON and OFF inputs. Such an arrangement of ON or OFF regions can arise from an initially uniform input distribution by correlation-based plasticity of feed-forward synapses.

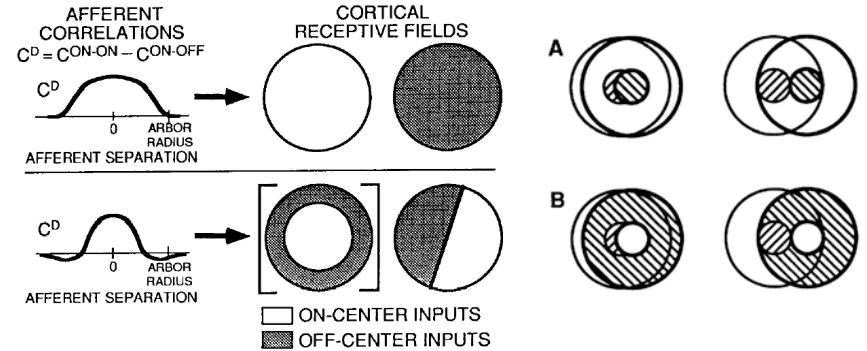


**Miller (1994) Development of simple cell receptive fields**

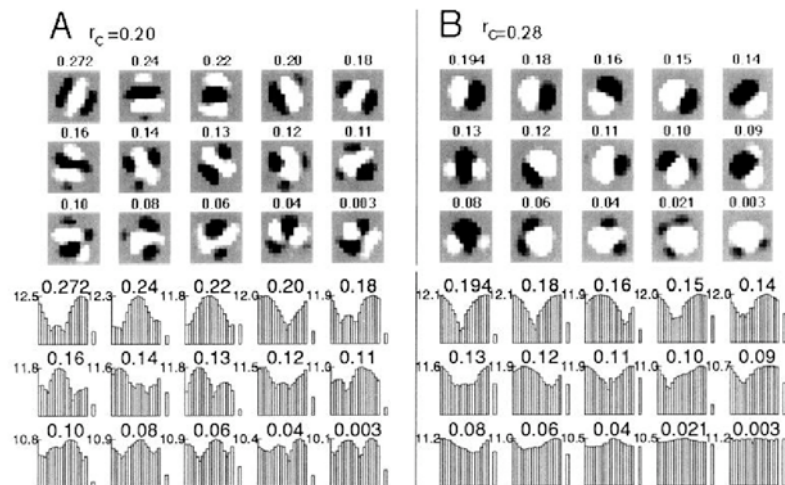
Consider a situation where two initially equivalent input projections (ON- and OFF-center LGN) compete to control a single output layer (V1 simple cells). The early development of a difference between the two projections can be studied by linear models.



**Differential covariance within and between input projections.**



**Receptive fields and orientation tuning.**

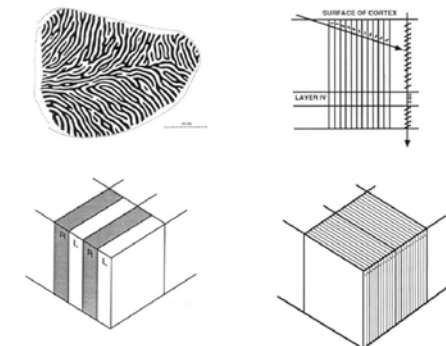


**7.2 Ocular dominance and orientation selectivity**

Unsupervised Hebbian learning helps us understand the development of neuronal selectivity and the formation of cortical maps. In many cases, neural selectivities are arranged across the cortical surface in an orderly fashion known as a "cortical map".

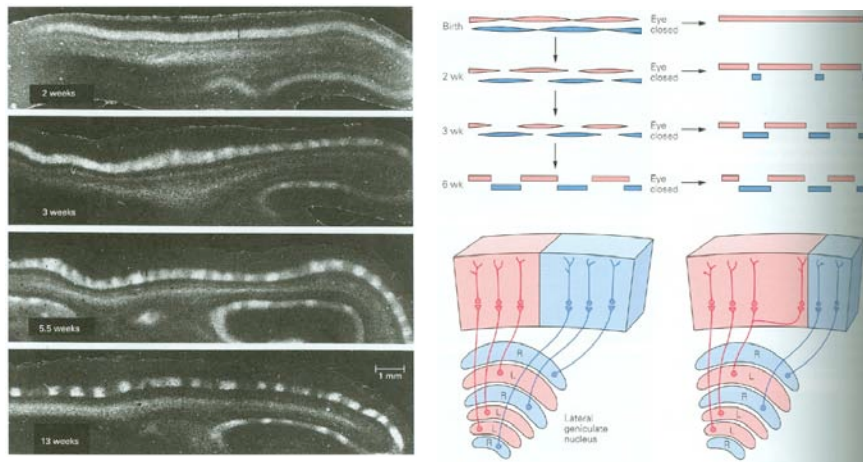
Cortical maps are established during development by both activity-independent and activity-dependent mechanisms. The standard view is that the initial targeting of axons is independent of activity, and simply determines the proper layer to innervate and establishes a coarse order of projections. Activity-dependent mechanisms then refine this initial order and create and preserve the mature selectivities and cortical maps.

The best known examples are ocular dominance stripes and orientation columns in primary visual cortex, but the same principle applies many other sensory cortical areas.

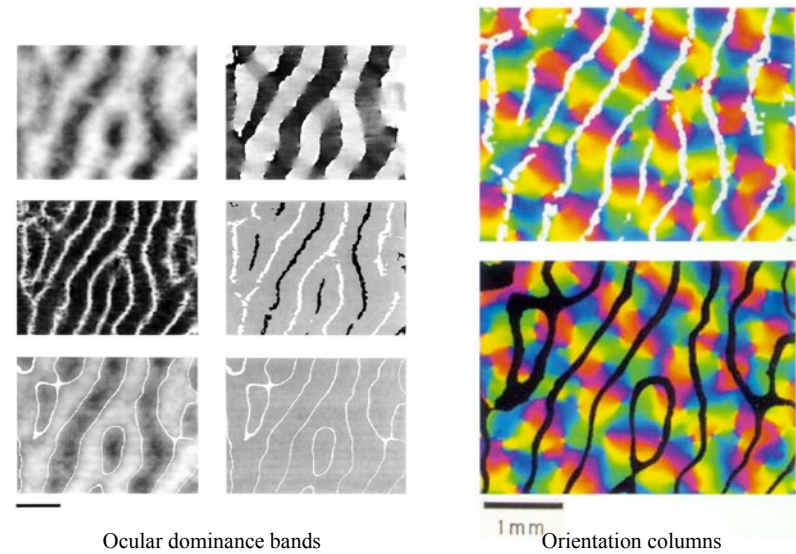




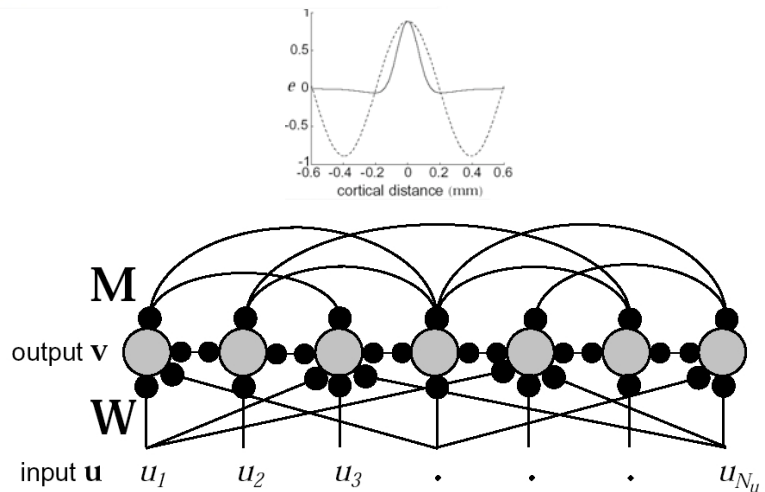
### Development of OD bands in kittens



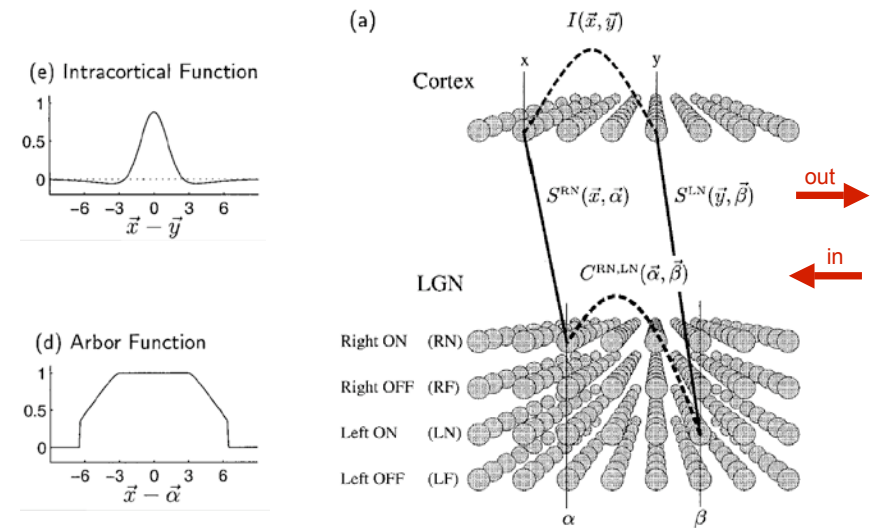
### Segregation patterns revealed by intrinsic signal imaging

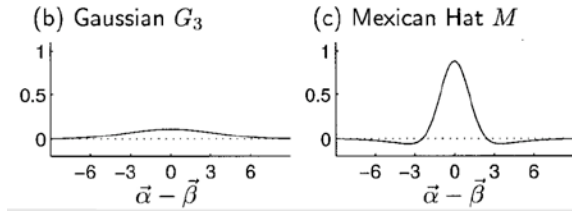


Consider a multi-unit input level mapping onto a multi-output level. Different output neurons must come to represent different aspects of the input. Recurrent inhibition in the output network can ensure that output units develop different selectivities. For example, a "Mexican hat" connectivity ensures (i) that neighboring units have similar selectivities and (ii) that more distant neurons have different selectivities.

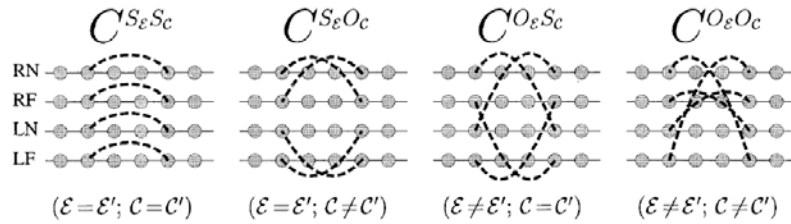


### Erwin & Miller (1998) Orientation and ocular dominance maps





(a) Symmetries reduces the 16 correlation functions to 4 types



To ensure that each output neuron represents one particular input correlation (RN, RF, LN, LF), we impose a *competitive constraint* and require that the sum of all synaptic weights onto each output neuron be constant.

$$\sum_{E, C, \bar{\alpha}} S^{EC}(\bar{x}, \bar{\alpha}, t) = const$$

The results of the Hebbian learning are discussed in terms of *differential weight functions*  $S$  and *differential correlation functions*  $C$ :

$S^{SUM}(\bar{\alpha}, \bar{x})$  and  $C^{SUM}(\bar{\alpha}, \bar{\beta})$ : all cell types

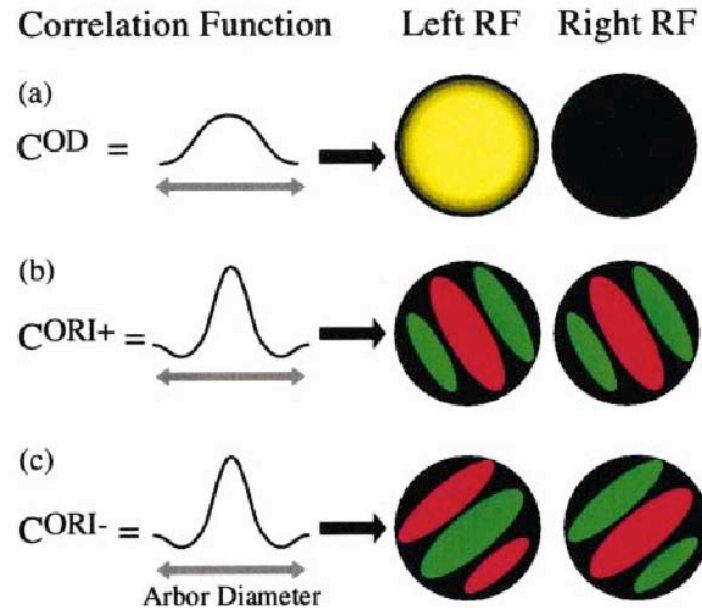
$S^{OD}(\bar{\alpha}, \bar{x})$  and  $C^{OD}(\bar{\alpha}, \bar{\beta})$ : right-eye minus left-eye

$S^{ORI+}(\bar{\alpha}, \bar{x})$  and  $C^{ORI+}(\bar{\alpha}, \bar{\beta})$ : on-center minus off-center

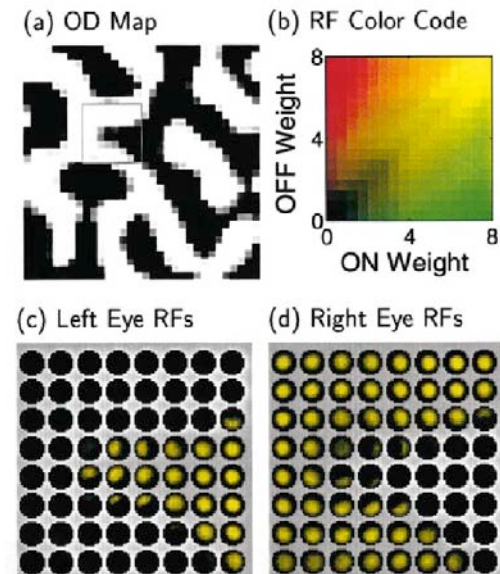
$S^{ORI-}(\bar{\alpha}, \bar{x})$  and  $C^{ORI-}(\bar{\alpha}, \bar{\beta})$ : right-eye on/off difference minus left-eye on-off difference

Table 1. Definitions of the composite variables

Composite weight variables $S^\mu$	Composite correlation functions $C^\mu$
$S^{SUM} = (S^{RN} + S^{RF}) + (S^{LN} + S^{LF})$	$C^{SUM} = (C^{SESc} + C^{SEOc}) + (C^{OESc} + C^{OEOc})$
$S^{OD} = (S^{RN} + S^{LF}) - (S^{LN} + S^{RF})$	$C^{OD} = (C^{SESc} + C^{SEOc}) - (C^{OESc} + C^{OEOc})$
$S^{ORI+} = (S^{RN} - S^{RF}) + (S^{LN} - S^{LF})$	$C^{ORI+} = (C^{SESc} - C^{SEOc}) + (C^{OESc} - C^{OEOc})$
$S^{ORI-} = (S^{RN} - S^{RF}) - (S^{LN} - S^{LF})$	$C^{ORI-} = (C^{SESc} - C^{SEOc}) - (C^{OESc} - C^{OEOc})$

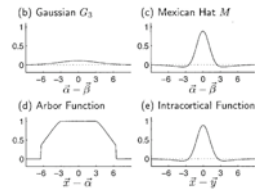


Development Dominated by  $C^{OD}$



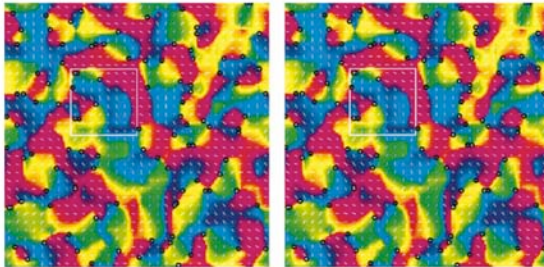
Development dominated by ORI

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

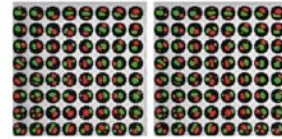


Development of a single ORI mode

(a) Left and right eye ORI maps for  $C^{ORI+}$  dominant



(b) RFs for  $C^{ORI+}$  dominant

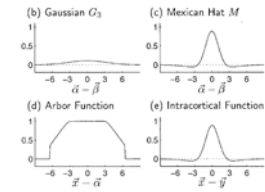


(c) RFs for  $C^{ORI-}$  dominant

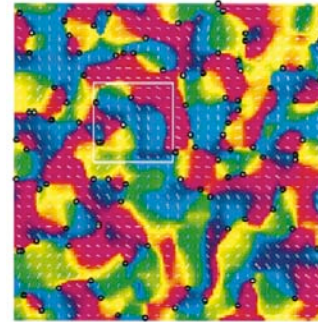


Combined development

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



(a) ORI Preference Map,  $d=1.0$



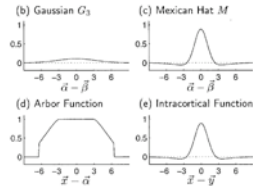
(c)  $d = 0.5$



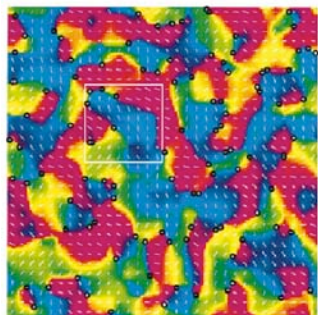
(d)  $d = 1.0$



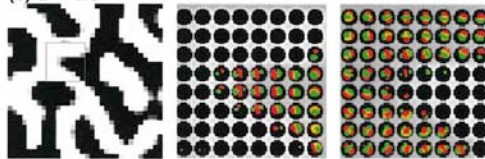
$$C^{OD} = d G_3$$



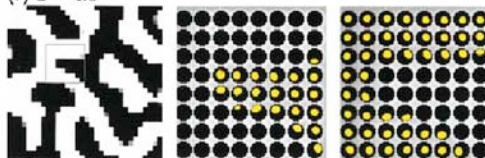
(b) ORI Preference Map,  $d=1.6$



(e)  $d = 1.6$



(f)  $d = 4.0$

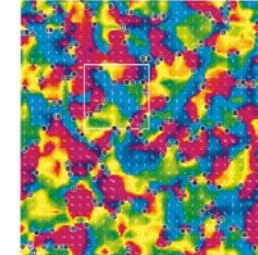


Two stage development

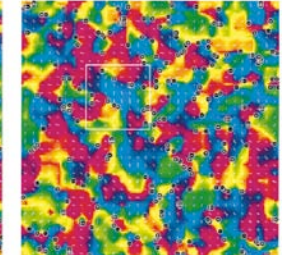
ON vs OFF

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

(a) Stage 1, Left-Eye ORI map ( $t_1 = 26$ )



(b) Stage 1, Right-Eye ORI map ( $t_1 = 26$ )

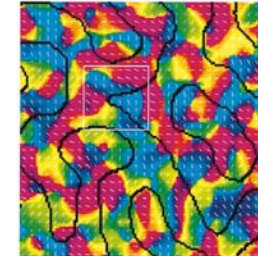


left vs right

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

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

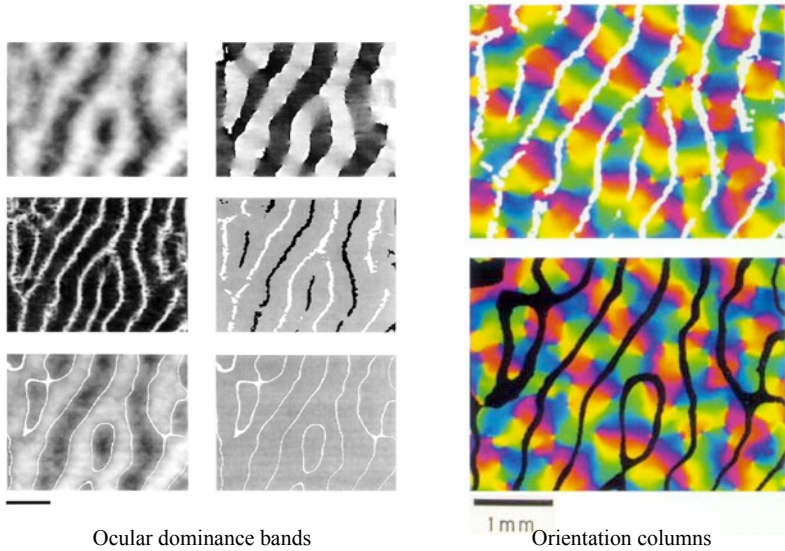
(c) Stage 2, Binocular ORI map



(d) Stage 2, OD map



## Segregation patterns revealed by intrinsic signal imaging



Ocular dominance bands

Orientation columns

## 7.3 Conclusion

Ocular dominance requires:

Same eye inputs more strongly correlated than opposite eye inputs, especially at small separations. Must be true for aggregate input from both center types (ON and OFF).

Orientation-selectivity requires:

Same center input more strongly correlated than opposite center input, at small separations (within arbor radius), and correlated less strongly at large separations. Must be true for aggregate input from both eyes (right and left).

These requirements are met by a wide class of correlation functions. The development of orientation and ocular dominance maps is best achieved sequentially.

The biological development may also be sequential: visual responses in the retina will exhibit ON-OFF correlations even before the two eyes move in a coordinated fashion. LEFT-RIGHT correlations will arise only after the eyes move together.

