

# Emergence of spatiotemporal receptive fields and its application to motion detection

Stefan Wimbauer, Wulfram Gerstner, J. Leo van Hemmen

Physik-Department, Technische Universität München, D-85747 Garching bei München, Germany

Received: 6 April 1994 / Accepted in revised form: 30 June 1994

**Abstract.** A model of motion sensitivity as observed in some cells of area V1 of the visual cortex is proposed. Motion sensitivity is achieved by a combination of different spatiotemporal receptive fields, in particular, spatial and temporal differentiators. The receptive fields emerge if a Hebbian learning rule is applied to the network. Similar to a Linsker model the network has a spatially convergent, linear feedforward structure. Additionally, however, delays omnipresent in the brain are incorporated in the model. The emerging spatiotemporal receptive fields are derived explicitly by extending the approach of MacKay and Miller. The response characteristic of the network is calculated in frequency space and shows that the network can be considered as a space-time filter for motion in one direction. The emergence of *different* types of receptive field requires certain structural constraints regarding the spatial and temporal arborisation. These requirements can be derived from the theoretical analysis and might be compared with neuroanatomical data. In this way an explicit link between structure and function of the network is established.

## 1 Introduction

Motion signals are among the most prominent features that are extracted from the optical input received by the photoreceptors in the retina. The importance of motion information is reflected by the fact that neurons of two areas in the mammalian visual cortex, viz., the middle temporal area (MT or V5) and the medial superior temporal area (MST), seem to be primarily tuned to the detection of motion of whole patterns (Movshon et al. 1985). Neurons that are sensitive to more basic motion signals can be observed earlier in the visual pathway, in particular, in layers 4A and 6 of area V1 (Zeki and Shipp 1988). These neurons respond to motion in *one* direction within a small area of the visual field only.

Numerous models have been developed to understand motion sensitivity. These models can be divided roughly

into four different classes (for a review see Nakayama 1985 and Sereno 1992):

- (i) Models based on a principle first introduced by Hasenstaen and Reichardt (Reichardt 1957; Reichardt and Poggio 1976). In these so-called Reichardt detectors a direction sensitive output is achieved by delayed comparison of two neighbouring receptive fields (van Santen and Sperling 1985).
- (ii) Gradient models where the velocity is calculated by dividing the temporal derivative of the light intensity by the spatial gradient  $v_x = -\frac{dI}{dt} / \frac{dI}{dx}$ . Marr and Ullmann have implemented a special version of a gradient model which is used to compute the direction of motion of an intensity edge (Marr 1982; Marr and Ullmann 1981).
- (iii) Energy models that perform a spatial and a temporal frequency filtering so as to detect velocity (Watson and Ahumada 1985; Adelson and Bergen 1985). These sort of elementary motion detectors are particularly suitable as building blocks for a model of coherent motion (see e.g. Yuille and Grzywacz 1988).
- (iv) Biological models based on identified neural microcircuits that might show a direction dependent response (Maex and Orban 1991, 1992).

The approach presented here contains elements of all four model classes. Our aim is an explanation of some basic principles of motion sensitivity as observed in area V1. We make, however, no attempt to model the underlying neural structure explicitly. Rather, we base our considerations on the concept of *spatiotemporal* receptive fields similar to the approach taken in energy models. It is shown that this type of receptive field emerges as the result of a Hebbian learning process. This way of tackling the problem bears the advantage of being biologically plausible and, at the same time, analytically tractable. Our analysis will proceed in two steps. First, the emergence of spatiotemporal receptive fields is investigated. In the second step we ask the question of how these receptive fields have to be combined to account for motion sensitivity.

The most prominent model to describe the emergence of receptive fields has been developed by Linsker (1986a-c). He introduced a linear feedforward network consisting of

neurons locally connected to neurons of the previous layer. The application of a simple Hebb rule results in the emergence of center-surround and other types of receptive field if the parameters are in an appropriate regime.

Linsker limited his approach to the case of purely spatial receptive fields, a restriction which is also found in most of the experimental work. Since the introduction of the receptive field concept by Hartline (1938) most researchers have investigated the *spatial* structure of receptive fields only. A prominent example in this line of research is the discovery of simple and complex cells in the cat's visual cortex by Hubel and Wiesel (1962).

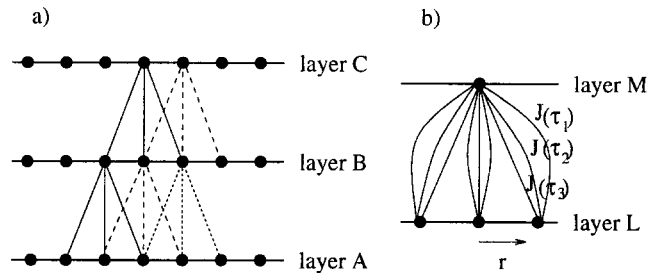
During recent years, however, there has been growing experimental evidence that the temporal characteristic of the receptive field is of great importance for visual processing as well (Dinse et al. 1991; Eckhorn et al. 1993). A spatiotemporal receptive field is characterized by the fact that it changes its spatial response in dependence upon the time that elapses between the arrival of the input to the receptive field and its output. Eckhorn et al. (1993) have developed an elaborate technique, the RF-cinematogram, to study this characteristic explicitly. An important question regarding the temporal aspects of receptive fields refers to the possible role of feedback in the visual cortex (Dinse et al. 1991; Celebrini et al. 1993). While feedback seems to be omnipresent in the cortex, we demonstrate that interesting temporal aspects of the receptive field can also emerge in a purely feedforward structure. Thus, feedback does not have to be incorporated necessarily.

In order to model temporal behaviour we extend the network and learning rule introduced by Linsker to the case of delayed signal transmission. One of the key findings of this article is that a spatial and temporal differentiator are the dominant receptive fields that emerge within the model. A spatial differentiator compares two neighbouring parts of the optical field, whereas a temporal differentiator performs a comparison operation of inputs to the receptive field at subsequent times. The architecture of the network, in particular, the spatial and temporal arborisation, determines whether a spatial or a temporal differentiator is learnt. Thus, we have an intricate relation between structure and function of the network. Motion sensitivity arises by combining the outputs of a spatial and a temporal differentiator. This is done in subsequent layers of the network.

The paper is organized as follows. Section 2 introduces a Hebbian rule that allows learning spatiotemporal correlations. The dominant receptive fields that emerge during learning are calculated. In Sect. 3 possible mechanisms that lead to direction sensitivity are investigated. We perform our analysis both in frequency and in real space. We conclude our considerations with a short discussion on the biological aspects of our model in Sect. 4.

## 2 The emergence of spatiotemporal receptive fields

This section aims at deriving analytically how spatial and temporal differentiators are learnt in a modified Linsker network. One has to keep in mind that these types of receptive field are the essential building blocks for deriving our model of motion sensitivity in Sect. 3.



**Fig. 1.** Structure of an extended Linsker network with axonal delays: **a** Synapses from layer A to B are fixed and constant, whereas synapses from layers B to C are learnt according to a Hebbian rule. **b** Two neurons are not only linked by one synapse as in Linsker's original model but by many synapses with different delays  $\tau$

### 2.1 Network structure and learning rule

Our model is an extended version of the linear feedforward network introduced by Linsker (1986a-c). The essential characteristic of Linsker's network is its spatially convergent structure (see Fig. 1a). That is, a neuron in layer M with  $M=B$  or  $C$  receives inputs from many, say  $n = 10^3$ , neurons of the previous layer L with  $L=A$  or  $B$ , respectively. The neurons of layer L to which a given neuron in layer M is linked are chosen according to a normal distribution in dependence upon the distance  $|r|$  (see Fig. 1b).

In contrast to Linsker's approach two neurons of our model are not linked by a single synapse but by several, say  $l \approx 10 - 20$ , synapses with different delays. For a given pair of neurons  $(i, j)$  with  $i \in M$  and  $j \in L$  the  $l$  delays  $\tau_{ij}$  are sampled from a normal distribution with a mean delay  $\tau_0$ . For the sake of simplicity, the mean and width of the temporal distribution are assumed to be independent of the distance  $|r|$ . Summing up, the structure of the network is completely characterized by three parameters for each layer: the width of the spatial distribution, the mean delay and the width of the temporal distribution of the synapses.

Synapses with a range of different delays have proven to be useful in other model contexts as well, e.g. in an associative network for temporal sequences (Herz et al. 1988, 1989). There are different sources of a delayed response in a neural system. In particular, axonal transmission times, the low pass characteristic of many synapses, and passive transport along the dendrite may altogether account for effective latency times of 10 ms but also up to 100 ms (see e.g. Dinse et al. 1993). In a Linsker network synapses are adjusted according to a Hebbian learning rule. In our network with delays, we have to modify the standard Hebbian rule appropriately in order to store correlations both in space and time.

Let us take a closer look at a single neuron in layer M that receives inputs from neurons in the previous layer L. If we assume a linear transfer function, the activity  $S_{Mi}(t)$  of neuron  $i$  in layer M at time  $t$  is a weighted sum of the activities  $S_{Lj}(t - \tau_{ij})$  in layer L. The sum has to be taken over space, indicated by an index  $j$  and over all possible delays  $\tau_{ij}$

$$S_{M_i}(t) = \sum_j \sum_{\tau_{ij}} J_{ij}(\tau_{ij}) S_{L_j}(t - \tau_{ij}). \quad (1)$$

Since the index  $ij$  of the delays  $\tau_{ij}$  will become irrelevant when adopting a continuum representation in Sect. 2.3 we will drop it in the following.

We have chosen an analogue neuron representation where  $S_{M_i}(t)$  and  $S_{L_i}(t)$  may take arbitrary positive or negative values. Obviously, this is an extremely crude description of a single neuron. The biological quantity, which  $S$  is related to most closely is the postsynaptic membrane potential, where positive and negative values of  $S$  denote potentials above or below the resting potentials. These potentials are summed in the soma and evoke a spike that originates from the soma and will produce another postsynaptic potential at subsequent neurons. By choosing a linear form as in (1), we have neglected both the explicit spike structure of the axonal membrane potential and the non-linear transfer characteristic of a real neuron. Since the synapses  $J_{ij}(\tau)$  depend both on a spatial distance  $ij$  and a temporal distance  $\tau$  they are well suited to store spatial *and* temporal correlations according to a Hebbian rule.

Correlations in a network that describes elements of the visual pathway can arise from two different sources. They may stem either from correlations within the visual input to the network or from fixed synapses that are not modified during learning. In the following we will concentrate on the latter. To be specific, we keep all synapses from layer A to layer B unmodified and constant. The Hebbian rule is applied to synapses from layer B to layer C only. During learning, random and hence uncorrelated patterns are fed into the input layer. The fixed synapses between A and B will introduce Gaussian correlations at layer B. At each learning period that lasts for a time  $\Lambda$ , the synapse  $J_{ij}(\tau)$  is modified according to the Hebbian rule

$$\frac{1}{\Lambda} \Delta J_{ij}(\tau, t) = \epsilon \frac{1}{\Lambda} \int_0^\Lambda ds \left\{ [S_{C_j}(t-s) - c_2] \times [S_{B_i}(t-s-\tau) - c_1] + c_3 \right\}. \quad (2)$$

That is, the presynaptic activity of neuron  $i$  in layer B at time  $t-s-\tau$  is correlated with the postsynaptic activity of neuron  $j$  in layer C at time  $t-s$  with  $s \geq 0$  sampling the past. In this way the signal transmission time  $\tau$  of the presynaptic pulse is taken into account. In particular, the change of the synaptic efficacy is based on the information locally available to the synapse at time  $t$  (Herz et al. 1988, 1989).

Using (1) one obtains

$$\begin{aligned} \frac{1}{\Lambda} \Delta J_{ij}(\tau, t) &= \epsilon \sum_k \sum_{\tau'} J_{jk}(\tau', t) \\ &\left[ \frac{1}{\Lambda} \int_0^\Lambda ds (S_{Bk}(t-s-\tau') - \bar{S}_B) \right. \\ &\left. \times (S_{B_i}(t-s-\tau) - \bar{S}_B) + k_2 \right] + k_1 \\ &= \epsilon \sum_k \sum_{\tau'} J_{jk}(\tau') [Q(i, k; \tau, \tau'; t) + k_2] + k_1 \end{aligned} \quad (3)$$

where

$$\begin{aligned} k_1 &= c_2 c_1 - c_2 \bar{S}_B + c_3 \\ k_2 &= \bar{S}_B^2 - c_1 \bar{S}_B, \end{aligned} \quad (4)$$

$\bar{S}_B$  denotes the time averaged activity that is assumed to be the same for all neurons in layer B, and  $Q$  is the correlation matrix

$$Q(i, k; \tau, \tau'; t) = \frac{1}{\Lambda} \int_0^\Lambda ds (S_{Bk}(t-s-\tau') - \bar{S}_B) \times (S_{B_i}(t-s-\tau) - \bar{S}_B). \quad (5)$$

We assume that the random patterns fluctuate fast as compared with the rate of change of the couplings. Therefore, a self-averaging of the correlation product over  $s$  in (2), (3) and (5) seems to be adequate. Due to the averaging and the random character of the input patterns, the correlation function  $Q$  no longer depends on the time  $t$ ,

$$Q(i, k; \tau, \tau'; t) = Q(i, k; \tau, \tau'). \quad (6)$$

The above separation of time scales allows us to rewrite (3) as a differential equation for the slow change of the couplings that has to be contrasted with the fast fluctuations of the input patterns. Thus, we obtain

$$\frac{dJ_{ij}(\tau, t)}{dt} = \epsilon \sum_k \sum_{\tau'} J_{jk}(\tau', t) [Q(i, k; \tau, \tau') + k_2] + k_1. \quad (7)$$

According to the differential (7) an infinite increase of the couplings might occur. To avoid such biologically implausible behaviour, an upper and a lower bound of the couplings is introduced

$$-J_{max} \leq J_{ij}(\tau) \leq J_{max} \quad \text{for all } i, j, \text{ and } \tau. \quad (8)$$

## 2.2 Outline of the analysis

A detailed analysis of a purely spatial Linsker network has been performed by MacKay and Miller (1990). Many of their arguments can be applied to the spatiotemporal case as well. Following these authors we can solve the differential equation (7) by analysing the fixed points of (7) and calculating the eigenvalues and eigenfunctions of  $(Q(i, k; \tau, \tau') + k_2)$ . To this end we rewrite (7) as a matrix equation

$$\dot{\mathbf{J}} - \epsilon (\hat{Q} + k_2 \hat{I}) \mathbf{J} = k_1 \mathbf{n}, \quad (9)$$

where  $\mathbf{J}$  is a vector,  $\hat{Q}$  a matrix in space and time,

$$I(i, j, \tau, \tau') = 1 \quad \text{for all } i, j, \tau, \tau', \quad (10)$$

and  $\mathbf{n}$  denotes the vector

$$n(i, \tau) = 1 \quad \text{for all } i, \tau. \quad (11)$$

The fixed points of the equation are

$$\begin{aligned} \mathbf{J}^{FP} &= -\frac{k_1}{\epsilon} (\hat{Q} + k_2 \hat{I})^{-1} \mathbf{n} \\ &= -\frac{k_1}{\epsilon} \sum_a \frac{\mathbf{e}_a \cdot \mathbf{n}}{\lambda_a} \mathbf{e}_a = \sum_a J_a^{FP} \mathbf{e}_a. \end{aligned} \quad (12)$$

$\mathbf{e}_a$  and  $\lambda_a$  are the eigenvectors and eigenvalues of  $\hat{Q} + k_2 \hat{I}$ .

As a solution of the differential equation (9) one obtains a linear combination of a special solution plus a solution to the homogeneous equation. A special solution is the fixed point of (9). Thus, we end up with

$$\mathbf{J}(t) = \sum_a [J_a^{FP} + (J_a(0) - J_a^{FP})e^{\lambda_a t}] \mathbf{e}_a, \quad (13)$$

where  $J_a^{FP}$  and  $J_a(0)$  are the components in direction of the eigenvector  $\mathbf{e}_a$  of the fixed point vector  $\mathbf{J}^{FP}$  and the initial vector  $\mathbf{J}(0)$ .

If the couplings are distributed randomly at the beginning, all eigenvectors start with equal amplitude. The components of  $\mathbf{J}$  belonging to the largest eigenvalues, however, will grow fastest and reach the upper and lower bound  $\pm J_{max}$  first. Thus, the largest eigenvalue determines the emerging receptive field.

A different situation may arise if one component  $J_a^{FP}$  of the fixed point vector in the direction of an eigenvector  $\mathbf{e}_a$  exceeds both all other components of the fixed point vector and the components of the initial weight vector  $\mathbf{J}(0)$ ; in particular, the fixed point component belonging to the dominant eigenvalue. Such a component receives a head start and may reach the upper and lower bounds before the component corresponding to the maximal eigenvalue starts dominating the dynamics. As was pointed out by MacKay and Miller, taking a head start into account is necessary to explain the emergence of center-surround receptive fields, as observed by Linsker. In the following, however, we will neglect this mechanism and concentrate on a search for the dominant eigenvalues and the respective spatiotemporal eigenfunctions of  $\hat{Q} + k_2 \hat{I}$ . These eigenfunctions are then identified as the receptive fields of the network.

### 2.3 Calculation of the dominant receptive fields

To calculate the dominant receptive fields, that is, the eigenfunctions of  $\hat{Q} + k_2 \hat{I}$  corresponding to the largest positive eigenvalue, we proceed as follows. First, we adopt a continuum representation to facilitate the analysis. Thereupon, we derive the correlation function  $Q$  for random input patterns and fixed synapses from layer A to B. The eigenfunctions are then calculated in two steps. First, the case  $k_2 = k_1 = 0$  is discussed. After that, the influence of a negative constant  $k_2$  is investigated.

To formulate the problem in a continuum representation, we have to introduce a distribution of the couplings in space and time. An important assumption that is made at this point is that the spatial and the temporal distribution are independent of each other

$$P(\mathbf{r}, \tau) = P(\mathbf{r})P(\tau) \quad (14)$$

We choose

$$P(\tau) = \begin{cases} \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{(\tau - \tau_0)^2}{2T}\right) & \text{for } \tau \geq 0 \\ 0 & \text{for } \tau < 0 \end{cases} \quad (15)$$

For  $\sqrt{T} \ll \tau_0$  one can approximate  $P(\tau)$  by

$$P(\tau) = \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{(\tau - \tau_0)^2}{2T}\right) \quad \text{for all } \tau \quad (16)$$

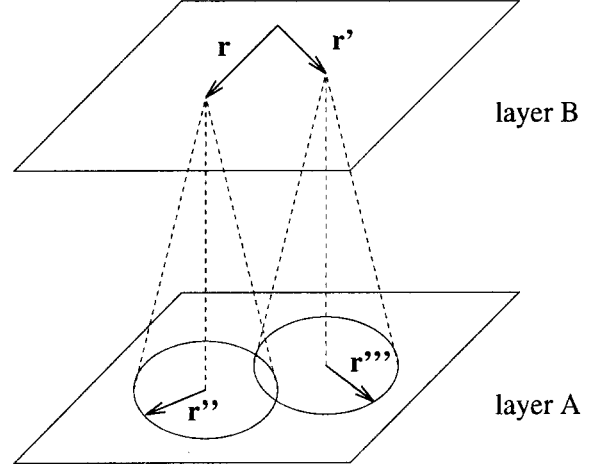


Fig. 2. Correlations are introduced by fixed synapses that link a given neuron in layer B with neurons of the previous layer A that lie in a circular area directly underneath. Gaussian correlations in space and time reflect normal distributions of the couplings with respect to  $|\mathbf{r}|$  and  $\tau$

A Gaussian form is chosen for the spatial distribution as well (cf. Fig. 1b),

$$P(\mathbf{r}) = \frac{1}{2\pi A} \exp\left(-\frac{|\mathbf{r}|^2}{2A}\right) \quad (17)$$

The activity of a neuron can now be calculated by

$$S_M(\mathbf{r}_0, t) = \int d^2\tau \int d\tau P(\mathbf{r} - \mathbf{r}_0)P(\tau) \times J(\mathbf{r} - \mathbf{r}_0, \tau) S_L(\mathbf{r}, t - \tau) \quad (18)$$

and the differential equation that describes the learning process takes the form

$$\frac{dJ(\mathbf{r} - \mathbf{r}_0, \tau, t)}{dt} = \epsilon \left[ \int d^2\tau' \int d\tau' P(\mathbf{r}' - \mathbf{r}_0)P(\tau') \times J(\mathbf{r}' - \mathbf{r}_0, \tau', t) (Q(\mathbf{r}, \mathbf{r}'; \tau, \tau') + k_2) \right] + \epsilon k_1 \quad (19)$$

Let us derive the explicit form of the correlation function  $Q$ . Since the input patterns are uncorrelated in space and time and the averaging time is long on the short time-scale of input pattern fluctuations, the correlation function of the input layer is a simple product of two delta functions,

$$Q_A(\mathbf{r}, \mathbf{r}'; \tau, \tau') = \frac{1}{A} \int_0^A ds (S_A(\mathbf{r}', t - s - \tau') - \bar{S}_A) \times (S_A(\mathbf{r}, t - s - \tau) - \bar{S}_A) = \Delta S_A^2 \delta(\mathbf{r} - \mathbf{r}') \delta(\tau - \tau'), \quad (20)$$

where  $\Delta S_A^2 = \langle (S_A(\mathbf{r}, t) - \bar{S}_A)^2 \rangle$  and where the brackets denote averaging over space and time. For convenience we set  $\Delta S_A^2 = 1$  in the following.

Correlations in layer B are introduced by the set of fixed synapses between layer A and B (see Fig. 2)

$$Q_B(\mathbf{r}, \mathbf{r}'; \tau, \tau') = \int d^2\tau'' \int d^2\tau''' \int d\tau'' \int d\tau''' P(\mathbf{r}'')P(\mathbf{r}''') \times P(\tau'')P(\tau''') J(\mathbf{r}'', \tau'') J(\mathbf{r}''', \tau''')$$

$$\begin{aligned} & \times Q_A(\mathbf{r} + \mathbf{r}'', \mathbf{r}' + \mathbf{r}''', \tau + \tau'', \tau' + \tau''') \\ & = Q_B(\mathbf{r}, \mathbf{r}') Q_B(\tau, \tau'). \end{aligned} \quad (21)$$

The fact that  $Q$  separates into a spatial and a temporal part will facilitate the calculation below.

With  $J(\mathbf{r}, \tau) = 1$  from A to B one obtains

$$\begin{aligned} Q_B(\mathbf{r}, \mathbf{r}') &= \int d^2 r'' P(\mathbf{r}'') P(\mathbf{r} - \mathbf{r}' + \mathbf{r}'') \\ &\sim \exp\left(-\frac{(\mathbf{r} - \mathbf{r}')^2}{2C_B}\right) \end{aligned} \quad (22)$$

$$\begin{aligned} Q_B(\tau, \tau') &= \int d\tau'' P(\tau'') P(\tau - \tau' + \tau'') \\ &\sim \exp\left(-\frac{(\tau - \tau')^2}{2Z_B}\right) \end{aligned} \quad (23)$$

with  $C_B = 2A_A$  and  $Z_B = 2T_A$ .

Now we are ready to solve the eigenvalue problem

$$\begin{aligned} \lambda J(\mathbf{r} - \mathbf{r}_0, \tau) &= \epsilon \left[ \int d^2 r' \int d\tau' P(\mathbf{r}' - \mathbf{r}_0) P(\tau') \right. \\ &\quad \times (Q_B(\mathbf{r}, \mathbf{r}'; \tau, \tau') + k_2) J(\mathbf{r}' - \mathbf{r}_0, \tau') \\ &\quad \left. + \epsilon k_1 \right] \end{aligned} \quad (24)$$

The point is that the constant  $k_1$  appears only in the formula of the fixed point (cf. (12)) and the effects of a large  $\mathbf{J}^{FP}$  are omitted in our considerations, so that we can set  $k_1 = 0$ .

In a first step the eigenfunctions for  $k_2 = 0$  are calculated. In this case the problem simplifies considerably, because  $Q$  is a product of a spatial and a temporal part, and the eigenproblem can be solved for space and time separately. The solution for the couplings can then be formulated as a product of a spatial and a temporal part

$$J(\mathbf{r} - \mathbf{r}_0, \tau) = J(\mathbf{r} - \mathbf{r}_0) J(\tau) \quad (25)$$

The same is true for the eigenvalues

$$\lambda = \lambda_r \lambda_\tau \quad (26)$$

Let us formulate the spatial part of the eigenvalue problem<sup>1</sup>

$$\lambda_r J(\mathbf{r}) = \int d^2 r' Q_B(\mathbf{r}, \mathbf{r}') P(\mathbf{r}') J(\mathbf{r}') \quad (27)$$

with

$$Q_B(\mathbf{r}, \mathbf{r}') = \exp\left(-\frac{(\mathbf{r} - \mathbf{r}')^2}{2C_B}\right) \quad (28)$$

and

$$P(\mathbf{r}') = \frac{1}{2\pi A_B} \exp\left(-\frac{|\mathbf{r}'|^2}{2A_B}\right) \quad (29)$$

The solutions of this equation have been derived by MacKay and Miller (1990), and the three leading eigenfunctions are summarized in Table 1.

Similarly, the temporal eigenvalue problem can be written as

$$\lambda_\tau J(\tau) = \int d\tau' Q_B(\tau, \tau') P(\tau') J(\tau') \quad (30)$$

<sup>1</sup> Since the problem is invariant with respect to translations, we put  $\mathbf{r}_0 = 0$ .

$J_1(\mathbf{r}) = \exp\left(-\frac{r^2}{2R_B}\right)$	$\lambda_{1r} = \frac{C_B}{A_B} \left(\frac{R_B - C_B}{R_B}\right)$
$J_2(\mathbf{r}) = r \cos \Theta \exp\left(-\frac{r^2}{2R_B}\right)$	$\lambda_{2r} = \frac{C_B}{A_B} \left(\frac{R_B - C_B}{R_B}\right)^2$
$J_3(\mathbf{r}) = \left(1 - \frac{r^2}{r_0^2}\right) \exp\left(-\frac{r^2}{2R_B}\right)$	$\lambda_{3r} = \frac{C_B}{A_B} \left(\frac{R_B - C_B}{R_B}\right)^3$
with $R_B = \frac{1}{2} C_B \left(1 + \sqrt{1 + 4\frac{A_B}{C_B}}\right)$	
and $r_0^2 = \frac{A_B R_B}{2A_B + R_B}$	

Table 1. Spatial part of the eigenfunctions

$J_1(\tau) = \exp\left(-\frac{(\tau - \tau_0)^2}{2W_B}\right)$	$\lambda_{1\tau} = \frac{Z_B}{T_B} \left(\frac{W_B - Z_B}{W_B}\right)^{\frac{1}{2}}$
$J_2(\tau) = (\tau - \tau_0) \exp\left(-\frac{(\tau - \tau_0)^2}{2W_B}\right)$	$\lambda_{2\tau} = \frac{Z_B}{T_B} \left(\frac{W_B - Z_B}{W_B}\right)^{\frac{3}{2}}$
$J_3(\tau) = \left(1 - \frac{(\tau - \tau_0)^2}{v_0^2}\right) \exp\left(-\frac{(\tau - \tau_0)^2}{2W_B}\right)$	$\lambda_{3\tau} = \frac{Z_B}{T_B} \left(\frac{W_B - Z_B}{W_B}\right)^{\frac{5}{2}}$
with $W_B = \frac{1}{2} Z_B \left(1 + \sqrt{1 + 4\frac{T_B}{Z_B}}\right)$	
and $v_0^2 = \frac{W_B T_B (\sqrt{T_B + W_B} - \sqrt{T_B})}{(T_B + W_B)^{\frac{3}{2}} - T_B^{\frac{3}{2}}}$	

Table 2. Temporal part of the eigenfunctions

with

$$Q_B(\tau, \tau') = \exp\left(-\frac{(\tau - \tau')^2}{2Z_B}\right) \quad (31)$$

and

$$P(\tau) = \frac{1}{\sqrt{2\pi T_B}} \exp\left(-\frac{(\tau - \tau_0)^2}{2T_B}\right) \quad (32)$$

See Table 2 for the leading eigenfunctions and eigenvalues. An explicit derivation of the temporal part of the eigenfunctions and eigenvalues that proceeds along the same line as the derivation in the spatial case can be found in the Appendix.

Putting things together, we find that the three leading eigenfunctions of (24) for  $k_2 = 0$  are

$$\begin{aligned} J^{ss}(\mathbf{r}, \tau) &= \exp\left(-\frac{r^2}{2R_B}\right) \exp\left(-\frac{(\tau - \tau_0)^2}{2W_B}\right) \\ \lambda^{ss} &= \frac{C_B}{A_B} \left(\frac{R_B - C_B}{R_B}\right) \frac{Z_B}{T_B} \left(\frac{W_B - Z_B}{W_B}\right)^{\frac{1}{2}} \\ J^{os}(\mathbf{r}, \tau) &= r \cos \Theta \exp\left(-\frac{r^2}{2R_B}\right) \exp\left(-\frac{(\tau - \tau_0)^2}{2W_B}\right) \end{aligned} \quad (33)$$

$$\lambda^{os} = \frac{C_B}{A_B} \left( \frac{R_B - C_B}{R_B} \right)^2 \frac{Z_B}{T_B} \left( \frac{W_B - Z_B}{W_B} \right)^{\frac{1}{2}} \quad (34)$$

$$J^{so}(\mathbf{r}, \tau) = \exp\left(-\frac{r^2}{2R_B}\right) (\tau - \tau_0) \exp\left(-\frac{(\tau - \tau_0)^2}{2W_B}\right)$$

$$\lambda^{so} = \frac{C_B}{A_B} \left( \frac{R_B - C_B}{R_B} \right) \frac{Z_B}{T_B} \left( \frac{W_B - Z_B}{W_B} \right)^{\frac{3}{2}} \quad (35)$$

Thus, for  $k_2 = 0$  the eigenfunction  $J^{ss}(\mathbf{r}, \tau)$  emerges as the dominant receptive field during learning. Since  $J^{ss}$  is symmetric both in space and time, it does not show any functional properties besides averaging over space and time.

In the following we will therefore investigate whether for  $k_2 \neq 0$ , other types of receptive fields that show more favorable characteristics are learnt in a Linsker network when a Hebbian rule is applied.

For  $k_2 > 0$  we can use the Perron-Frobenius theorem (Bellman 1970; MacKay and Miller 1990; Seneta 1990), that states that for a matrix whose entries are all non-negative, the components of the principal eigenvector all have the same sign. Therefore, as in the case  $k_2 = 0$ , an eigenfunction of the  $J^{ss}$  type is learnt, and no additional functional properties are gained.

For  $k_2 < 0$ , however, the situation changes dramatically. To clarify how the order of the eigenfunctions is rearranged we go to the limit  $k_2 \ll 0$ , in particular  $|k_2| \gg \lambda^{ss}$ . In this case  $Q$  can be considered as a perturbation of  $k_2$ . The eigenfunctions of

$$\lambda J(\mathbf{r}, \tau) = \int d^2 r' \int d\tau' k_2 P(\mathbf{r}') P(\tau') J(\mathbf{r}', \tau') \quad (36)$$

are  $J_0(\mathbf{r}, \tau) = 1$  corresponding to an eigenvalue  $\lambda_0 = k_2$  and an arbitrary, degenerate set of functions orthogonal to  $J_0(\mathbf{r}, \tau) = 1$  corresponding to an eigenvalue  $\lambda = 0$ . Since  $|\lambda_0| \gg \lambda^{ss}$ , the eigenfunction that is symmetric in both space and time and which one obtains for the perturbed problem is a slight modification of  $J_0(\mathbf{r}, \tau) = 1$ . Hence, it corresponds to a large and negative eigenvalue. Therefore the symmetric receptive field without zeros will be suppressed [see Eq. (13)] and other types of receptive fields can emerge during learning.

In our search for the eigenfunctions corresponding to the leading eigenvalue, we will now examine the class of functions that are either odd in space with respect to the origin or in time with respect to  $\tau_0$  or both. To this end we rewrite the eigenvalue problem (24)

$$\begin{aligned} \lambda J(\mathbf{r}, \tau) = \epsilon \left[ \int d^2 r' \int d\tau' Q(\mathbf{r}, \mathbf{r}') Q(\tau, \tau') \right. \\ \times P(\mathbf{r}') P(\tau') J(\mathbf{r}', \tau') \\ \left. + k_2 \int d^2 r' \int d\tau' P(\mathbf{r}') P(\tau') J(\mathbf{r}', \tau') \right] \quad (37) \end{aligned}$$

Since according to our assumption  $J(\mathbf{r}', \tau')$  is an odd function the second part of the right-hand side in (37) will vanish for this class of functions. Therefore, these types of receptive fields and their eigenvalues are independent of  $k_2$  and can be found by solving the special case  $k_2 = 0$ , as we did above.

In particular,  $J^{os}$  (34) and  $J^{so}$  (35), which are odd in the spatial part with one zero and symmetric in the temporal part and vice versa, correspond to the largest eigenvalues of the subclass of odd functions and are therefore good candidates for the dominant receptive fields as well.

The only question that remains is whether an eigenfunction with an even number of zeros in the spatial or the temporal part corresponds to an eigenvalue larger than the ones of  $J^{os}$  or  $J^{so}$ , what seems to be improbable regarding the ordering of the eigenfunctions in the case  $k_2 = 0$ . We have checked this point numerically, and it turns out that  $J^{os}$  or  $J^{so}$  actually correspond to the largest eigenvalue for  $|k_2| \gg \lambda_0$  and are therefore learnt, if a spatiotemporal Hebbian rule is applied to a Linsker network. An alternative approach would be to perform degenerate perturbation theory similar to the approach taken in MacKay and Miller (1990). To summarize,  $J^{os}$  and  $J^{so}$  are the dominant eigenfunctions of our problem. In the following we will investigate the properties of these functions.

#### 2.4 Properties of the emerging receptive fields

The receptive field  $J^{os}$  changes its sign under reflection along one spatial axis with a positive sign on one side and negative sign on the other. The temporal part of  $J^{os}$  is positive for all delays  $\tau$ . Therefore, it compares two neighbouring regions in space while performing an average over time. Thus,  $J^{os}$  is a spatial differentiator; see Fig. 3a. ,

$J^{so}$ , on the other hand, can be interpreted as a temporal differentiator since this type of receptive field is asymmetric with respect to  $\tau_0$  in the temporal part and does not change sign in the spatial part. Hence,  $J^{so}$  compares the input to the receptive field at subsequent time steps; see Fig. 3b.

The question of whether a spatial or a temporal differentiator is learnt depends on the size of the respective eigenvalues. In particular, a spatial differentiator is learnt, if

$$\lambda^{os} > \lambda^{so} \quad (38)$$

This is equivalent to

$$\frac{A_B}{C_B} > \frac{T_B}{Z_B} \quad (39)$$

$A_B$  and  $T_B$  denote the width of the spatial and the temporal part of the receptive field in layer B. On the other hand, since  $C_B = 2A_A$  [cf. (22)] and  $Z_B = 2T_A$  [cf. (23)],  $C_B$  and  $Z_B$  are functions of the widths of the receptive field in the preceding layer A, and we find that (39) boils down to

$$\frac{A_B}{A_A} > \frac{T_B}{T_A} \Leftrightarrow \frac{A_B}{T_B} > \frac{A_A}{T_A} \quad (40)$$

In conjunction with the mean delay  $\tau_0$ , the widths  $A$  of the spatial distribution and the widths  $T$  of the temporal distribution of the couplings for the different layers provide a complete description of the network. Thus, the size of the eigenvalues  $\lambda^{os}$  and  $\lambda^{so}$  and, hence, the answer to the question of whether a spatial or a temporal differentiator emerges during learning are determined entirely by the architecture of the network, in particular by the way in which the spatial and temporal widths of the receptive field change from

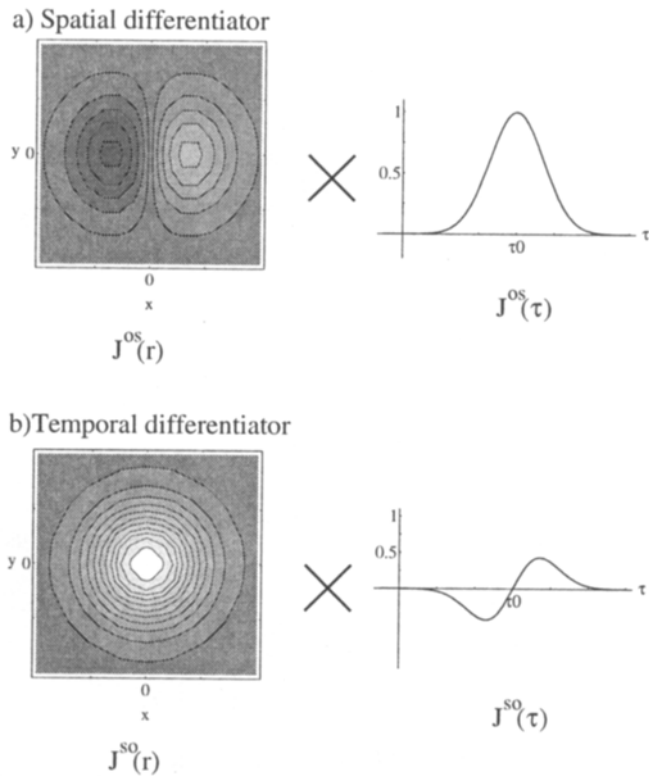


Fig. 3. A spatial and a temporal differentiator emerges as the dominant receptive fields when a Hebbian learning rule is applied to the spatiotemporal Linsker network. The spatial and the temporal part are shown respectively

layer to layer. In this way we have derived an explicit link between the structure and the function of the network.

To summarize the results of this section: from a more abstract point of view, the algorithm performed by the network, that is, spatial or a temporal differentiation, is not determined by the software, i.e. the learning rule, but by the hardware, in particular, the “anatomy” of the network. According to Zeki (1993) this seems to be a general principle governing the organisation of the visual cortex. Different algorithms require a different anatomical machinery and, hence, have to be performed in different areas of the visual cortex. This strong relation of algorithm and network structure is identified by Zeki as the main reason for functional segregation taking place along the visual pathway.

### 3 Modelling motion sensitive cells

In this section we return to the question of how the emergence of motion sensitivity can be explained within the framework of an extended Linsker network. Here, motion sensitivity is understood in the following specific sense: a neuron responds to a pattern, like a moving edge, only if the movement is in one particular direction and remains silent if the pattern moves in the opposite direction.

Furthermore, we have to address the problem of how the network responds to motion patterns of different contrast. In particular, it seems to be a reasonable requirement for a motion detector that its output does not depend on the sign of

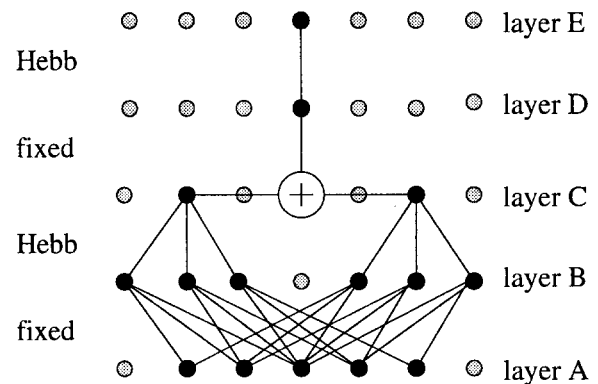


Fig. 4. Structure of the model network: due to variations in the arborisation of the synapses between layers B and C, a spatial differentiator can be learnt on one side and a temporal differentiator on the other. The Gaussian correlations in layer B are introduced by fixed synapses between layers A and B. For the sake of simplicity, temporal but no spatial arborisation is assumed between layer C and D and between layers D and E. Again, fixed synapses between layers C and D cause Gaussian correlations in time in layers D that result in the emergence of a temporal differentiator between layers D and E

the intensity gradient in the direction of motion. Thus, for example, our network should respond to a moving black-white and a white-black edge in the same way.

Similarly to the energy model of Watson and Ahumada (1985), the essential building block of our model is the linear combination of a spatial and a temporal filter. These filters can now be identified with the receptive fields  $J^{os}$  and  $J^{so}$ , the spatial and temporal differentiator, derived in the previous section.

#### 3.1 A model network for the emergence of motion sensitivity

The network we propose in the following has to be considered as a prototype only, i.e. it is one of many possible architectures that may account for motion sensitivity by linking a spatial and a temporal differentiator. Our main intention has been to keep the network as simple as possible in order to make an analytical treatment feasible.

For the sake of simplicity, we make the following assumption concerning our network. We suppose that two layers linked by *fixed* synapses (from layer A to B and from layer C to D in Fig. 4) alternate with layers connected by synapses that show *Hebb* plasticity (from layer B to C and from layer D to E). This supposition can be interpreted as an abstraction of the fact that synapses in different regions of the cortex may learn at different velocities.

We assume that during learning strong random fluctuations are present in every layer of the network. These induce Gaussian correlations via the fixed synapses, which we assume to be stronger than correlations induced by earlier layers. So, technically speaking, the layers of fixed synapses combined with high internal noise help to avoid higher-order correlations that might be introduced by the previously learnt receptive fields themselves. In this way the analytic derivations of Sect. 2 remain applicable to our model network.

During learning a spatial and a temporal differentiator emerge at the synapses from layer B to layer C, where we assume that these two types of receptive fields lie *next to* each other. As described in Sect. 2.4, the character of the receptive field is determined by the size of the eigenvalues  $\lambda^{os}$  and  $\lambda^{so}$ , which in turn depends on the ratio  $A_B/A_A$  and  $T_B/T_A$ , respectively. These ratios reflect the way the spatial and temporal widths of the receptive field change from layer to layer. Therefore, slight variations of the network structure within one layer may cause a situation, where  $\lambda^{os}$  exceeds  $\lambda^{so}$  at one point, while it is smaller in a neighbouring region. This could explain the emergence of a spatial and a temporal differentiator next to each other.

The outputs of the spatial and the temporal differentiator are combined linearly in layer C. Whether the responses are added or subtracted will determine the direction of motion our model network is sensitive to. As will be demonstrated in the following subsection, a network consisting of layers A to C only suffices to explain the phenomena of motion sensitivity. However, the sign of the response in layer C is dependent on the contrast of the moving pattern.

This dependence is compensated by the subsequent layers D and E. For the sake of simplicity, we have assumed that no spatial arborisation takes place between layers C and D and layers D and E. As in previous layers, however, two neurons are linked by several synapses associated with different delay times. Synapses from layer C to layer D are again fixed and constant, whereas synapses from layer D to layer E show Hebbian plasticity. Thus, a temporal differentiator emerges between D and E during learning. The output neuron of layer E then responds to motion in one direction, independent of the contrast of the moving pattern.

We would like to stress once again that the model network of Fig. 4 is one of many possible networks that could explain motion sensitivity. For example, if one assumes additional spatial arborisation between layers C and D and layers D and E, the temporal differentiator between layers D and E might be replaced by a spatial differentiator. Furthermore, the linear combination of the temporal and spatial differentiator taking place in layer C might be performed equally well by spatial synapses linking layer C to an additional layer. These changes would not alter the functional properties of the network.

In the next subsection the response properties of the network are analysed in frequency and real space. It will be demonstrated that the network represents a spatiotemporal filter suitable for the detection of motion signals in one direction.

### 3.2 The response properties of the network in frequency and real space

The use of frequency space through a Fourier transformation will simplify our analysis considerably since the response of one neuron as described in (18) can be regarded as a convolution in space and time.

After factorizing the synaptic kernel  $J(\mathbf{r}-\mathbf{r}_0, \tau) = J(\mathbf{r}-\mathbf{r}_0)J(\tau)$  one obtains

$$S_M(\mathbf{r}_0, t) = \int d^2r \int d\tau P(\mathbf{r}-\mathbf{r}_0)J(\mathbf{r}-\mathbf{r}_0)$$

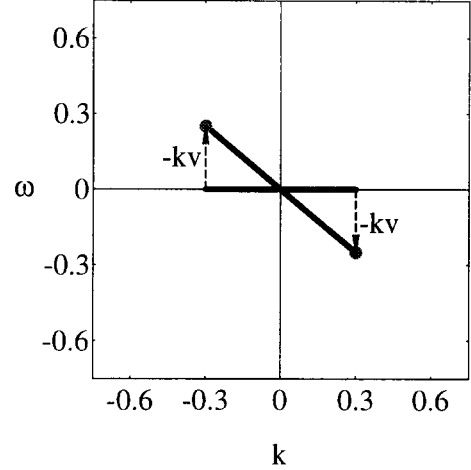


Fig. 5. A stationary pattern has a Fourier transform restricted to the  $\omega = 0$  axis. Movement in *positive*  $x$ -direction shifts the Fourier pattern into the 2nd and 4th quadrants

$$\begin{aligned} & \times P(\tau)J(\tau)S_L(\mathbf{r}, t - \tau) \\ & = \int dk \exp(-i\mathbf{k}\mathbf{r}_0) \int d\omega \exp(-i\omega t) \\ & \times [(2\pi)^3 (PJ)(\mathbf{k})(PJ)(\omega)S_L(\mathbf{k}, \omega)] \end{aligned} \quad (41)$$

or equivalently

$$S_M(\mathbf{k}, \omega) = (PJ)(\mathbf{k})(PJ)(\omega)S_L(\mathbf{k}, \omega), \quad (42)$$

where  $(PJ)(\mathbf{k})$  is the Fourier transform of  $P(-\mathbf{r})J(-\mathbf{r})$  and  $(PJ)(\omega)$ ,  $S_M(\mathbf{k}, \omega)$ ,  $S_L(\mathbf{k}, \omega)$  are the Fourier transforms of  $P(\tau)J(\tau)$ ,  $S_M(\mathbf{r}, t)$  and  $S_L(\mathbf{r}, t)$ , respectively.

Before calculating the explicit response of our network in frequency space, let us investigate how the Fourier transform of a pattern changes if it is moving into a specific direction, say, the direction of the  $x$ -axis.

The Fourier transform of the resting pattern  $S(\mathbf{r}, t)$  is  $S(\mathbf{k}, \omega)$ . Movement of the pattern at a constant velocity  $v_x$  then gives the Fourier transform

$$\begin{aligned} & \int S(x - v_x t, y, t) \exp[-i(k_x x + k_y y)] \\ & \times \exp(-i\omega t) dx dy dt \\ & = \int S(x, y, t) \exp[-i(k_x x + k_y y)] \\ & \times \exp[-it(\omega + k_x v_x)] dx dy dt \\ & = S(\mathbf{k}, \omega + k_x v_x) \end{aligned} \quad (43)$$

That is, motion means shifting the Fourier distribution along the spatial frequency  $k_x$  into the temporal frequency domain  $\omega$  by a value  $-k_x v_x$ ; see Fig. 5. Therefore, a network that is expected to show direction sensitivity in the  $x$ -direction has to respond in two opposite quadrants of the  $\omega k_x$ -plane only, whereas the response in the two remaining quadrants has to be zero.

The output of our model network in frequency space is now derived for layers C and E simply by multiplying and adding the responses of subsequent layers. The relevant Fourier transforms are summarized in Table 3.



Real space		Frequency space	
$P(\mathbf{r})$	$= \frac{1}{2\pi A} \exp\left(-\frac{r^2}{2A}\right)$	$P(\mathbf{k})$	$= \exp\left(-\frac{A}{2} k^2\right)$
$P(\tau)$	$= \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{(\tau-\tau_0)^2}{2T}\right)$	$P(\omega)$	$= \exp\left(-\frac{T}{2}\omega^2\right) \exp(i\omega\tau_0)$
Spatial differentiator			
$P(\mathbf{r})J^{os}(\mathbf{r})$	$= \frac{A+R}{4AR} r \cos \Theta$ $\cdot \exp\left(-\frac{r^2}{2A}\right) \exp\left(-\frac{r^2}{2R}\right)$	$(PJ)^{os}(\mathbf{k})$	$= i \frac{\pi AR}{2(A+R)} k \cos \phi$ $\cdot \exp\left(-\frac{AR}{2(A+R)} k^2\right)$
$P(\tau)J^{os}(\tau)$	$= \sqrt{\frac{T+W}{2\pi WT}} \exp\left(-\frac{(\tau-\tau_0)^2}{2W}\right)$ $\cdot \exp\left(-\frac{(\tau-\tau_0)^2}{2T}\right)$	$(PJ)^{os}(\omega)$	$= \exp(i\omega\tau_0) \exp\left(-\frac{WT}{T+W}\omega^2\right)$
Temporal differentiator			
$P(\mathbf{r})J^{so}(\mathbf{r})$	$= \frac{A+R}{4AR} \exp\left(-\frac{r^2}{2A}\right) \exp\left(-\frac{r^2}{2R}\right)$	$(PJ)^{so}(\mathbf{k})$	$= \exp\left(-\frac{AR}{2(A+R)} k^2\right)$
$P(\tau)J^{so}(\tau)$	$= \frac{T+W}{2WT} (\tau - \tau_0)$ $\cdot \exp\left(-\frac{(\tau-\tau_0)^2}{2W}\right) \exp\left(-\frac{(\tau-\tau_0)^2}{2T}\right)$	$(PJ)^{so}(\omega)$	$= i \left(\frac{WT}{W+T}\right)^{\frac{1}{2}} \exp(i\omega\tau_0)$ $\cdot \omega \exp\left(-\frac{WT}{T+W}\omega^2\right)$

**Table 3.** Fourier transforms relevant to the detection of motion signals

The distribution of the input connections from A to B of the temporal and spatial differentiator is denoted by  $P_A^I$  and  $P_A^{II}$ , respectively. Multiplication with  $(PJ)_B^{so}$  and  $(PJ)_B^{os}$  yields the input to layer C. The output of layer C is the sum of the spatial and temporal differentiator. Thus, we have

$$S_C(\mathbf{k}, \omega) = [P_A^I(\mathbf{k})P_A^I(\omega)(PJ)_B^{so}(\mathbf{k})(PJ)_B^{so}(\omega) + P_A^{II}(\mathbf{k})P_A^{II}(\omega)(PJ)_B^{os}(\mathbf{k})(PJ)_B^{os}(\omega)] \times S_A(\mathbf{k}, \omega) \quad (44)$$

A similar calculation for layer E gives

$$S_E(\mathbf{k}, \omega) = [P_A^I(\mathbf{k})P_A^I(\omega)(PJ)_B^{so}(\mathbf{k})(PJ)_B^{so}(\omega) + P_A^{II}(\mathbf{k})P_A^{II}(\omega)(PJ)_B^{os}(\mathbf{k})(PJ)_B^{os}(\omega)] \times P_C(\omega)(PJ)_D^{so}(\omega)S_A(\mathbf{k}, \omega) \quad (45)$$

The response kernels in the  $\omega k_x$ -plane are plotted for both cases in Fig. 6. We find that neurons in both layer C and layer E respond to signals in the 2nd and 4th quadrants of the  $\omega k_x$ -plane only, whereas signals in the 1st and 3rd quadrants are suppressed. Comparison of the filter properties with the signal characteristics of a moving pattern in frequency space shows that the system is sensitive to motion in the positive  $x$ -direction; cf. Fig. 5.

Note, however, the change of sign between the 2nd and 4th quadrants in the case of a layer C output (Fig. 6). As mentioned in Sect. 3.1, this reflects the fact that the response of layer C depends on the contrast of the moving pattern. We would like to contrast this with the response of a layer E neuron, two layers further up in the pathway, which shows no change of sign in frequency space. In other words, the output is independent of the contrast of the moving pattern.

To clarify this point, let us investigate the response of the network to a moving black-white or white-black edge in Fourier space<sup>2</sup>

$$S_A(x, t) = \pm \text{sgn}(x - v_x t) \text{ with } \text{sgn} = \begin{cases} +1 & \text{for } x \geq 0 \\ -1 & \text{for } x < 0 \end{cases}, \quad (46)$$

where the  $\pm$  sign denotes the two types of edge.

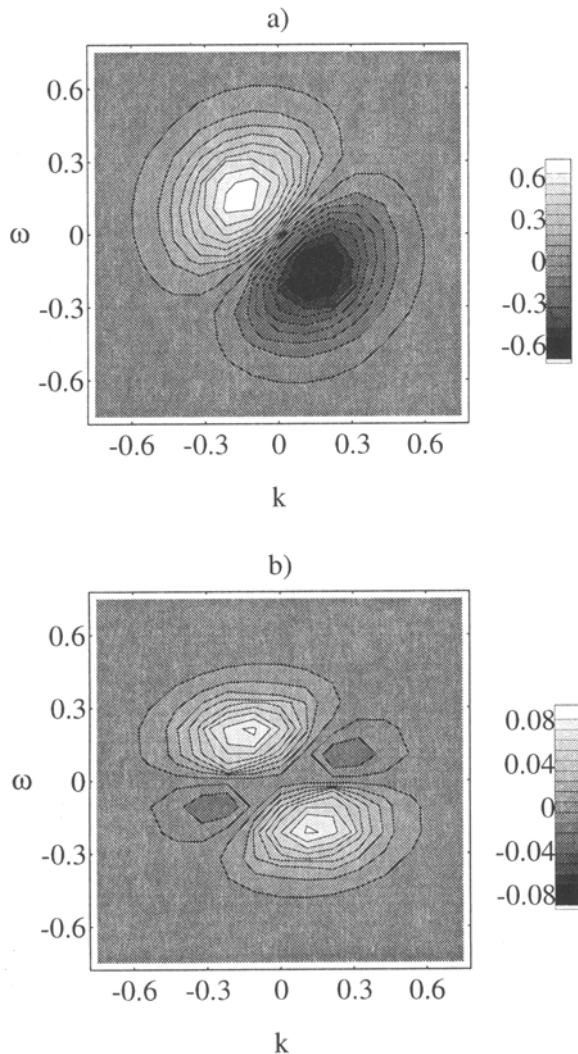
Taking the Fourier transform of (46), one obtains

$$S_A(k_x, \omega) = \mp i \frac{2}{k_x} \delta(\omega + k_x) \quad (47)$$

So the stimulus has different signs in the  $k_x < 0$  and  $k_x > 0$  half-plane, and inserting (47) into (44) for layer C results in a response whose sign depends on the contrast of the pattern. On the other hand, in order to calculate  $S_E(\mathbf{k}, \omega)$  one multiplies a response kernel that is positive in the whole  $\omega k$ -plane (see Fig. 6) with a stimulus that changes sign at  $k_x = 0$ . Therefore,  $S_E(\mathbf{k}, \omega)$  is negative (positive) in the 2nd and positive (negative) in the 4th quadrant for the rightward motion of a white-black (black-white) edge. Thus, the response to both types of edges has a positive component that can be extracted by a threshold element. That is, the sensitivity of the output neuron is independent of the contrast of the moving pattern.

By taking the inverse Fourier transformation of  $S_C(\mathbf{k}, \omega)$  and  $S_E(\mathbf{k}, \omega)$  we could now derive the output of the network as a function of time. Instead, we will verify our analytical results in frequency space by simulating the response of layers C and E to the movement of an edge in a network of discrete neurons. The results of this simulation in real space

<sup>2</sup> Since we consider motion in the  $x$ -direction, we neglect Fourier components in the  $k_y$ -direction.



**Fig. 6.** Response kernel of the model network in frequency space (a) as seen in layer C and (b) in layer E. Comparison with Fig. 5 reveals that both a network up to layer C and a network up to layer E represent a space-time filter sensitive for rightward motion. However, the response kernel of layer C changes sign between the 2nd and the 4th quadrants. Therefore, the response depends on the contrast of the moving pattern. This contrast dependence is compensated by the subsequent layers D and E, as can be recognized from (b)

are summarized in Fig. 7. A black-to-white edge moving to the right causes a positive activity in layer C, and a white-to-black edge moving in the same direction evokes a negative activity in this layer, whereas there is only a weak response of the network to motion in the opposite direction. Thus, our simulation confirms the contrast dependence of the activity in layer C. In layer E, however, the situation is different. Here the response has both a positive and negative contribution for the two contrast types. If we pass this signal through an additional threshold, we obtain a positive component which is independent of the contrast of the moving pattern but highly sensitive to its direction.

It is worth mentioning that neurons corresponding to an output in layer C as well as neurons corresponding to an output in layer E can be observed experimentally. Albus (1980)

reports on experiments in the cat's striate cortex that aim at investigating the effects of contrast reversal on the output of direction sensitive neurons. Whereas one group of neurons responds to motion in one direction independently of the contrast of the moving pattern as in layer E, another group shows a response characteristic similar to the output of layer C. In these cells the preferred direction of motion is reversed if the contrast of the stimulus is reversed. However, the response for one type of contrast is normally much weaker than that for the other type. Passing the output of layer C through a threshold element reproduces exactly such behaviour.

The examples which we have considered so far dealt with motion in the  $x$ -direction exclusively. Since we have chosen the spatial differentiator in such a way that it changes its sign along the  $x$ -axis, our model network is most sensitive to motion in this direction.

Finally, let us suppose our network is stimulated by a contrast edge moving in an *arbitrary* direction  $\mathbf{v} = (v \cos \Theta, v \sin \Theta)$ . A glance at Eq. (34) shows that the response of the spatial differentiator from layer B to C is diminished by a factor  $\cos \Theta$ . In particular, it goes to zero for motion in the  $y$ -direction. On the other hand, the response of the temporal differentiator is independent of the direction of motion. Thus, the response in C and E changes gradually from a motion detector to a purely temporal differentiator as  $\Theta$  changes from 0 to  $\pi/2$ .

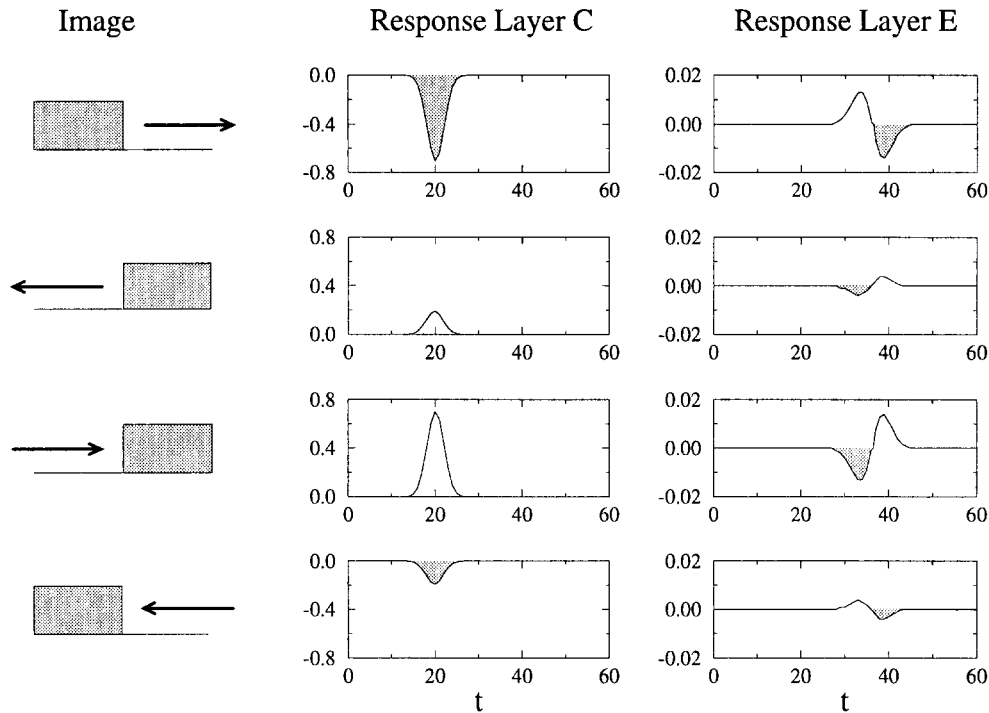
#### 4 Conclusions and biological discussion

We have developed a simple linear feedforward network that generalizes an architecture introduced by Linsker to the case with delays. Within this framework the emergence of direction-sensitive cells can be understood as the result of a combination of different types of time-dependent receptive fields, in particular, a spatial and a temporal differentiator. These receptive fields are learnt in our network, if a Hebbian rule that stores correlations in space *and* time is applied.

Analytically, the receptive fields can be interpreted as the eigenfunctions of the spatiotemporal correlation function  $Q$  corresponding to the leading positive eigenvalue. The question of whether a spatial or a temporal differentiator emerges during learning is determined entirely by the widths of the spatial arborisation and the distribution of the delays, that is, by the architecture of the network.

Obviously, our model has shortcomings with respect to biological details. In particular, we restrict ourselves to a purely linear feedforward network, neglecting feedback connections and non-linearities caused by the synaptic transfer function and postsynaptic integration. Furthermore, the complicated spike response of single neurons is replaced by a simple analogue neuron representation. We do not distinguish between excitatory and inhibitory synapses but use abstract synapses and neurons that may have both signs. Finally, preprocessing in the retina, in particular, the transient response characteristic of M ganglion cells, is neglected completely.

The aim of this study, however, was not towards a detailed neurobiological model of the visual pathway, but



**Fig. 7.** Simulation results for a moving black-white and a white-black edge: the network is sensitive to rightward motion in both layers C and E. Whereas the sign of the response depends on the contrast of the edge for layer C, the output neuron in layer E responds both positively and negatively for a black-white as well as a white-black edge. Passing the output through an additional threshold element yields a positive response for both cases

rather towards a principle issue. We wanted to clarify the relation between structural properties and the functional characteristics of the network. Indeed, one important result of our investigations is that an explanation of motion sensitivity can be based on the *linear* concept of spatiotemporal receptive fields. These linear components have to be completed, at the very end, by a simple threshold element. While non-linear response characteristics or feedback loops are known to be present all over the cortex, they are not central for our model of motion sensitivity of cells in V1<sup>3</sup>. Only recently, Jagadeesh et al. (1993) performed experiments supporting the idea that linear mechanisms play an important role in direction sensitivity. They have demonstrated by intracellular recordings that the changes in the membrane potential evoked by a moving stimulus are predicted accurately by the linear summation of responses to stationary stimuli.

A central issue of our derivation is the combination of a spatial and a temporal differentiator situated next to each other. Is there experimental evidence that nature actually computes motion signals in this manner? Physiological studies of Baker and Cynader (1988) and other authors revealed that motion sensitive neurons in V1 are independently tuned to spatial and temporal frequency - a result that seems to support our assumption. Furthermore, our analysis shows that local variations in the spatial and temporal arborisation of the synapses are necessary prerequisites for the emergence of a structure with a temporal and a spatial differentiator next to each other. These variations may be reflected neuroanatom-

ically by the patches of motion-sensitive cells in layer 4A of V1 projecting to V5 (Shipp and Zeki 1989) that can be revealed by the injection of the retrograde tracer horseradish peroxidase (HRP) to V5.

A major assumption of our discussion concerns the importance of Hebbian learning during development. It is, however, still an open problem, to what extent the structure of the visual cortex has to be explained primarily by learning during early development (Rakic 1977) or by genetic determination. To answer this question, more experimental and theoretical work seems to be necessary.

*Acknowledgement.* S.W. gratefully acknowledges financial support from the Hanns Seidel Stiftung and the Studienstiftung des deutschen Volkes. W.G. has been supported by the Deutsche Forschungsgemeinschaft (grant numbers He 1729/2-1 & 2-2).

## Appendix

Following the arguments for the derivation of the spatial eigenfunctions by MacKay and Miller (1990), we solve the temporal eigenproblem

$$\lambda_{\tau} J(\tau) = \int d\tau' Q(\tau, \tau') P(\tau') J(\tau') \quad (48)$$

with

$$Q_B(\tau, \tau') = \exp\left(-\frac{(\tau - \tau')^2}{2Z_B}\right) \quad (49)$$

and

<sup>3</sup> A different point of view regarding this problem has been advocated by Dinse et al. (1991).

$$P(\tau) = \frac{1}{\sqrt{2\pi T_B}} \exp\left(-\frac{(\tau - \tau_0)^2}{2T_B}\right) \quad (50)$$

The integral in (48) can be considered as a convolution of  $Q$  with  $(PJ)$ . So by going to a Fourier representation, the right-hand side of (48) becomes a product

$$\lambda_\tau J(\omega) = Q(\omega)(PJ)(\omega) \quad (51)$$

As an ansatz we take a Gaussian multiplied by a polynomial

$$J(\tau) = \exp\left(-\frac{(\tau - \tau_0)^2}{2W}\right) L(\tau - \tau_0) \quad (52)$$

with  $L$  being of the form  $L(\tau) = 1$ ,  $L(\tau) = \tau$ , or  $L(\tau) = 1 + a\tau^2$ . The expressions for  $W$  and  $a$  are then obtained by comparing coefficients in Fourier space.

## References

- Adelson EH and Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2:284-299
- Albus K (1980) The detection of movement direction and effects of contrast reversal in the cat's striate cortex. *Vision Res* 20:289-293
- Baker CL and Cynader MS (1988) Space-time separability of direction selectivity in cat striate cortex neurons. *Vision Res* 28: 803-812
- Bellman R (1970) Introduction to matrix analysis, 2nd Edition (McGraw-Hill, New York) p. 208
- Celebrini S, Thorpe S, Trotter Y, Imbert M (1993) Dynamics of orientation coding in area V1 of the awake primate. *Vis Neurosci* 10:811-825
- Dinse HR, Krüger K, Mallot HA, Best J (1991) Temporal structure of cortical information processing: Cortical architecture, oscillations, and non-separability of spatio-temporal receptive field organization. In: *Neuronal Cooperativity*, edited by J. Krüger (Springer-Verlag, Berlin Heidelberg) pp 68-104
- Dinse HR, Spengler F, Godde B, Hartfiel B (1993) Dynamic aspects of cortical function: processing and plasticity in different sensory modalities. In: *Brain Theory*, edited by A. Aertsen
- Eckhorn R, Krause F, Nelson JI (1993) The RF-cinematogram: A cross-correlation technique for mapping several visual fields at once. *Biol Cybern* 69:37-55
- Hartline HK (1938) The response of single optic nerve fibres of the vertebrate eye to illumination of the retina. *Am J Physiol* 121:400-415
- Herz AVM, Sulzer B, Kühn R, van Hemmen JL (1988) The Hebb rule: storing static patterns and dynamic objects in an associative neural network. *Europhys Lett* 7:663-669
- Herz AVM, Sulzer B, Kühn R, van Hemmen JL (1989) Hebbian learning reconsidered: Representation of static and dynamic objects in associative neural nets. *Biol Cybern* 60:457-467
- Hubel DH and Wiesel TN (1962) Receptive fields, binocular integration and functional architecture in the cat's visual cortex. *J Physiol Lond* 160:106-154
- Jagadeesh B, Wheat HS, Ferster D (1993) Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex. *Science* 262:1901-1904
- Linsker R (1986a) From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proc Natl Acad Sci USA* 83:7508-7512
- Linsker R (1986b) From basic network principles to neural architecture: Emergence of orientation-selective cells. *Proc Natl Acad Sci USA* 83:9390-8394
- Linsker R (1986c) From basic network principles to neural architecture: Emergence of orientation columns. *Proc Natl Acad Sci USA* 83:8779-8783
- MacKay DJC, Miller KD (1990) Analysis of Linsker's application of Hebbian rules to linear networks. *Network* 1:257-297
- Maex R and Orban G (1991) Subtraction inhibition combined with a spiking threshold accounts for cortical direction selectivity. *Proc Natl Acad Sci USA* 88:3549-3553
- Maex R and Orban G (1992) A model circuit for cortical temporal low-pass filtering. *Neural Comp* 4: 932-945
- Marr D (1982) *Vision*. Freeman, San Francisco
- Marr D and Ullmann S (1981) Directional selectivity and its use in early visual processing. *Proc R Soc London Ser B* 211:151-180
- Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1985) The analysis of moving visual patterns. In: *Pattern recognition mechanisms*, edited by Chagas C, Gattas R and Gross CG (Vatican Press, Rome) pp 117-151
- Nakayama K (1985) Biological image motion processing: a review. *Vision Res* 25:625-660
- Rakic P (1977) Prenatal development of the visual system in the rhesus monkey. *Philos Trans R Soc Lond Ser B* 278:245-260
- Reichardt W (1957) Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems. *Z Naturforsch Teil B* 12:447-457
- Reichardt W and Poggio T (1976) Visual control of orientation behaviour in the fly, Part I and II. *Quart. Rev Biophysics* 9, 311-438
- van Santen JPH and Sperling G (1985) Elaborated Reichardt detectors. *J Opt Soc Am A* 2:300-321
- Seneta E (1990) *Nonnegative matrices* (Wiley, New York)
- Sereno ME (1992) *Neural computation of pattern motion* (MIT Press, Cambridge)
- Shipp S and Zeki S (1989) The organization of connections between areas V5 and V1 in macaque monkey visual cortex. *Eur J Neurosci* 1:309-332
- Watson AB and Ahumada AJ (1985) Model of human visual motion sensing. *J Opt Soc Am A* 2:322-342
- Yuille AL and Grzywacz NM (1988) A computational theory for the perception of coherent visual motion. *Nature* 333: 71-74
- Zeki S and Shipp S (1988) The functional logic of cortical connections. *Nature* 335: 311-317
- Zeki S (1993) *A vision of the brain* (Blackwell, Oxford)