Kybernetik 14, 85—100 (1973)

Self-Organization of Orientation Sensitive Cells in the Striate Cortex

Chr. von der Malsburg
Max-Planck-Institut für Biophysikalische Chemie, Göttingen, FRG

Received: June 10, 1973

Abstract

A nerve net model for the visual cortex of higher vertebrates is presented. A simple learning procedure is shown to be sufficient for the organization of some essential functional properties of single units. The rather special assumptions usually made in the literature regarding preorganization of the visual cortex are thereby avoided. The model consists of 338 neurones forming a sheet analogous to the cortex. The neurones are connected randomly to a "retina" of 19 cells. Nine different stimuli in the form of light bars were applied. The afferent connections were modified according to a mechanism of synaptic training. After twenty presentations of all the stimuli individual cortical neurones became sensitive to only one orientation. Neurones with the same or similar orientation sensitivity tended to appear in clusters, which are analogous to cortical columns. The system was shown to be insensitive to a background of disturbing input excitations during learning. After learning it was able to repair small defects introduced into the wiring and was relatively insensitive to stimuli not used during training.

I. Introduction

The task of the cortex for the processing of visual information is different from that of the peripheral optical system. Whereas eye, retina and lateral geniculate body (LGB) transform the images in a "photographic" way, i.e. preserving essentially the spatial arrangement of the retinal image, the cortex transforms this geometry into a space of concepts.

Within the last decade electrophysiology took the first steps into discovering the way in which the visual cortex performs this transformation. This paper is mainly concerned with the following features, which have been found in the primary visual cortex (area 17) of cat and monkey (Hubel and Wiesel, 1962, 1963, 1968).

1. There are neurones, which are selectively sensitive to the presentation of light bars and edges of a certain orientation (Hubel and Wiesel, 1962).

2. The neurones seem to be organized in "functional columns", i.e. the neurones lying within one cylinder vertical to the cortical surface are sensitive to the same orientation (Hubel and Wiesel, 1963).

3. Neighbouring columns tend to respond to stimuli of similar orientation (Hubel and Wiesel, 1963, 1968).

Although these findings are interesting in themselves, they will yield their full potential profit only if two questions are answered:

I. For what reason and to what end is area 17 organized in this way?

II. By which mechanisms are these neuronal properties determined?

Ad I: The fibres of the optic radiation are most sensitive to phasic changes of light intensity. Such intensity changes are brought about by eye movements, which scan the receptive fields of individual retinal and geniculate neurones over the light and dark contours of an image. Moving light edges and bars are therefore the most important stimuli which lead to geniculate output, i.e. cortical input. This may be considered one of the conditions for the existence of edge and bar detectors at the first cortical levels.

Ad II: The only mechanism proposed so far in the literature is genetical predetermination of the required circuitry (Hubel and Wiesel, 1963). This view has several disadvantages.

First, it would cost the system an immense volume of genetic information to tell all the terminal branches of the afferent axons with which cortical neurone they have to make contact.

Second, a rigid, genetically determined circuit would not have a high degree of plasticity. Such plasticity was demonstrated by experiments, in which the trigger features of visual cortical cells of young kitten could be determined in various ways by visual experience (Hirsch and Spinelli, 1970; Blakemore and Cooper, 1970; Blakemore and Mitchell, 1973).

Finally, plasticity, i.e. a process of self-organization should be possible in later stages of information processing by the brain, when it has to deal with situations not foreseen by nature.

The aim of this paper is to propose a mechanism of self-organization of the visual cortex which is able to explain in a simple way the facts 1 to 3 above and which also reduces the genetical problem to a reasonable level.

II. The Model

We describe now a model structure for the visual cortex and its specific afferents, which is in principle in accord with the known anatomical data and which has the minimum degree of complication required for the purpose of this paper.

a) The Elements

The model consists of a network of cells. The information transmitting signal is thought to be the discharge rate of the cells, averaged over a small interval of time. Thus the signal we employ here is a smooth function of time. This avoids the problem of artificial pulse — synchronization caused by the quantization of time, which is necessary in computer simulations. The cells make contact via connections (synapses). The connections can differ in weight, and are characterized by a number called strength of connection. The strength of connection from a cell A to a cell B will be denoted by p_{AB}. This term would include excitatory effects such as the sum of the postsynaptic potentials caused in a cell B by all the synapses of cell A on the dendrites and cell body of B. No assumptions are being made about the morphological variables which might determine the strength of connection. It may be different effectiveness or position of single synapses or merely a variable number of synapses between cells A and B (probably both). There are excitatory (E) and inhibitory (I) cells which have positive and negative strengths of connection respectively.

It is a simplification to characterize the connection between two cells by one number (or actually two numbers, as there are two directions, $A \to B$ and $B \to A$). In reality there is at least the complication of a variable synaptic delay and of different time-courses of excitation in cell B caused by one pulse of cell A. But the simplification of one number is sufficient for our purpose. The excitation per second caused by A in B is equal to the output of A (i.e. its firing frequency) times the strength of connection $A \to B$. This may, of course, be negative or positive.

It is furthermore assumed that the excitation and inhibition of one cell caused by all its presynaptic elements are summed linearly. The resulting quantity constitutes the input to the cell. The internal state of the cell is described by a quantity $H(t)$, which is called "excitatory state" (ES). As output signal of the cell we take that part of its ES, which exceeds a threshold. It shall be denoted by $H^*(t)$ (see Table 1). The cells influence each other via their output signals. The total input into a cell k is $\sum_l p_{lk} H_l^*(t)$, the p_{lk} being the strength of connection from cell l to cell k.

In the absence of input to a cell its ES decays exponentially with time. This can be described by the differential equation $d/dt\, H(t) = -\alpha H(t)$. The decay constant α is introduced to represent two phenomena, the decay of the postsynaptic potentials and the post-excitatory polarization following each action potential. A single decay constant is a crude but adequate approximation to these two precesses. The cells used here are very simple models of real neurones. The most conspicuous simplification is the linear dependency of the output signal on the input, as long as the threshold is exceeded. This implies in particular, that there is no intrinsic upper limitation to the output signal, as is imposed in reality by an absolute refractive period. Such an unlimited output can lead to instability in a network, which contains circular excitatory pathways. Therefore, in this model, instabilities can only be avoided by a properly devised inhibitory system within the network. In real systems there seems also to exist a limiting mechanism apart from the absolute refractive period, as cells rarely fire with maximum frequency for an extended period of time (above 100 msec).

b) The Wiring of the Model

The cells of the model form a two dimensional arrangement, the "cortical plane". Their distribution is uniform, E- and I-cells have equal density, although the relative proportions do not matter. The strength of connection between any two cells is a function $f(x)$ of their distance x. This function should be monotonically decreasing, e.g. bell-shaped. It is characterized by a range R and an amplitude A. There are functions $f(x)$ with different range R and amplitude A according to the different types of pairs of pre- and postsynaptic elements: $E \to E$, $E \to I$, $I \to E$ and $I \to I$, (see Fig. 1):

$f_{EE}(x)$ with amplitude $A_{EE} > 0$ and range R_{EE},

$f_{EI}(x)$ with amplitude $A_{EI} > 0$ and range $R_{EI} = R_{EE}$,

$f_{IE}(x)$ with amplitude $A_{IE} < 0$ and range $R_{IE} > R_{EE}$.

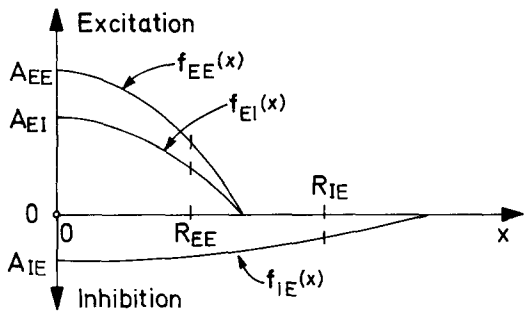

Fig. 1. Schematic representation of the dependency of the intracortical connection strengths on cell distance x. Explanation of symbols in the text

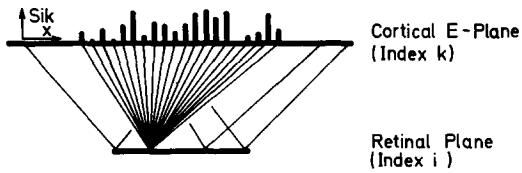

Fig. 2. Schematic drawing to show the organization of the afferents. The lower horizontal line represents the retinal plane or a cross section through the bundle of afferent fibres. The upper horizontal line represents a section of the cortical plane. The vertical bars symolize the different strengths of connection from one fibre i to many of the cortical E-cells k. The distribution of the heights of the bars is random. The connection of the other afferent fibres to one cortical cell are not shown

No connection is assumed here between the I-cells ($A_{II}=0$). This assumption and the restriction to $R_{EI}=R_{EE}$ do not lead to a loss of features essential for this paper. The intracortical connections described will serve to organize columns. For this it is essential that R_{IE} be larger than R_{EE}, as will become evident later on.

How does this model cortex compare with the cortex found in higher vertebrates? Firstly, it has only two kinds of cells, E- and I-cells. We do not try here to identify them with two of the many classes of neurones described in actual cortex. This identification may, in fact, be impossible, as perhaps no single cell of the cortex has the field of innervation we look for. The model could in this case be saved by the existence of multicellular units: clusters of cells integrated by strong mutual excitation. The individual arborizations of these cells could then add up to give the postulated field of innervation. It was Colonnier, who introduced the concept of local fields of innervation to explain columnar organization. He also discussed histological evidence for this scheme (Colonnier, 1966).

It should be emphasized, that the model described here certainly corresponds to only a part of the real cortex; for example long-range excitatory connections within the cortex are not represented. Also there may be several systems like the one described here, which occupy the same space and which are weakly linked to each other.

c) *The Afferent Organization*

There is a set of afferent fibres which provide the input to the model cortex. The fibres have circular receptive fields within a small area of the retina. Where the retina is hit by the light of a stimulus, it switches its optical fibres to an active state, i.e. a state of constant firing. All the other fibres are silent. The model retina is thus again a simplification as only one type of cells (on-cells) without a center-surround organization and no off-cells are assumed. This simplification is possible, as only static light bars will be used as stimuli. Each afferent fibre projects to an area of the cortical plane and connects to all the E-cells within that area. A possible connection to I-cells is left out for the sake of simplicity.

Up to this point the wiring of the system is homogeneous: Each element is entirely equivalent to its neighbours of the same class, if one considers only relative coordinates. We now add an element of irregularity. Let s_{ik} be the strength of connection between the fibre i and the E-cell k. It is chosen to be an element of a set of random numbers (Fig. 2). This set was arbitrarily assumed to have uniform distribution in an interval $[0, s]$. No correlation is assumed between the numbers s_{ik} for different k. (We will introduce later a correlation between the s_{ik} for different values of i.) If $A_i^*(t)$ denotes the signal on the afferent fibre i, then the afferent input into the cell k is $\sum_i s_{ik} A_i^*(t)$.

A word has to be said about retinotopic organization here: It is well known, that there is a continuous mapping from the retina onto the visual cortex. Suppose, that this is also true for the model in the sense that the receptive field position on the retina and the geometrical centers of the fields of projection of the afferent fibres correspond to a continuous mapping. But the probabilistic distribution of connection strengths within the field of projection will lead to a random scatter in the "centers of gravity" of the fields of projection and the retinotopic organization will be upset on a small scale. For the present study one can forget retinotopic organization altogether since the fields of projection of neighbouring retinal cells overlap considerably. Suppose that the small

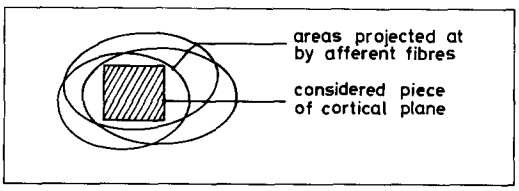

Fig. 3. View onto the cortical plane, showing the overlap of the fields of projection of different fibres. The hatched area lies entirely within the region of overlap and so the information about the exact positions of the receptive fields of the afferent fibres is lost within this area

piece of cortex considered here lies entirely within the region of overlap of the fibres coming from the small piece of relevant retina, as is illustrated by Fig. 3. Within this small cortical area all afferent fibres can then be regarded as equivalent in spite of their different position on the retina.

Several experimental findings on cat and monkey suggest, that also in reality retinotopical organization gives way to random scatter on a small scale: If one records during one cortical electrode penetration from successive neurones, one will find a large random scatter superimposed on the slow systematic displacement due to retinotopic organization (Hubel and Wiesel, 1962; Albus, 1973). In addition, if one maps the two receptive fields of binocular units, one finds a disparity in their positions which changes unsystematically from cell to cell (Joshua and Bishop, 1970). This suggests a still larger individual scatter in the course of single afferent fibres, as the position of the receptive field corresponds to an average position of the fibres constituting it. An additional argument is the rather irregular course of afferent fibres, seen on Golgi pictures (Ruiz-Marcos and Valverde, 1970; Ramon y Cajal, 1955, p. 613).

d) The Learning Principle

The system as it is described up to now is not yet able to explain the experimental facts stated in the introduction. For this, a process of self-organization is required, i.e. the system has to have the possibility to modify itself. This is done in the following way: if during a stimulation the afferent fibre i is active and if the stimulus leads to the firing of the E-cell k, then s_{ik}, the strength of connection between fibre i and cell k is increased by an increment Δs_{ik}. This corresponds to synaptic learning as it was proposed earlier in one form or another (Hebb, 1949; Uttley, 1970; Brindley, 1969; Marr, 1971). The learning principle as defined here leaves the question open by which mechanism such type of synaptic learning may be brought about: by a chemical change within the existing synapses, by a change of their position, by an increase in the number of synapses or by a change of their dimensions. There are morphological data which support the last two alternatives (Cragg, 1968; Møllgaard *et al.*, 1971).

The principle, as it was just stated, leaves one main problem. If s_{ik} is increased by a constant amount Δs at each time when a coincidence $i-k$ takes place, then this will lead to synaptic strengths which will grow forever and eventually will cause instability of the circuit. One way out would be to let the s_{ik} saturate: the increments get smaller and smaller as s_{ik} approaches a maximum value. For this model we choose a different solution, which is stated in the form of a learning principle:

> if there is a coincidence of activity in an afferent fibre i and a cortical E-cell k, then s_{ik}, the strength of connection between the two, is increased to $s_{ik}+\Delta s$, Δs being proportional to the signal on the afferent fibre i and to the output signal of the E-cell k. Then all the s_{jk} leading to the same cortical cell k are renormalized to keep the sum $\sum_j s_{jk}$ constant.

This last step could correspond to the idea, that the total synaptic strength converging on one neurone is limited by the dendritic surface available. It means, that some s_{ik} are increased at the expense of others.

III. The Function of the Model

a) Basic Equations of Evolution

What functional states are there for the model network described in the last section and how will it be influenced by stimulation? To answer these questions we have to write down the equations, which gowern the evolution of the system. They are summarized in Table 1 (compare Grossberg, 1972). At first sight they look like linear differential equations. But H_k^* is a nonlinear function of H_k and this nonlinearity is essential, i.e. one cannot get rid of it by approximations. There are no mathematical ways to solve these equations in a closed form and that is the reason we had to do numerical calculations on a computer.

In this paper we are interested only in static stimuli, i.e. stimuli which are switched on for a moment and switched off again. Ideally the network's response will be an initial transient settling down to a

Table 1. Equations of evolution in time

$H_k(t)$	Excitatory state (ES) of cell k at time t
θ_k	Threshold of cell k
$H_k^*(t)$	Signal of cell k
	$H_k^*(t) = \begin{cases} H_k(t) - \theta_k & \text{if } H_k(t) > \theta_k \\ \text{zero} & \text{otherwise} \end{cases}$
$A_i^*(t)$	Signal of afferent fibre i
N	Number of cortical cells
M	Number of afferent fibres
α_k	Decay constant of ES of cortical cell k
s_{ik}	Strength of connection between fibre i and cell k
p_{lk}	Strength of connection from cell l to cell k

$$\frac{d}{dt} H_k(t) = -\alpha_k H_k(t) + \sum_{l=1}^{N} p_{lk} H_l^*(t) + \sum_{i=1}^{M} s_{ik} A_i^*(t), \quad k = 1, \ldots, N$$

Table 2. Stationary equations

E_k	ES of E-cell number k
I_k	ES of I-cell number k
E_k^*, I_k^*	Corresponding signals
N	Number of E-cells and number of I-cells
M	Number of afferent fibres
Strengths of Connections:	
$p_{lk} > 0$	from E-cell l to E-cell k
$q_{lk} > 0$	from I-cell l to E-cell k
$r_{lk} > 0$	from E-cell l to I-cell k
$s_{ik} > 0$	from afferent fibre i to E-cell k

$$\left.\begin{aligned} &\text{a)} \quad E_k = \sum_{\substack{l \neq k \\ l=1}}^{N} p_{lk} E_l^* - \sum_{l=1}^{N} q_{lk} I_l^* + \sum_{i=1}^{M} s_{ik} A_i^* \\ &\text{b)} \quad I_k = \sum_{l=1}^{N} r_{lk} E_l^* \end{aligned}\right\} \quad k = 1, \ldots, N$$

steady state, which lasts until the end of the stimulus period. As the details of the switching-on and -off periods are irrelevant for this paper, we will restrict our argument to the steady state, however short it might be in reality.

The specialization of the equation of Table 1 to the steady state, or $dH_k/dt = 0$ reads:

$$\alpha_k H_k = \sum_{l=1}^{N} p_{lk} H_l^* + \sum_{i=1}^{M} s_{ik} A_i^*, \quad k = 1, \ldots, N.$$

Here one can divide by α_k and absorb the factor $1/\alpha_k$ on the right side into the definition of the coefficients p_{lk} and s_{ik} giving p'_{lk} and s'_{ik}:

$$H_k = \sum_{l=1}^{N} p'_{lk} H_l^* + \sum_{i=1}^{M} s'_{ik} A_i^*, \quad k = 1, \ldots, N.$$

These equations can be made more explicit, as according to the conventions of the model many of the p_{lk} and s_{ik} vanish and others are negative. By the introduction of more specialized symbols the final set of equations of Table 2 is obtained. (The primes are dropped again, as no confusion can arise.)

b) Specification of Details

Many of the definitions used in the description of the model above were unprecise, because too many details would have obscured the principle. These details have now to be specified. Some of these features will still be found to be oversimplified, but their special form was dictated by the necessity to economize computer time and space.

A hexagonal array was chosen for E- and I-cells (see Fig. 4). For every E-cell there is a corresponding I-cell occupying the same place. The hexagonal arrangement has the advantage of giving to each cell an almost circular surround of neighbouring cells. The total number of cells was chosen to be $2 \times 169 = 338$ cells, giving a network of a major diameter of 15 cells (Fig. 6). The threshold of all the cells was made equal to 1.

The wiring of the network is explained in Fig. 4: Each active E-cell directly excites the immediately neighbouring E-cells with strength p and the immediately neighbouring I-cells (including the one occupying the same place as the active cell) with strength r. Each active I-cell inhibits with strength q its next-to-immediate neighbours amongst the E-cells only.

These distributions are rather crude representations of the bell shaped curves of Fig. 1. The fact, that

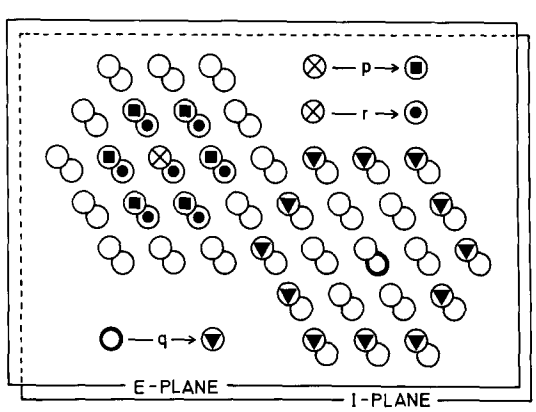

Fig. 4. A small part of the simulated cortex, showing the hexagonal array of the E-cells (upper plane) and the I-cells (lower plane). The different symbols are used to designate those cells which are connected with strengths, p, r and q. Every cell is connected with its neighbors in the same way (except at the borders)

Table 3. Numerical parameters (Definitions see text and Table 2)

$N = 169$
$M = 19$
$p = 0.4$
$q = 0.3$
$s = 0.25$
$r = 0.286$
$h = 0.05$

the I-cells do not inhibit their directly neighbouring E-cells has no major consequences on the function of the model, as was checked in separate calculations. It is a measure of economy.

The afferents are made up to 19 fibres. Each fibre is thought to fan out into 169 branches to contact the 169 E-cells. Correspondingly, a matrix of 19×169 numbers representing the strengths of connection had to be specified. This matrix was derived from a set s'_{ik} ($i = 1, \ldots, 19$; $k = 1, \ldots, 169$) of random numbers with a flat distribution within an interval $[0, s]$. (For the values of s and all the other numerical parameters see Table 3.)

With the numbers s'_{ik} the total afferent synaptic strength s'_k leading to the E-cell k can be written $s'_k = \sum_{i=1}^{19} s'_{ik}$. This number has to be a constant during learning. For the sake of simplicity, all the s'_k were changed to their mean value, which is $19 \cdot s/2$. This was done by renormalizing the s'_{ik} according to:

$$s_{ik} = s'_{ik} \cdot 19 \cdot \frac{s}{2} / s'_k, \quad i = 1, \ldots, 19; \quad k = 1, \ldots, 169.$$

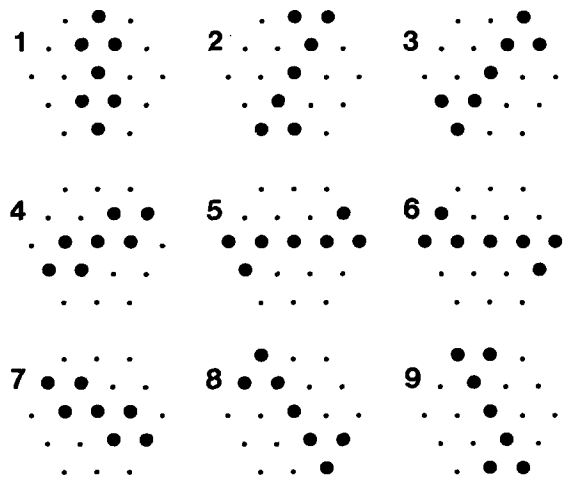

Fig. 5. The standard set of stimuli used on the model "retina". Large and small dots represent active and non-active fibres respectively

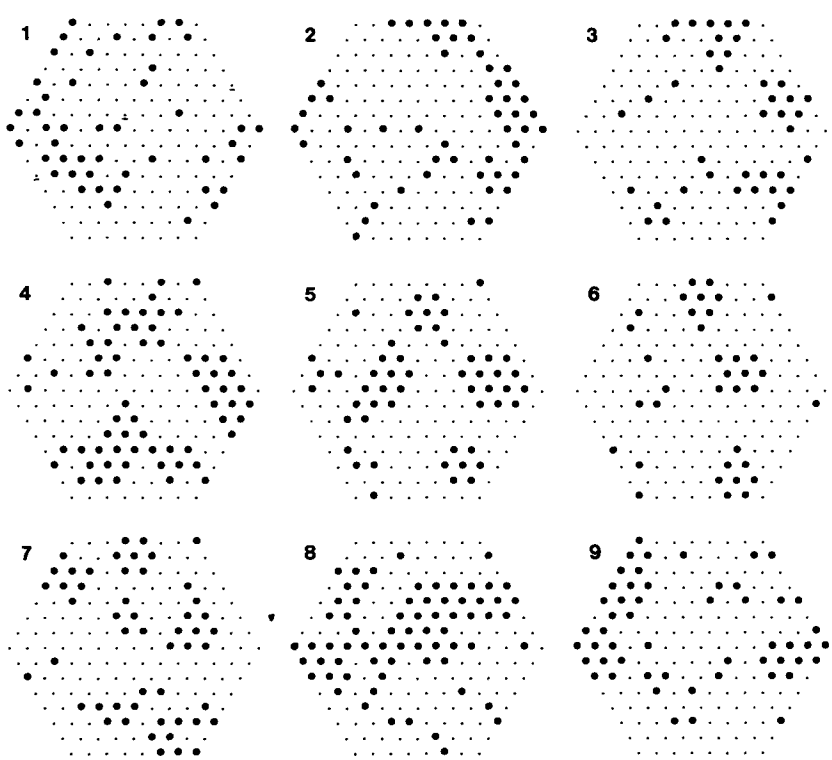

Fig. 6. The reactions of the cortical E-cells to the stimuli of Fig. 5. Large dots represent firing cells. The numbers to the upper left correspond to those of Fig. 5

The s_{ik} thus derived are the starting values of the afferent strengths of connection. All the simulations of this paper are based on the same initial set of synaptic strengths.

In each learning step the s_{ik} values are increased to

$$s'_{ik} = s_{ik} + h \cdot A_i^* \cdot E_k^*$$

after the stimulation. Then the s'_{ik} are renormalized in the way just described to give the s_{ik} of a next generation.

The stimuli always consisted of seven active afferent fibres, i.e. seven of the 19 A_i^* in equation a) of Table 2 were set to 1, the others to zero.

There was a standard set of nine stimuli, which was employed mainly. This set is shown in Fig. 5. It was chosen to represent light bars of different orientation.

It should however be emphasized that in the absence of retinotopical organization the important property of the stimuli in Fig. 5 is not their geometrical arrangement but rather their relationships established by mutual overlap.

c) *The Procedure of the Numerical Calculations*

The solutions to the equations of Table 2 were found by numerical calculation on a UNIVAC 1108. The method we employed was stepwise approximation by an iterative procedure. That means, the equations we really used were those of Table 1, the different steps of the approximation corresponding to their solution at consecutive time steps. Not every set of parameters p, q, r and s lead to stable solutions. Those finally employed were found partly by trial and error.

The solution is not reached in a monotone and quick way. The ES of the cells will first follow a course of damped oscillations and then approach slowly their final values. To save time, this slow approximation was stopped after 20 iterative ("time"-) steps and the result taken as approximate solution to the equation of Table 2. By then the change from step to step in the ES of most cells was smaller than 0.5%. After the solution was found, learning was done by the described manipulations on the s_{ik}. All the other parameters were held fixed.

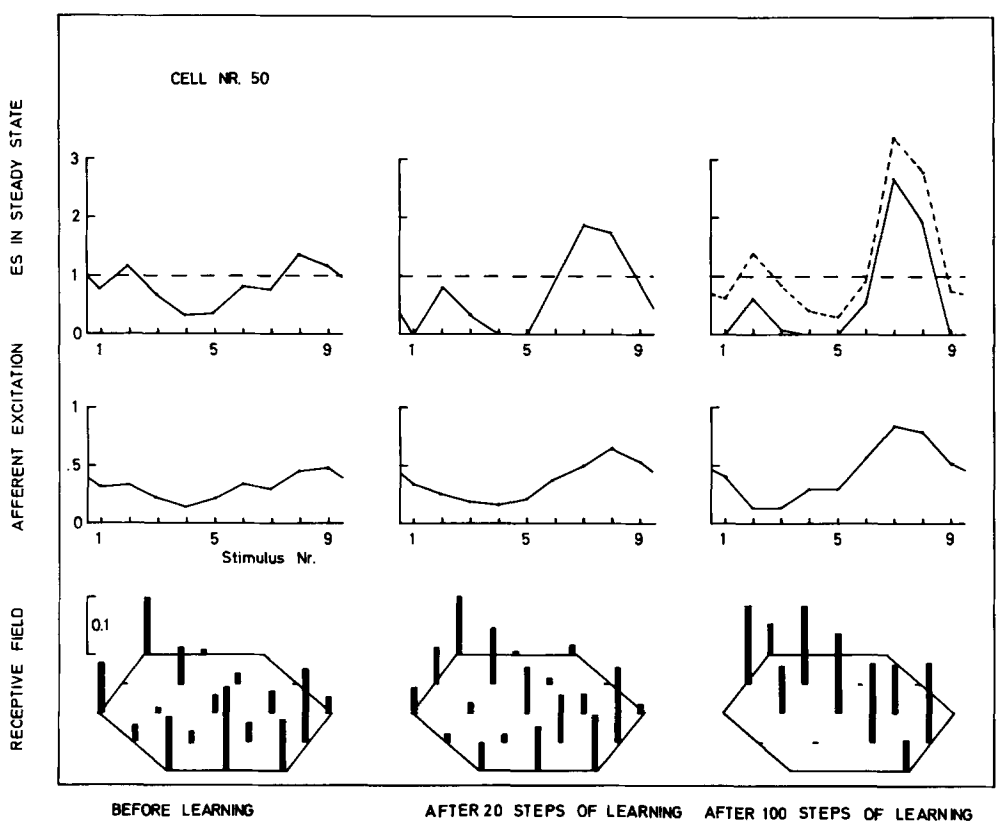

Fig. 7. Receptive field organization, afferent excitation and ES of the E-cell No. 50 in the steady state response to the nine stimuli. Left, middle and right column correspond to the system without learning, with 20 steps and with 100 steps of learning, respectively. The heights of the vertical bars in the hexagon in lower row represent the connection strengths s_{ik} of the 19 retinal fibres to the cell. Their arrangement corresponds to Fig. 5. The position of this cell is underlined in the first diagram of Fig. 6 in the fifth line from the top

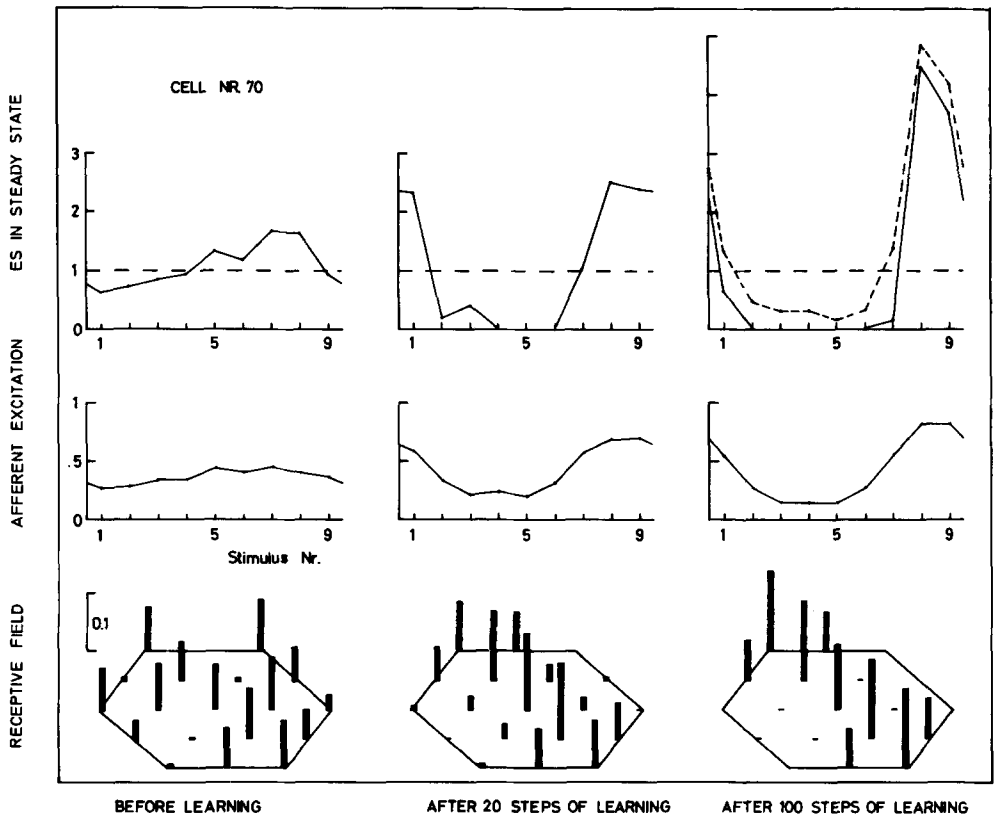

Fig. 8. Receptive field organization, afferent excitation and ES of the *E*-cell No. 70. Its position on the cortex is in the seventh line from the top (see Fig. 6, 1). For explanation see Fig. 7

d) The Results without Learning

Figure 6 shows in a qualitative way the reaction of the network to the nine stimuli (Fig. 5), before any learning took place. The small and large dots represent *E*-cells with ES below and above threshold respectively. The *I*-cells are not shown. As can be seen, the cells have already the clear tendency to fire in clusters. To get a more quantitative impression of the reaction of the cells consider the left column of diagrams in Figs. 7, 8 and 9, which summarize the behaviour of three typical cells, Nos. 50, 70 and 120 (for their positions see Fig. 6). The vertical bars in the hexagon in the bottom row show the connection strengths s_{ik} of the 19 afferent fibres to the cell. Their hexagonal arrangement corresponds to the one in Fig. 5. For each stimulus the afferent excitation (which corresponds to the sum of seven of the bars) is plotted in the middle row of Figs. 7–9 against stimulus number. Notice, that the points in these diagrams tend to form a continuous line, i.e. neighbouring points are correlated. This is a consequence of the retinal overlap between neighbouring stimuli.

The upper graphs of the left columns in Figs. 7–9 show the ES of the particular cells in response to the nine stimuli. This plot could also be called the cells orientational tuning curve. Its details are determined roughly by the afferent excitation, although it is modified by what goes on in the cortical neighbourhood.

Table 4. Classification of cells according to reaction type

a) Classification of tuning curves

	No response	Unimodal	Multimodal
Before learning	12	87	70
20 steps of learning	43	118	8
100 steps of learning	21	147	1

b) Width of unimodal tuning curves (n is the number of neighboring stimuli to which the cells responded)

n	1	2	3	4	5	6	7
Before learning	20	24	18	19	5	—	1
20 steps of learning	24	19	45	25	5	—	—
100 steps of learning	8	43	64	25	7	—	—

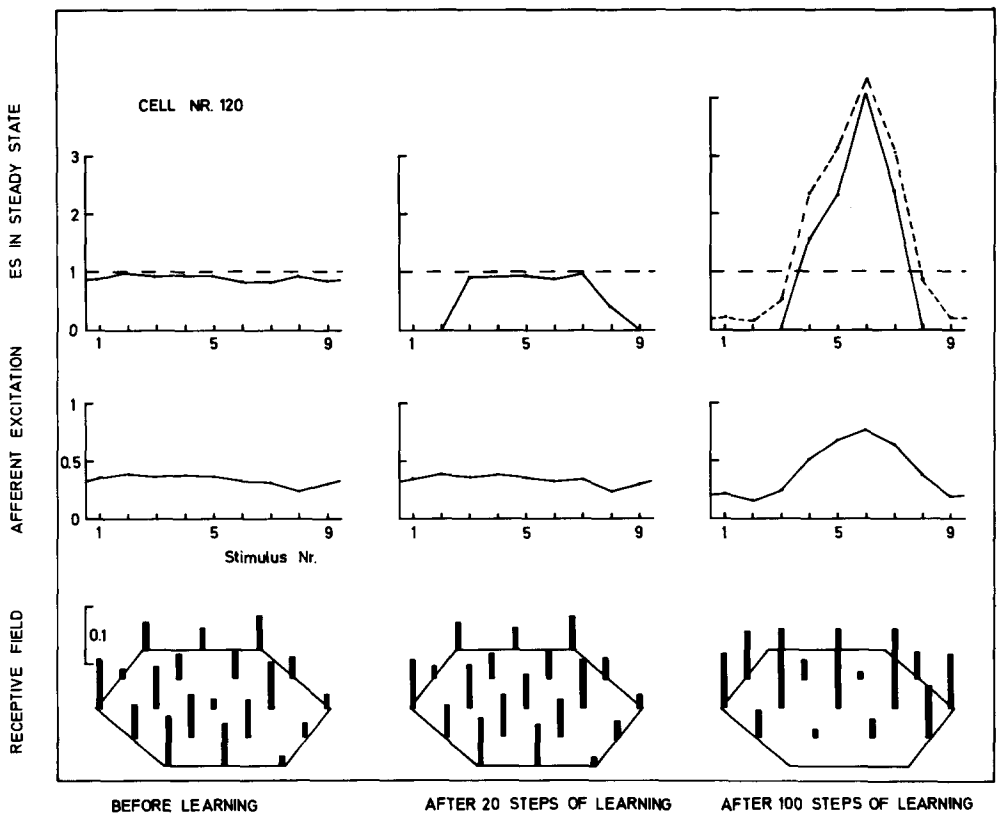

Fig. 9. Receptive field organization, afferent excitation and ES of the E-cell No. 120. Position in the eleventh line from the top (Fig. 6, 1). For explanation see Fig. 7

A summary of the main features of the 169 tuning curves is shown in Table 4. Twelve cells never reach threshold. Seventy of them react to stimuli in a multimodal fashion, i.e. they fire within separate regions. This is exemplified by the cell in Fig. 7 (left column). A large fraction, eighty-seven, of the cells, could already be called orientation sensitive, as their tuning curves are unimodal, although not very sharp (e.g. cell number 70, Fig. 8). In summary one can say, that although there is no systematics in the organization of the afferents, we already get a tendency to firing in clusters ("columnar organization") and a considerable fraction of the cells (51 %) with unimodal orientation tuning ("orientation specific units"), although the tuning curves may still be comparatively flat.

e) The Results after a Learning Phase

A step of learning consists of one presentation of all the nine different stimuli and the learning manipulations on the s_{ik} subsequent to each presentation.

During the learning phase the sequence of presentation of the stimuli was 1, 6, 2, 7, 3, 8, 4, 9 instead of the natural one (1, 2, 3, ...). This was to avoid special effects resulting from the consecutive presentation of two maximally overlapping stimuli.

The behaviour of the system after 20 steps of learning is shown in Fig. 10. The cells now show a much stronger tendency to fire in clusters than they did before learning (compare Fig. 6). This can be interpreted in the following way: the intracortical connections give the cells a natural tendency to fire in separate clusters. At first the cells are disturbed in this tendency by the afferent excitation, which at the beginning does not favour clusters at all. Nevertheless the ES of the cortical cells will be highest, if they can exchange maximal intracortical excitation and minimal intracortical inhibition, which is the case with firing in patterns of small clusters. Now, the higher the ES of a cortical cell, the stronger will be its influence on the afferent organization via learning. Finally those patterns of afferent organization will persist, which favour cortical clusters.

If the statistics of the afferent stimulations are stationary, the learning process in the system will saturate after some time. This saturation is not yet reached after 20 learning steps. Therefore the reactions

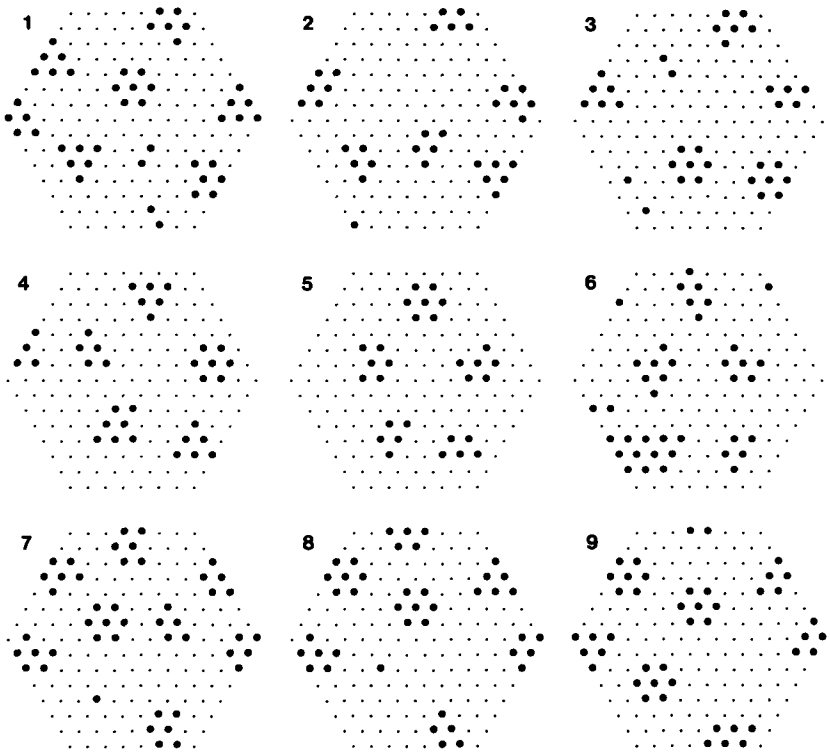

Fig. 10. Reaction of the cortical *E*-cells after 20 steps of learning. The figure corresponds to Fig. 6. In this and the following figure there is no learning between the nine stimulations shown

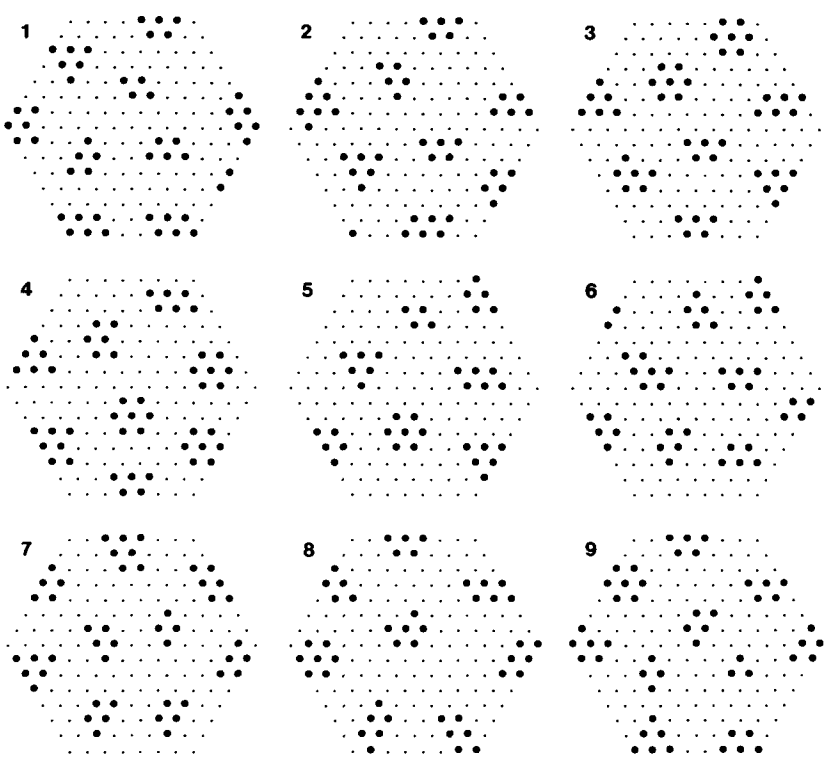

Fig. 11. Reaction of the cortical *E*-cells after 100 steps of learning

of the system continue to change. Figure 11 shows the responses of the E-cells after 100 steps of learning, of which the last 40 were accelerated by doubling the learning constant h (its value being 0.1 instead of 0.05).

To see the effect of learning on the level of the single cells, consider the middle and right columns in Figs. 7–9. In many cases the tuning curves are now much steeper, the cells being either strongly excited or little excited (e.g. Figs. 7 and 8). This behaviour is more pronounced with more learning (see right columns). The cell of Fig. 7 had at first two sensitive regions, one of which disappeared within the first 20 steps of learning. In the case of Fig. 9 there was no reaction of the cell to any of the stimuli up to learning step 20. Later on, however, it was occasionally excited to the firing level by its neighbours. This led to a sudden modulation of its tuning curve, which is very steep after 100 learning steps.

The changes in the tuning curves of the cells with learning are the result of positive feedback: Whenever a cell fires, it will strengthen those afferent connections which were exciting the cell. This leads to increased afferent excitation of the cell, when the same stimulus is applied the next time. Increased afferent excitation will in its turn make the cell fire more strongly in response to the stimulus, and this will lead to accelerated learning as long as saturation is far. The corresponding modifications of the afferent organizations and of the excitation curves are apparent in Figs. 7–9, bottom row. In the end, only those s_{ik} persist which are used by the effective stimuli (compare Fig. 5 for the stimuli). The decrease of the unused connection strengths is a consequence of the condition that their sum be constant.

The chain of positive feedback described in the last paragraph is modified by the intracortical excitation and inhibition, which are added to the afferent excitation. This is the point, where intracortical dynamics enters and leads to the clustering of firing.

The behaviour of all the E-cells with learning is summed up in Table 4. It shows, that less and less cells have multimodal tuning curves (70 before learning, 8 after 20 steps and one after 100 steps). Note also, that very sharp tuning curves, i.e. reaction of a cell to only a single stimulus, seems not to be favoured by the system after extended learning (24 after 20 steps and only 8 after 100 steps).

The broken lines in the upper right graphs of Figs. 7–9 show the excitation of the cells alone, no inhibition being subtracted. The inhibition is the difference between the two curves of each graph. As one can see, the tuning curves are made a bit narrower by the inhibition. In 20 cases a second separate sensitive region of a cell was suppressed by the inhibition, as in the example of the cell of Fig. 7. Therefore in this model the inhibition takes part in the organization of the "cortex" in an essential way (apart from its stabilizing function), although it is homogeneously distributed and not modified by learning at all.

Two of the three aims we set for the model in the introduction of this paper are now accomplished: First, clusters of cortical activity are brought about by intracortical dynamics. Second, organization of orientation specific units is brought about by a learning strategy rather than by genetical determination.

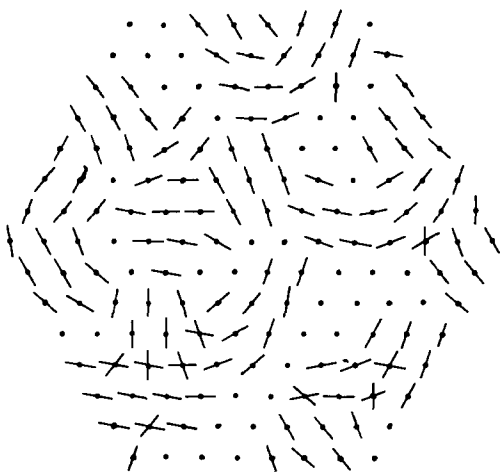

Fig. 12. View onto the cortex. Each bar indicates the optimal orientation of the E-cell (for definition see text). Dots without a bar are cells which never reacted to the standard set of stimuli. Two bars indicate two separate sensitive regions

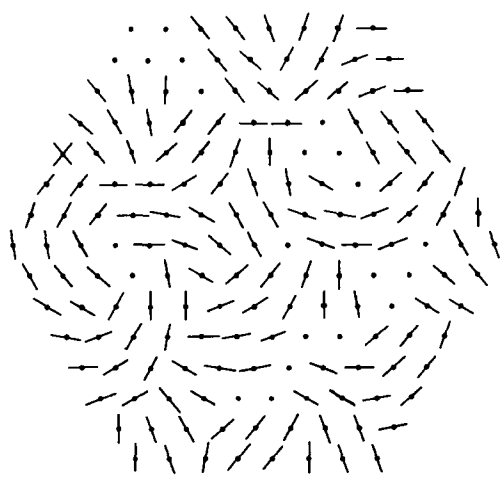

Fig. 13. View onto the cortex after 100 steps of learning

The third task was an explanation of the fact, that neighbouring cells have the tendency to react to stimuli of neighbouring orientation (Hubel and Wiesel, 1963, 1968). This now turns out to be a natural consequence of the existence of clusters and of their influence on the afferent synaptic strength, as can be seen directly by inspection of Fig. 12.

Each bar in this figure indicates the median of the orientations to which the corresponding cell responded. If, for instance, a cell fired in response to stimuli 1, 2 and 3, the orientation corresponding to stimulus 2 is plotted, regardless of the magnitude of the three answers. If the cell fired in response to stimuli 1 and 2, the plotted orientation lies halfway between those for 1 and 2. Two crossing bars indicate a "bimodal" reaction of the cell, i.e. responses to two orientations separated by one or more uneffective orientations.

It can be seen that the probability of similar orientations in adjacent cells is high. This tendency is emphasized in Fig. 13, which shows the optimal orientations after 100 steps of learning. There are even series of cells with continuously turning orientations (e.g. the seventh line in Fig. 13) as described in the literature (Hubel and Wiesel, 1968).

f) The Effect of Non-Standard Stimuli

An important question one can pose now is how the trained system will react to stimuli which it does not know yet. To answer this question, the system was tested with 45 different stimuli m_i, $i=1,\ldots,45$. As the standard stimuli n_k, $k=1,\ldots,9$ of Fig. 5, the m_i consisted of seven retinal points each. The m_i were charaterized by the maximal overlap $V_{i\max}$ they had with any of the n_k:

$$V_{i\max} = \max_{k=1,\ldots,9} (m_i \cap n_k).$$

The m_i were chosen to form five groups of equal $V_{\max}$, containing nine stimuli each. $V_{\max}$ varied from 2 in the first group to 6 in the fifth group. As the model retina is so small, no stimuli with $V_{\max}=1$ could be found. Within one group the stimuli were chosen to be as different as possible.

To judge the effect of a stimulus on the system, the mean output signal E of the E-cells was computed:

$$E = \frac{1}{N} \sum_k E_k^* .$$

In Fig. 14 the average of E for the nine stimuli in one group is plotted against the maximum overlap $V_{i\max}$ with the standard stimuli. The overlap 7 means the standard set itself. The flat curve, (○ — ○) is the effect of the stimuli on the "naive" network, before learning of any stimuli took place. No significant

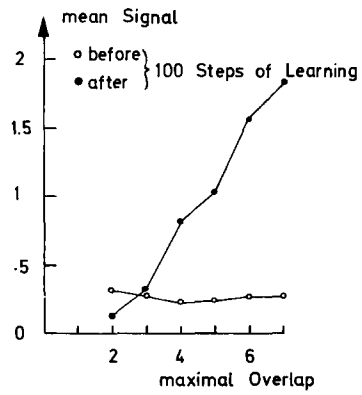

Fig. 14. The mean output signal of the E-cells in response to different sets of stimuli. The stimuli of each set have a maximal overlap with stimuli of the standard set. This overlap is calculated as number of fibres the stimuli have in common and it is shown on the abscissa. An overlap of seven corresponds to the standard set

differences between the groups are found. The steeper curve (● — ●) shows the effect of the different stimulus sets after the network went through 100 steps of learning the standard stimuli, which, in Fig. 14, corresponds to the group of overlap 7. The stimuli most similar to the stimuli used for training are clearly favoured. The response becomes smaller the less similar the test stimuli are to the standard set. In the case of overlap 2 there is even a suppression in comparison with the sensitivity before training. This indicates, that the system responds less to "new" stimuli once it has been tought a given set.

It should be noted, that the mean output signal of the cells increased after learning (e.g. from 0.25 to 1.8 for the standard stimuli in Fig. 14). A more realistic model should be able to keep the mean excitation of the network constant in spite of learning. This could be done by an adapting inhibitory system. The trained network would then actively suppress the response to all stimuli, with which it was not trained. With this modification, even more than in the present model, the network could be regarded as an effective filter.

g) The Sensitivity to Nonspecific Input

Up to now the task of the model was fairly simple: a set of stimuli, which could be characterized by one parameter (their orientation) was presented. The problem for each cortical cell was to become selectively sensitive to a small range of the stimulus parameter only. In reality there are certainly different sources of perturbations:

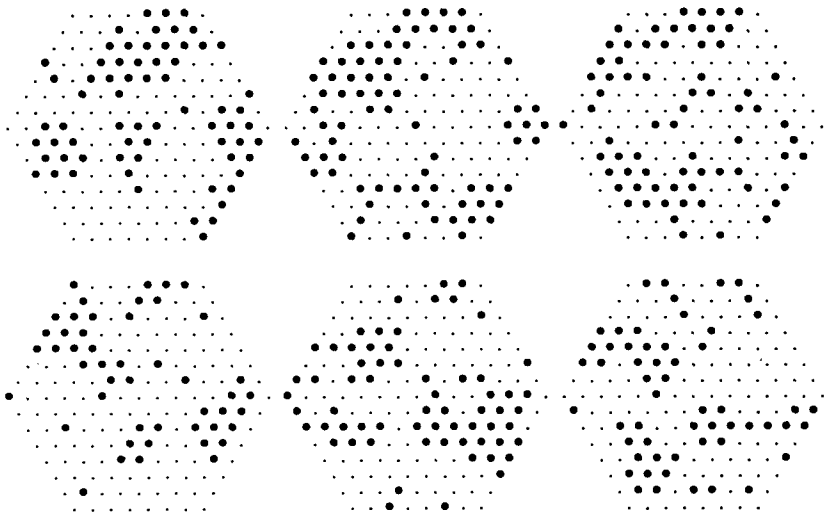

Fig. 15. Six reactions of the E-cells to the same stimulus (1 of Fig. 5) before learning. The differences are produced by extra random input

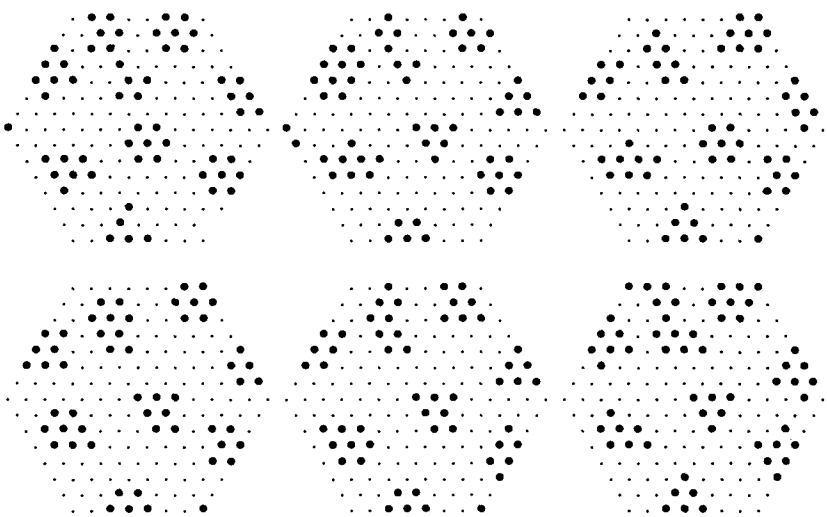

Fig. 16. Six reactions of the E-cells to stimulus 1 after 20 steps of learning

1. There is a very large number of different stimuli, which make up a large percentage of the information flow, but each individual stimulus occurs so rarely that no cortical cell would get specialized to it.

2. Those stimuli occurring frequently have small variations of composition, which can be regarded as a disturbance.

3. There is a nonspecific input to the cells, which is not related to the visual information.

To test the ability of the model to work in the presence of perturbations the following experiment was done. A random number t_k was added to the specific afferent excitation $\sum_i s_{ik} A_i^*$ (see Table 2) received by cell number k ($k = 1, \ldots, 169$). There was a new set of random numbers for every stimulation.

The first generation of the s_{ik} was chosen from the interval [0, 0.175]; the t_k from the interval [0, 0.525], three times as large. Consequently the mean of the expressions $\sum_i s_{ik} A_i^*$ is 0.613 ± 0.095 (root mean square deviation) and the mean of the t_k is 0.263 ± 0.153. Their sum, $\sum_i s_{ik} A_i^* + t_k$, has a mean of 0.875 (as before) and its r.m.s. deviation is ± 0.180, which is largely determined by the perturbation t_k.

The most important feature of the afferent excitatory input to the cortical cells is its differentiation

between different cells during one stimulation and for one cell between different stimulations. These differences have to be detected and enhanced by learning. However in the present experiment they are completely buried under the differences produced by the perturbative random excitation t_k.

How serious the perturbation is can be assessed from Fig. 15, which shows the reactions of the untrained system to six times the same stimulus (number one of Fig. 5). Without the perturbation the reactions would be identical. After 20 steps of learning (involving nine stimuli each and with $h=0.1$, compare Table 3) the picture is quite different: Fig. 16, also six presentations of stimulus one, shows much less variations.

To measure the decrease in variability from Fig. 15 to Fig. 16 by a number, the entropy of these variations was calculated. If E-cell k fired n times during m presentations of the same stimulus, then the probability of k to react can be defined as $p_k = n/m$, and h_k, the corresponding entropy can be calculated

$$h_k = -p_k \operatorname{ld} p_k - (1-p_k) \operatorname{ld}(1-p_k)$$

(ld is the binary logarithm). If k fires half the time, then the entroy is maximum, $h_k = 1.0$. The mean entropy of all the E-cells then is

$$H = \frac{1}{N} \sum_k h_k,$$

where N is the number of cells.

The entropy of the variability of reactions to one and the same stimulus before learning (Fig. 15) is $H=0.674$ and after 20 steps of learning (Fig. 16) it is $H=0.203$, considerably reduced.

In conclusion one can say: If there is a systematic structure in the information arriving at the cortical cells, it can be detected and enhanced by the learning system even if it is buried in nonstructured, random excitation.

h) Redundancy of Information Storage

It has been demonstrated, that the connectivity between two points of nervous tissue can change after simultaneous stimulation of these points (e.g. Bliss and Gardner-Medwin, 1970). These changes may be interpreted as synaptic learning. Unfortunately they were never longlasting and after some hours or sometimes days the connectivity had decayed to its previous value. This fact, which was also found in other preparations, has always been a serious argument against the interpretation of synaptic conditioning as a basis for permanent memory.

With our model we performed an experiment, which may be relevant to this question. In the system as it was left after 100 steps of learning, twelve arbitrarily chosen afferent fibres were strengthened by increasing the corresponding synaptic strengths s_{ik} to triple value, and the synaptic input to the cortical cells was renormalized to keep their total synaptic input constant (s. Section IIIb). This reduced the increased numbers s_{ik} to a somewhat lower value. Then the system was allowed to learn for 40 more steps (with $h=0.1$). It was found, that most of the connection strengths were brought back to a level, which was close to the value before the added increase. Two of the increased s_{ik} were connected to cells which never reacted before, during or after this experiment. Consequently no learning took place and these s_{ik} stayed high. The sum of the remaining 10 s_{ik} was 0.963 before the experiment. After the increase and renormalization it was 2.351, and after the 40 steps of learning it was back to 1.026, although saturation had not yet been reached.

The explanation of this result is that the characteristics of a pattern, to which a cell responds, are determined by all the effective connections leading to the cell. If the strength of only one of these connections is changed, the optimal pattern of the cell will not be changed very much. This one connection will then readapt by the process of learning to its previous value.

The experiment shows, that the system has enough redundancy in its information storage to make it insensitive to such small defects as an arbitrary change in the strength of some connections. This kind of argument could be able to explain the failure of the experiments mentioned above: The conductivity changes artificially produced in the experiment may be considered as disturbances of the normal function of the neural structure and are therefore "repaired" by the nervous system.

IV. Discussion

It was the aim of this paper to show, that there is at least one way to explain a large part of the functional organization in the visual cortex of cat and monkey without depending completely on a genetically predetermined connectivity between the cortex and its afferent fibres.

Most of the principles used here have been described before in the literature: the form of intracortical connectivity (Fig. 1) (Colonnier, 1966; Wilson and Cowan, to be published), mechanisms of synaptical conditioning (Hebb, 1949; Rosenblatt, 1961; Brindley, 1969; Grossberg, 1972; Uttley, 1970; Marr, 1971), and random connections (Beurle, 1956; Rosenblatt, 1961; Marr, 1971). The equations used here are very similar to those of (Grossberg, 1972).

a) The Structure of the Model and Generalizations

This paper proposed two principal mechanisms: the development of pattern sensitive cortical cells by a self-organizing process involving synaptic learning, and the arrangement of functional columns as a consequence of intracortical connections rather than due to a predetermined distribution of afferents.

The two main directions of local intracortical fibre systems are tangential and vertical to the cortical surface. The model takes these two directions into account in a simplified way. The vertical connections are implicit in the representation of all the cells of one vertical cylinder by only two cells, an excitatory and an inhibitory. The underlying assumption is that all cells of this cylinder are so strongly connected by vertical fibres, that they fire virtually simultaneous under most conditions.

The horizontal connections, on the other hand, have been represented explicitly. In contrast to the histological picture they spread symmetrically in all directions. This can be justified by the fact, that each connection between two functional units represents an average over many fibres, which connect a multitude of individual cells. The firing of cortical cells in clusters is due to the excitatory horizontal connections. As a consequence of the homogeneity of intracortical connection the borders between such clusters are not fixed and may shift slightly from one stimulation to the next. The inhibitory and excitatory interaction between cortical neurones makes it necessary to think in terms of collective networks rather than of isolated neurones. This means, that the functional behaviour of a cortical neurone is not only a function of its receptive field in terms of its afferent input, but also of its intracortical connections.

To organize orientation specificity of the cortical cells a mechanism of adaptation, namely synaptic conditioning, is introduced. This mechanism is applied to a network with nonspecific, random interconnections and is able to transform it into a highly specialized system. Some of the special features of the "learning principle" employed here need discussion.

The total sum of synaptic strength converging onto one cell was kept constant: while some synapses grew stronger, others became weaker. This was introduced for several reasons. One is stability: a system with only growing excitatory synapses is unstable. A second reason is the requirement of high specificity of the cells: they should become insensitive to all stimuli for which they were not trained. Both of these functions can probably also be realized by replacing the principle of constant synaptic strengths by additional training of inhibitory connections.

Some of the limitations of the proposed model are caused not by the underlying principles but rather by the restricted number of cells and connections and the highly restricted sensory inputs used in the computer simulations. More cells would give the cortex more degrees of freedom and it could adapt to a greater range of stimuli. An obvious generalization would be the inclusion of moving stimuli. The important stimulus property selected for would then no longer be retinal position but rather a temporal sequence of positions. Another generalization would be the organization of a hierarchical system of feature detectors.

b) Comparison with Experiments

The learning principle was applied only to synapses of the afferent fibres. This may be taken as a special case of learning in early development such as the "learning" of binocularity (Wiesel and Hubel, 1965) or orientation tuning (Blakemore and Cooper, 1970).

Many neurones in the visual cortex of very young kittens are orientation sensitive before any visual experience has occurred, but they are not so sharply tuned as in adult animals (Hubel and Wiesel, 1963; Pettigrew, 1972). In the light of these findings it is interesting to note that also in the proposed model a high percentage of orientation sensitive cells is found before any training took place (Table 4). Narrowly tuned neuronal orientation specificity found during experiments on unexperienced kittens may be a consequence of fast learning during the experiments, a property also shown by the model (see below).

The model leads to one important "prediction". If only a restricted set of stimuli is presented during the training period, the cortical neurones will specialize to these stimuli and will thereafter become insensitive to all other stimuli. This corresponds to experimental findings on training of young animals: if kittens are raised in an environment consistling entirely of horizontal or vertical stripes, they become virtually blind for contours perpendicular to the orientation they had experienced during the training period (Blakemore and Cooper, 1970). When testing the visual cortex cells of such animals a highly significant anisotropy in the distribution of preferred orientations was found. It was later shown that an exposure time as short as one hour during a sensitive period was sufficient to induce the anisotropy of preferred orientations (Blakemore and Mitchell, 1973). This compares well with the quick convergence of the self-

organization in the model network, which became relatively insensitive to untrained stimuli after as little as a hundred presentations of the set of training stimuli. This phenomenon like the corresponding experiments on unexperienced kittens may be an analogue to imprinting rather then learning in the definition of ethologists.

Recent experiments in this laboratory make it doubtful, that many geniculate fibres, the receptive fields of which are arranged in a line parallel to the optimal orientation, converge on individual cortical "simple" cells. The excitatory input of such cells appears to have a receptive field the form and size of which correspond to those of individual retinal ganglion cells rather than to lines (Benevento, Creutzfeldt and Kuhnt, 1972). It seems that the temporal sequence of retinal excitation, i.e. the stimulus movement is a more important aspect of cortical organization. In order to test this, the dynamic aspects rather than steady state conditions may have to be investigated with our model.

The results of this theoretical study are encouraging as they show, that such simple assumptions about intracortical connections and the mechanism of synaptic conditioning as made in this model proved to be sufficient to explain some of the most striking functional properties of the visual cortex.

Acknowledgements. The author would like to thank Professor Dr. Otto Creutzfeldt for providing excellent working conditions and for helpful suggestions, Dr. H. Wässle for critical remarks on the manuscript and Dr. Jean Ennever for correcting my English. The numerical calculations have been done on the UNIVAC 1108 computer of the Gesellschaft für wissenschaftliche Datenverarbeitung, Göttingen.

References

Albus, K.: Topology of orientation sensitivity in the cortical areas 17 and 18 of the cat. Pflügers Arch. Suppl. to **339**, R 91 (1973)

Benevento, L. A., Creutzfeldt, O. D., Kuhnt, U.: Significance of intracortical inhibitions in the visual cortex. Nature (Lond.) **238**, 124—126 (1972)

Beurle, R. L.: Properties of a mass of cells capable of regenerating pulses. Phil. Trans. Roy. Soc. (Lond.) B **240**, 55—94 (1956)

Blakemore, C., Cooper, G. F.: Development of the brain depends on the visual environment. Nature (Lond.) **228**, 477—478 (1970)

Blakemore, C., Mitchell, D. E.: Environmental modification of the visual cortex and the neural basis of learning and memory. Nature (Lond.) **241**, 467—468 (1973)

Bliss, T. V. P., Gardner-Medwin, A. R.: Long-lasting increases of synaptic influence in the unanaesthetized hippocampus. J. Physiol. (Lond.) **216**, 32—33 P (1971)

Brindley, G. S.: Nerve net models of plausible size that perform many simple learning tasks. Proc. Roy. Soc. (Lond.) B **174**, 173—191 (1969)

Colonnier, H. L.: Structural design of the neocortex. In: Eccles, J. C. (Ed.): Brain and conscious experience, p. 1—21. Berlin-Heidelberg-New York: Springer 1966

Cragg, B. G.: Are there structural alterations in synapses related to functioning? Proc. Roy. Soc. B **171**, 319—323 (1968)

Grossberg, S.: Neural expectation: cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes. Kybernetik **10**, 49—57 (1972)

Hebb, D. O.: Organization of Behaviour. New York: John Wiley 1949

Hirsch, H. V. B., Spinelli, D. N.: Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. Science **168**, 869—871 (1970)

Hubel, D. H., Wiesel, T. N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. (Lond.) **160**, 106—154 (1962)

Hubel, D. H., Wiesel, T. N.: Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. J. Neurophysiol. **26**, 994—1002 (1963)

Hubel, D. H., Wiesel, T. N.: Receptive fields and functional architecture of monkey striate cortex. J. Physiol. (Lond.) **195**, 215—243 (1968)

Joshua, D. E., Bishop, P. O.: Binocular single vision and depth discrimination. Receptive field disparities for central and peripheral vision and binocular interactions on peripheral single units in cat striate cortex. Exp. Brain. Res. **10**, 389—416 (1970)

Marr, D.: Simple memory. Phil. Trans. Roy. Soc. (Lond.) B **262**, 23—81 (1971)

Møllgaard, K., Diamond, M. C., Bennett, E. L., Rosenzweig, M. R., Lindner, B.: Quantitative synaptic changes with differential experience in rat brain. Int. J. Neurosci. **2**, 113—128 (1971)

Pettigrew, J. D.: The importance of early visual experience for neurones of the developing geniculostriate system. Invest. Ophthal. **11**, 386—392 (1972)

Ramon y Cajal, S.: Histologie du système nerveux, Vol. II. Madrid: Consejo Superior de Investigationes Cientificas, Instituto Ramon y Cajal 1955

Rosenblatt, F.: Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. Washington D.C.: Spartan Books 1961

Ruiz-Marcos, A., Valverde, F.: Dynamic architecture of the visual cortex. Brain Res. **19**, 25—39 (1970)

Uttley, A. M.: The informon: a network for adaptive pattern recognition. J. theor. Biol. **27**, 31—67 (1970)

Wiesel, T. N., Hubel, D. H.: Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. J. Neurophysiol. **28**, 1029—1040 (1965)

Dr. Christoph von der Malsburg
Max-Planck-Institut für Biophysikalische Chemie
D-3400 Göttingen
Am Faßberg, Postfach 968
Federal Republic of Germany