Chapter 5

The influence of the spatial and temporal wave properties upon development

5.1 Introduction

In Chapter 4 we introduced the Keesing model of retinogeniculate pathway development, investigating in detail the effect of the initial weight bias upon development and the role of normalisation schemes in the network. In this chapter we use the same model to explore the influence of spatial and temporal wave properties upon the development of the retinogeniculate pathway. The chapter is divided into three main experimental sections. First, we examine the temporal properties of the waves by seeing how the rate of wave generation affects development. Second, we vary the width of waves to investigate the spatial properties of the waves. Third, we look at the correlations required between on- and off-centre retinal units to drive development of polarity segregation within units. The results from the model are then compared with the current biological data about retinal waves.

5.2 Exploring the temporal properties of retinal inputs

One of the useful features of this model is that it allows us to vary the temporal properties of the retinal waves to see how development is affected. The model results will be compared with the biological data concerning temporal information later in Section 5.6 after the model results have been presented. Intuitively, the temporal properties of the waves should affect development, especially the monocular status of units: the more often waves are present in each eye, the more likely that units in both eyes are jointly active. This in turn can cause postsynaptic units to remain binocular since presynaptic units in different eyes will be correlated to some degree. In this section, we investigate the temporal properties of the retinal waves by measuring how the rate of wave generation affects development of ocular dominance and topography in the geniculate. Before describing these experiments, the next section analytically describes how the rate of wave generation determines the probability of joint activity in the two eyes.

5.2.1 The relationship between the rate of wave generation and the probability of activity in both eyes

When the waves are independently generated in each eye, three different states describe the activity across the two eyes: either both eyes are active, one of the two eyes is active, or neither eye is active. To calculate the overall probability of both eyes being active at the same time (O_2), one eye being active on its own (O_1), or both eyes being quiet at the same time (O_0), it is first necessary to determine the average delay between one wave finishing and the next wave starting. This value, which shall be called S, depends on p_w, the rate at which new waves are generated.

After an eye has recovered from its refractory period, the probability that a new wave starts in an eye after 0 time steps is p_w. If the wave does not begin at time step 0, then it will start after time step 1 with probability $(1-p_w)p_w$. Similarly, the probability of a new wave starting at time step 2 is $(1-p_w)(1-p_w)p_w$. In general, the probability of a new wave starting at time step n is $(1-p_w)^n p_w$. The average time when the eye is silent between waves, S, is therefore given by the infinite sum:

$$\begin{aligned} S &= 0 \times p_w \\ &\quad + 1 \times (1-p_w)p_w \\ &\quad + 2 \times (1-p_w)(1-p_w)p_w \\ &\quad + \cdots \\ &\quad + n \times (1-p_w)^n p_w \\ &\quad + \cdots \\ S &= \sum_{k=0}^{\infty} k(1-p_w)^k p_w \end{aligned} \tag{5.1}$$

If we consider the geometric progression $y = \sum_{k=0}^{\infty} q^k$ and take its derivative with respect to q:

$$\begin{aligned} y &= \sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \\ \frac{dy}{dq} &= \sum_{k=0}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2} \end{aligned} \tag{5.2}$$

This can be substituted into the definition of S:

$$\begin{aligned} S &= \sum_{k=0}^{\infty} k(1-p_w)^k p_w \\ &= p_w \sum_{k=0}^{\infty} k(1-p_w)^{k-1}(1-p_w) \\ &= p_w(1-p_w) \sum_{k=0}^{\infty} k(1-p_w)^{k-1} \\ &= p_w(1-p_w)\left(\frac{1}{1-(1-p_w)}\right)^2 \qquad \text{(using 5.2 with } q = 1-p_w\text{)} \\ S &= \frac{p_w(1-p_w)}{p_w^2} = \frac{1-p_w}{p_w} \end{aligned} \tag{5.3}$$

Once S is known, the overall probability that an eye is active at any time step, O is given by:

$$O = \frac{T}{S+R+T} \tag{5.4}$$

where R is the refractory period, and T the number of time steps that the wave is present on the retina. Once O is calculated, the probabilities O_2, O_1 and O_0 are given by:

$$O_2 = O \times O = O^2 \tag{5.5}$$

$$O_1 = O(1-O) + (1-O)O = 2O(1-O) \tag{5.6}$$

$$O_0 = (1-O) \times (1-O) = (1-O)^2 \tag{5.7}$$

This analysis can be generalised to n eyes, all independently generating waves: in this case, the probability that i out of n eyes are simultaneously active, O_i, is given as:

$$O_i = (1-O)^{n-i} \times O^i \times C_i^n, \text{where} \quad C_i^n = \frac{n!}{i!(n-i)!} \tag{5.8}$$

Figure 5.1 shows how these probabilities vary with the probability that new waves are generated. The overall probabilities for some specific values of p_w are additionally listed in Table 5.1.

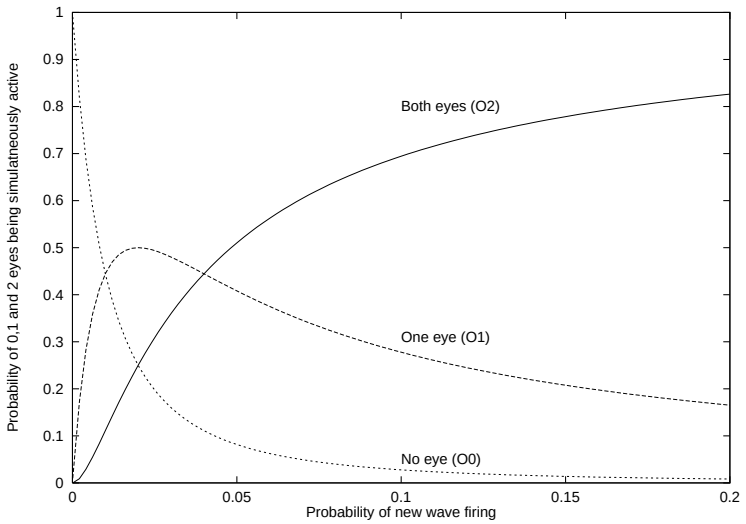


Figure 5.1: Overall probabilities of joint retinal activity as a function of the probability of wave generation. This graph shows the overall probabilities for values of p_w only in the range $[0, 0.2]$. Beyond this range, O_0 continues to decay to zero, with O_2 steadily climbing to a value of 0.96 as p_w tends to 1.0.

5.2.2 The effect of wave generation rate upon ocular dominance and topography

To assess the importance of the temporal properties of the waves upon development, a set of experiments were created using identical conditions to those used before (in Figure 4.9), but using different values for the probability of wave generation in both eyes. As shown in Figure 5.1, there are three main different ranges for the parameter p_w (actual values for p_w are given here in brackets assuming $T = 50, R = 1$):

p_w	O	O_0	O_1	O_2
0.002	0.091	0.826	0.165	0.008
0.003	0.130	0.756	0.227	0.017
0.005	0.200	0.640	0.320	0.040
0.020	0.500	0.250	0.500	0.250
0.200	0.909	0.008	0.165	0.826
0.500	0.962	0.001	0.074	0.925
0.800	0.976	0.001	0.047	0.952

Table 5.1: Theoretical probability of eye activity as a function of p_w. Measurements of the corresponding probabilities from the simulations of eye activity were always within four percent of the predicted values.

1. $O_0 > \max(O_1, O_2)$ $(p_w \leq 0.01)$. Here it is most likely that both eyes are quiet.

2. $O_1 > \max(O_0, O_2)$ $(0.01 < p_w < 0.04)$. In this range, the most likely situation is that one eye is active while the other eye is quiet.

3. $O_2 > \max(O_0, O_1)$ $(p_w \geq 0.04)$. In this last range (the widest of the three ranges), both eyes are likely to be jointly active most of the time.

A set of values for p_w were therefore chosen to cover each of the these three ranges. The values of p_w used in the experiments are listed in Table 5.1 along with the theoretical probability that both eyes are jointly active. (For comparison, all of the experiments described in Chapter 4 used values of $p_w = 0.02$ for each eye. This corresponds to a value of $O = 0.5$, meaning that, overall, each eye is active for fifty percent of the simulation.)

The correlations between presynaptic units for two extreme values of p_w are compared in Figure 5.2. Despite the large difference in the value of p_w, the only noticeable difference in the correlations are that the within-eye anticorrelations are slightly stronger for the larger value of p_w. Figure 5.3 shows the effect of varying p_w upon development. From this figure it can be seen that normal development depends upon the value of p_w. First, if p_w is too small, most of the units remain binocular and there is no topographic refinement of receptive fields. On the other hand, if p_w is too big, although most units become strongly monocular, the topography begins to break up, producing holes within the representation of the visual space for each row of postsynaptic units (for example, when $p_w = 0.2$).

5.2.3 Introducing the active-covariance rule

The case when p_w is too small is interesting, since it represents the situation when most of the time both eyes are quiet. This should produce ideal conditions for the development of monocularity because units in the two eyes are not likely to be active at the same time. Instead of the weights becoming monocular though, all of the weights in the network developed to the same value, producing completely binocular units with no retinotopic refinement.

An analysis of the covariance rule revealed the reason behind this lack of refinement of receptive fields. In this model, postsynaptic activity depends linearly on presynaptic activity. Given that most of the time the presynaptic units are quiet, the postsynaptic units are also usually quiet.

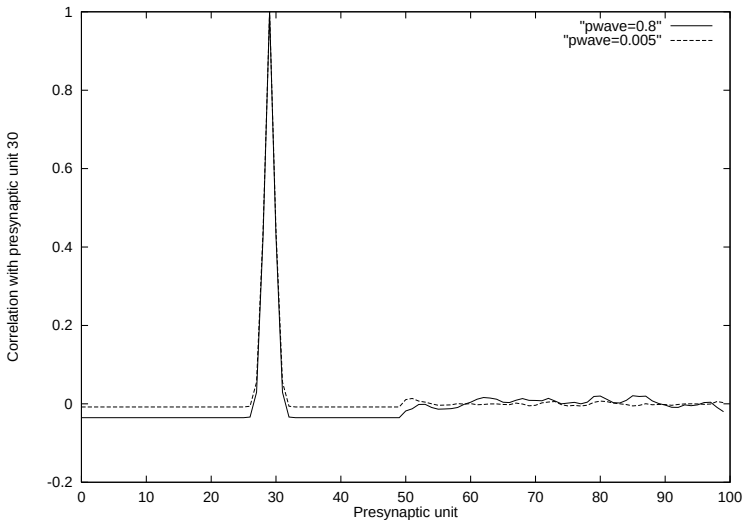


Figure 5.2: Correlation between presynaptic units as a function of the rate of wave generation. The figure shows the correlation between presynaptic unit 30 and all other presynaptic units for both $p_w = 0.8$ (solid line) and $p_w = 0.005$ (dotted line). The correlation, $Corr(i,j)$, between two presynaptic units i and j is defined as $Corr(i,j) = Cov(i,j)/\sqrt{Cov(i,i) \times Cov(j,j)}$ where $Cov(i,j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$ (the covariance between units i and j) and $\langle\ \rangle$ denotes averaging over all inputs. Presynaptic units 1–50 come from the left eye, and units 51-100 come from the right eye.

When both pre- and postsynaptic activity are below threshold, case four of the covariance rule (defined in Table 3.2 on page 38) increases the strength of connections between the two units. For networks with low overall probabilities of retinal waves being present, non-specific increase in all of the weights in the network, combined with the averaging effects of the growth rule, may cause all of the weights in the network to converge to the same value. This possibility is confirmed by the results presented in Table 5.3, which shows the amount of weight change specified by each case of the covariance rule. For low values of p_w, case four of the rule is responsible for most of the weight change in the network. As p_w increases, although the amount of weight change caused by case four stays roughly the same, the amount of weight change specified by the other cases increases considerably, so that case four is no longer dominant.

To test the hypothesis that case four of the covariance rule is responsible for averaging out weight values when p_w is small, the experiments were repeated with a slightly modified covariance rule. This rule was identical to the covariance rule except that when both pre- and postsynaptic activities were below their threshold levels (case four), no weight adaptation was performed. This rule is referred to here as the "active-covariance rule" to indicate that weight adaptation takes place only when either (or both) pre- or postsynaptic activity is above threshold. (This rule is similar to the rule presented by (Montague et al., 1991), although it was independently discovered during this thesis (Eglen, 1996).)

The effect of varying p_w upon development in a network using the active-covariance rule is shown in Figure 5.4. For all values of p_w chosen, network development with the active-covariance rule was either the same or better (by visual comparison) than with the covariance rule. Of particular interest, when both eyes are normally inactive ($p_w = 0.002, 0.003, 0.005$), the network develops

the normal patterns of ocular dominance and topography. A common feature of both rules however is that the average receptive field width increases inversely with p_w (see Table 5.2). These results must be interpreted with some caution however, due to the rather high standard deviation of receptive field widths in comparison to their mean values.

	Mean $\pm$ standard deviation	
p_w	Covariance	Active-cov.
0.002	N/A	11.384 ± 4.718
0.003	N/A	5.793 ± 5.770
0.005	N/A	3.762 ± 4.965
0.020	4.305 ± 4.956	1.591 ± 2.780
0.200	2.199 ± 0.869	0.889 ± 0.046
0.500	2.184 ± 1.001	0.887 ± 0.054
0.800	2.109 ± 0.679	0.896 ± 0.111

Table 5.2: Mean receptive field width for different values of p_w using either the covariance or active-covariance rule. For the entries listed N/A, all postsynaptic units were binocular, which precluded taking any receptive field measurements.

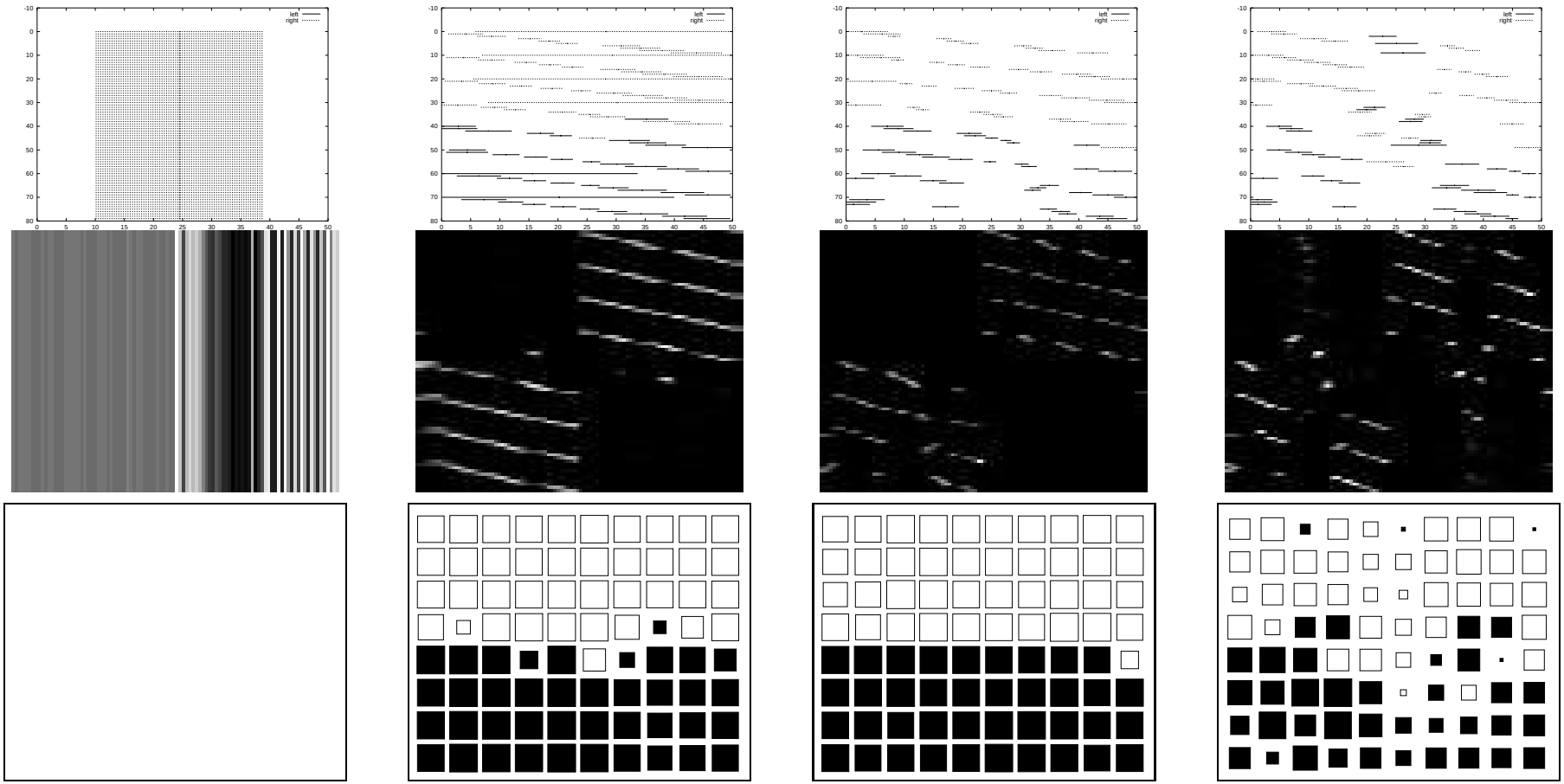
	Frequency of use / %				Weight change per epoch			
p_w	1	2	3	4	1	2	3	4
0.002	0.495	0.011	15.468	84.027	95.5920	-1.1990	-666.186	5830.838
0.003	0.763	0.015	23.032	76.191	164.7320	-1.4550	-1062.795	5279.894
0.005	1.103	0.018	32.381	66.498	257.7480	-1.7380	-1571.903	4600.636
0.020	1.062	1.732	23.810	73.396	1344.4520	-393.7110	-7218.146	4541.466
0.200	1.324	3.427	23.559	71.690	3108.4650	-819.3380	-13704.282	4365.681
0.500	1.459	3.540	25.339	69.661	3339.5660	-852.4340	-14409.363	4250.804
0.800	1.609	3.453	27.526	67.412	3448.7770	-846.1370	-14552.396	4185.087

(a) Covariance rule

	Frequency of use / %				Weight change per epoch			
p_w	1	2	3	4	1	2	3	4
0.002	0.382	0.124	11.815	87.680	112.7830	-24.6050	-806.365	6017.041
0.003	0.409	0.368	12.171	87.052	242.7040	-86.3470	-1652.527	5920.608
0.005	0.403	0.718	11.412	87.467	432.6690	-183.7100	-2872.521	5948.982
0.020	0.671	2.123	14.894	82.312	1513.0623	-562.6000	-8218.576	5538.130
0.200	1.063	3.689	18.945	76.303	3235.3990	-957.5940	-14278.994	4969.223
0.500	1.137	3.863	19.737	75.264	3473.8570	-999.0490	-15003.074	4875.800
0.800	1.249	3.813	21.323	73.615	3577.4920	-987.0899	-15108.254	4773.361

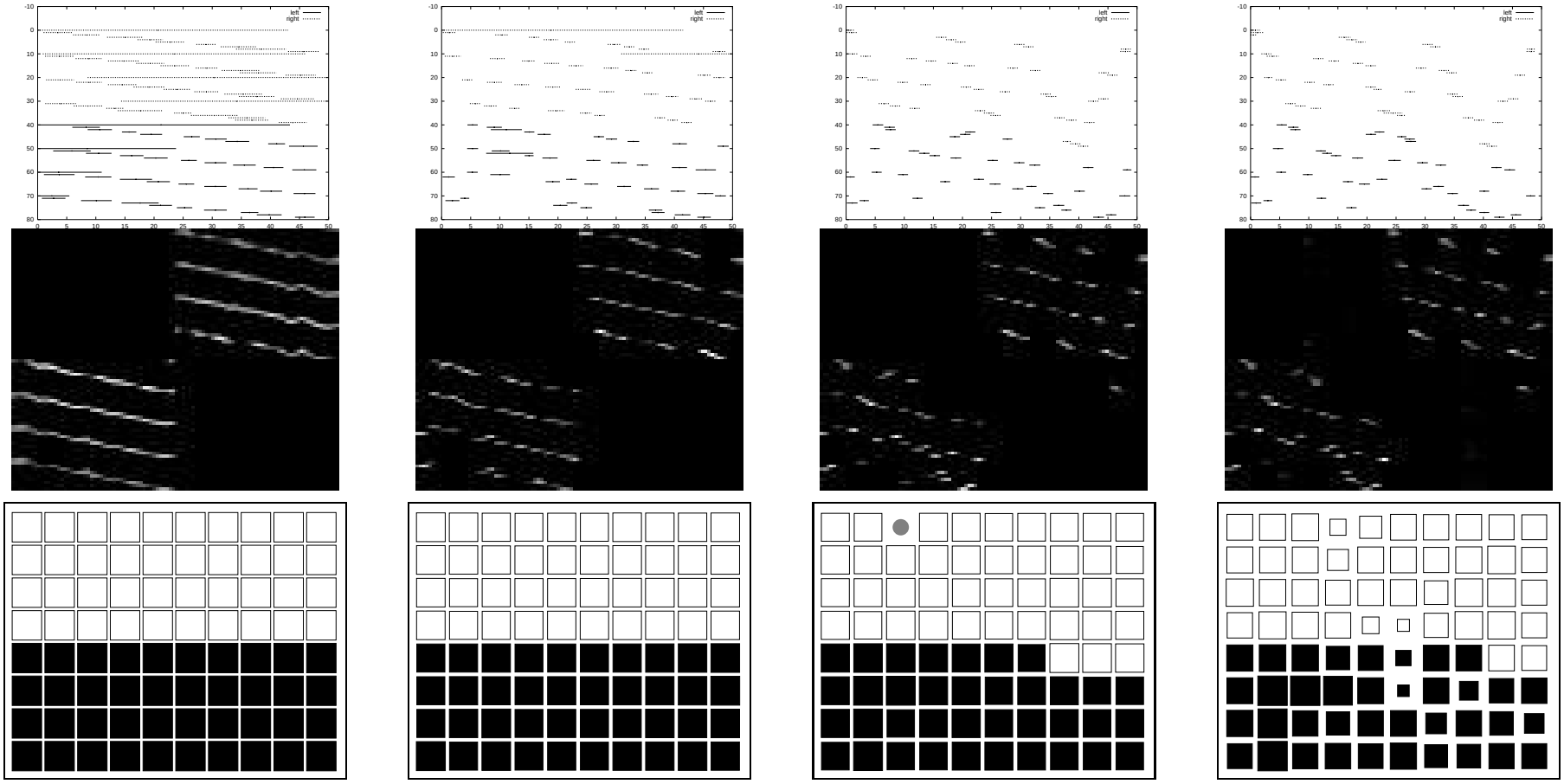
(b) Active-covariance rule

Table 5.3: The frequency of use and the amount of weight change per epoch for each case of the covariance and active-covariance rule. Cases refer to the cases of the covariance rule as shown in Table 3.2. For the active-covariance rule, although values of weight change for case four are given here, these weight changes were ignored.



(a) $p_w = 0.005$ (b) $p_w = 0.02$ (c) $p_w = 0.2$ (d) $p_w = 0.8$

Figure 5.3: Effect of varying p_w upon topography and ocular dominance – 1: development with the covariance rule. For each value of p_w we show (from top to bottom) the final topography plot, weight matrix and ocular dominance plot. When $p_w = 0.005$, all weights adapt to (almost) the same value, producing completely binocular units (hence the empty ocular dominance plot). The plot of the weight matrix for this experiment has exaggerated the small differences between weights, and should not be compared with the other weight matrices.



(a) $p_w = 0.005$ (b) $p_w = 0.02$ (c) $p_w = 0.2$ (d) $p_w = 0.8$

Figure 5.4: Effect of varying p_w upon topography and ocular dominance – 2: development with the active-covariance rule. For each value of p_w we show (from top to bottom) the topography plot, the weight matrix and the ocular dominance plot at the end of development.

5.3 Development of the LGN under conditions of monocular deprivation

The previous section investigated the effect of wave generation rate upon development. In these experiments, the probability of wave generation was the same for both eyes. Although to date there are no measurements comparing the properties of waves in the two eyes, it is reasonable to assume that in normal development the waves in the two eyes have similar spatio-temporal properties. However, experimental manipulations to the visual pathway have reduced the level of activity in one eye, revealing more details about the mechanisms underlying development of the visual system.

A common technique for reducing activity levels from one eye is simply to remove the eye early in development. The effect of enucleation upon the development of the visual cortex is fairly well established (Shatz & Stryker, 1978; Kalil, 1980): in such cases, the non-deprived eye occupies more of the postsynaptic sheet than normal at the expense of the deprived eye. Self-organising models of cortical development can reproduce the data from these experiments due to the competitive nature of the models: the more active presynaptic units tend to connect to more postsynaptic units (von der Malsburg, 1979; Goodhill & Willshaw, 1994).

Monocular deprivation also affects LGN development. Removal of one eye at day E44 (in the cat) allowed the remaining non-deprived eye to innervate the whole of the LGN. The only segregation found within the LGN was a novel lamination pattern consisting of a dorsal magnocellular division and a ventral parvocellular division (Garraghty et al., 1988).

To test whether this model can reproduce the results of monocular deprivation in the LGN, the probability of wave generation was lowered for the left eye in comparison to the right eye. Five different values of p_w (0.005, 0.003, 0.002, 0.001 and 0.000) were chosen for the left eye whilst p_w for the right eye was fixed at 0.020 for all five experiments. Given the low values of p_w and the results from Section 5.2.2, the active-covariance rule was used to modify weights. Initial weights were the same as shown in Figure 4.9, namely the left eye initially innervated the bottom two rows of the LGN and the right eye innervated the bottom four rows.

Figure 5.5 shows the final ocular dominance and topography plots for these experiments. As expected, the lower the value of p_w for the left eye, the fewer postsynaptic units connected to that eye. Some of the units in the bottom half of the LGN that are normally responsive to the left eye instead are innervated by the right eye. The size of layer A increases at the expense of layer A1, as layer A grows downwards into parts of the LGN that usually belong to layer A1. Table 5.4 quantifies the number of postsynaptic units that are responsive either to the left, right or neither eye. (All postsynaptic units developed to become either strongly monocular or dead.) This table shows that as p_w for the deprived eye decreases, the number of postsynaptic units responsive to the non-deprived eye increases.

Given the nature of the active-covariance rule, the basic result that the non-deprived eye occupies more of the LGN was expected. For units in the deprived eye, most of the time the active-covariance rule is specifying weight decrements (especially with case three of the rule), since the presynaptic activity is low but the postsynaptic activity is high due to the input activity from the non-deprived eye. However, there is one slight complication in the model as it stands. For the extreme case when the left eye never produces any waves ($p_w = 0.000$), there are still six units responsive to the left eye. This occurs simply because the presynaptic normalisation forces each

presynaptic unit to keep connections to the postsynaptic sheet, regardless of the activity level of the presynaptic units.

5.3.1 A weaker form of divisive normalisation

For the monocular deprivation simulations, divisive normalisation was used to fix the sum of weights from each presynaptic unit. Considering the divisive normalisation from Section 3.6.1 in more detail, there are two cases to consider when the weights are to be modified so that the sum of weights equals some target value:

- Multiplicative normalisation: when the sum is *less* than the target value, each weight is *multiplied* by some constant greater than one.
- Divisive normalisation: when the sum is *greater* than the target value, each weight is *divided* by some constant greater than one.

Both of these cases are called divisive normalisation, although strictly speaking only the second case divides each weight. Divisive normalisation therefore reduces the amount of weight strength back to some constant, whereas multiplicative normalisation increases the amount of weight strength. Arguments for normalisation methods in computational models are usually based on providing some form of resource limitation to prevent weights from increasing without bounds. However, by using both divisive and multiplicative normalisation, the constraint is stronger: not only does it prevent weights increasing without bounds, but it also prevents weights decreasing without bounds. This constraint is too strong, since most models also employ other techniques for ensuring weights do not decrease without bounds (such as ensuring each weight is non-negative).

A weaker form of normalisation was therefore used in the model by ignoring the multiplicative element of presynaptic normalisation. This effectively changes the normalisation for presynaptic unit i from that given in Section 4.2.1 to:

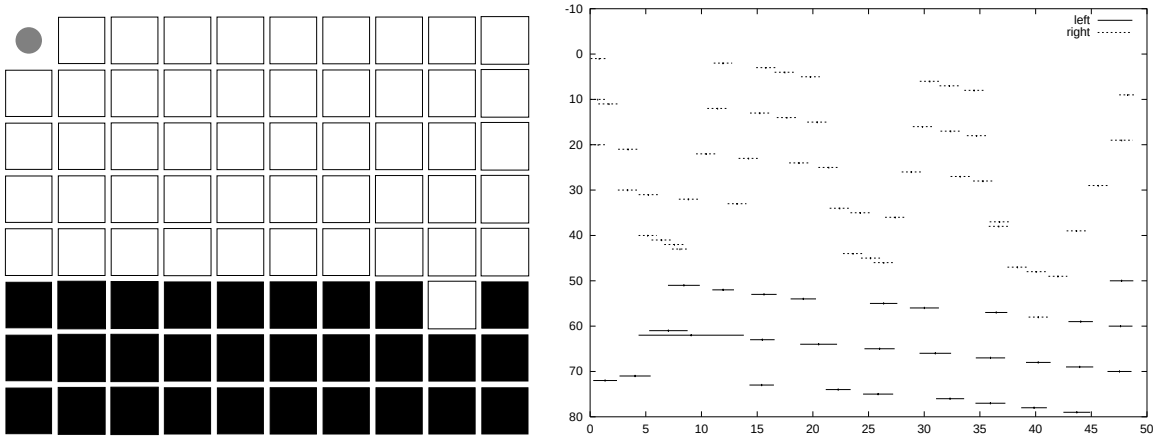
$$\sum_{j=1}^{N_{\text{post}}} w_{ij} \leq T_{\text{pre}}$$

The same set of monocular deprivation experiments were repeated using this new form of divisive normalisation. The results are shown in Figure 5.6, with counts of the number of postsynaptic units responding to each eye shown in the right-hand side of Table 5.4. The overall results are broadly similar to the experiments with both divisive and multiplicative normalisation: the less active the eye, the fewer postsynaptic units receive input solely from that eye. However, for very small values of p_w for the left eye (such as 0.000 or 0.002), no postsynaptic units are responsive to the left eye: all units are either responsive to the non-deprived right eye or are dead. The number of dead postsynaptic units also increased when multiplicative normalisation was ignored. The weaker form of divisive normalisation therefore produces better results than the divisive and multiplicative normalisation for replicating the outcome of monocular deprivation.

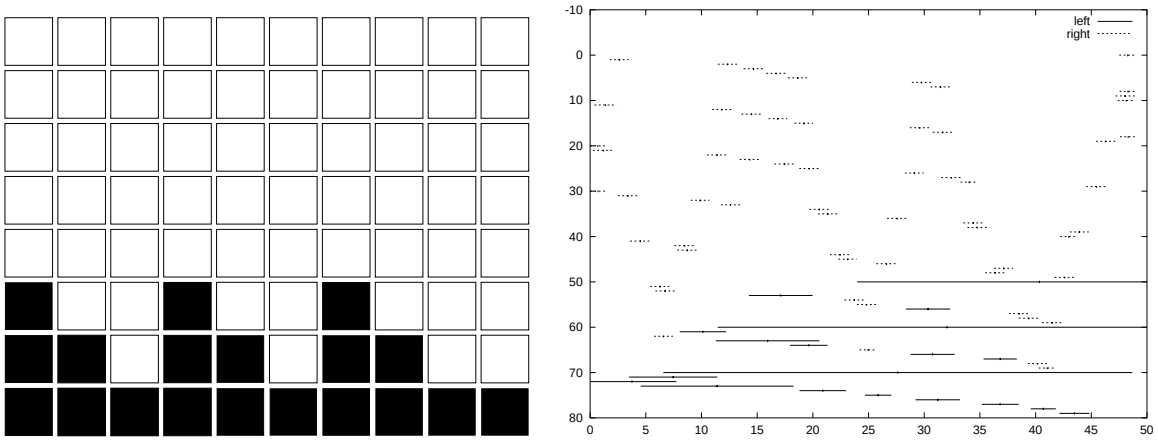
5.3.2 The nature of the retinotopic mapping

Regardless of the form of the divisive normalisation, the nature of the retinotopic mapping for the non-deprived eye is different to the normal mapping. As shown in the topographic plots of

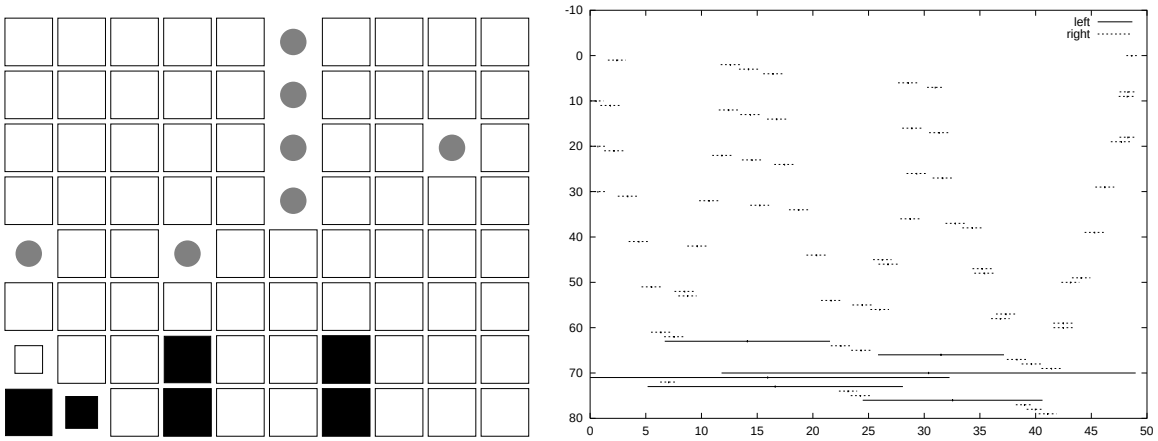
Multiplicative and divisive normalisation



(a) $p_w = 0.005$

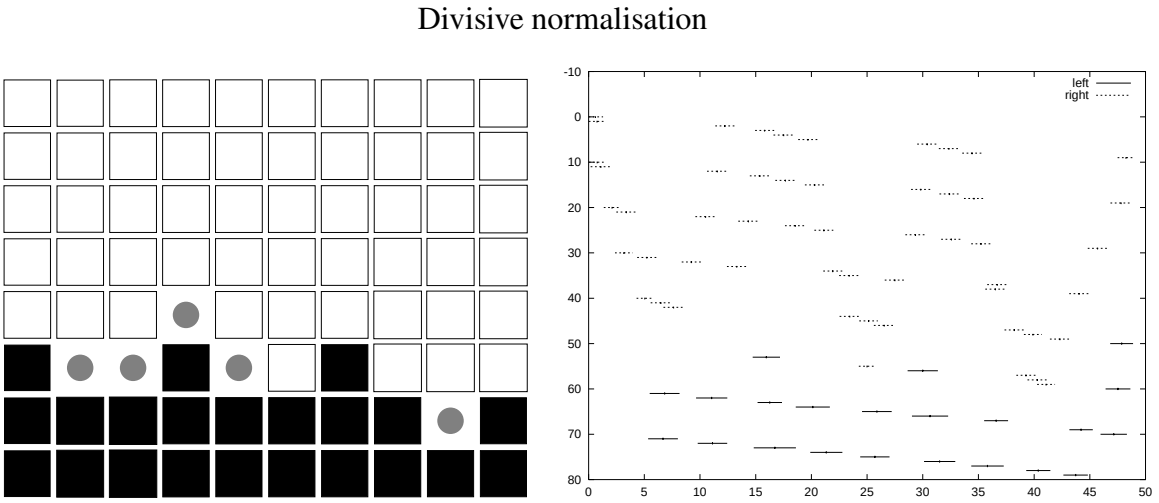


(b) $p_w = 0.002$

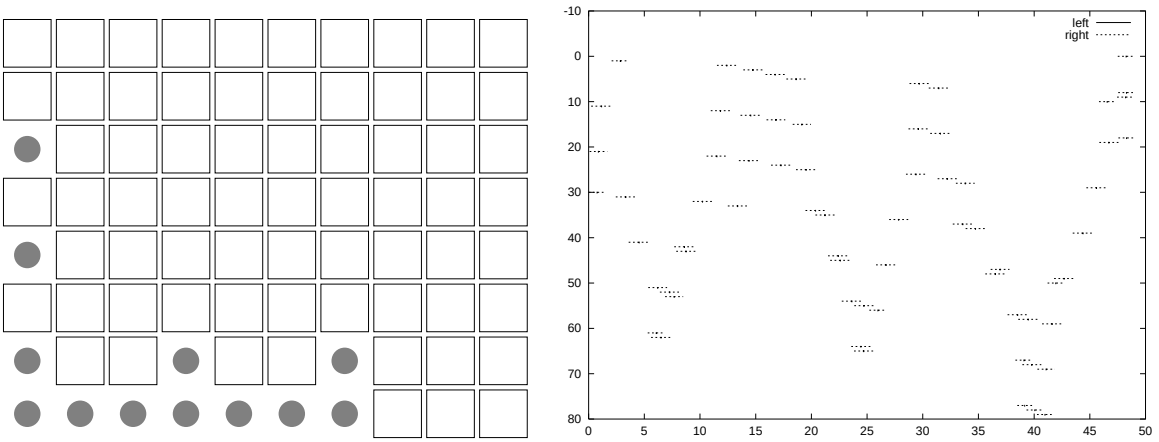


(c) $p_w = 0.000$

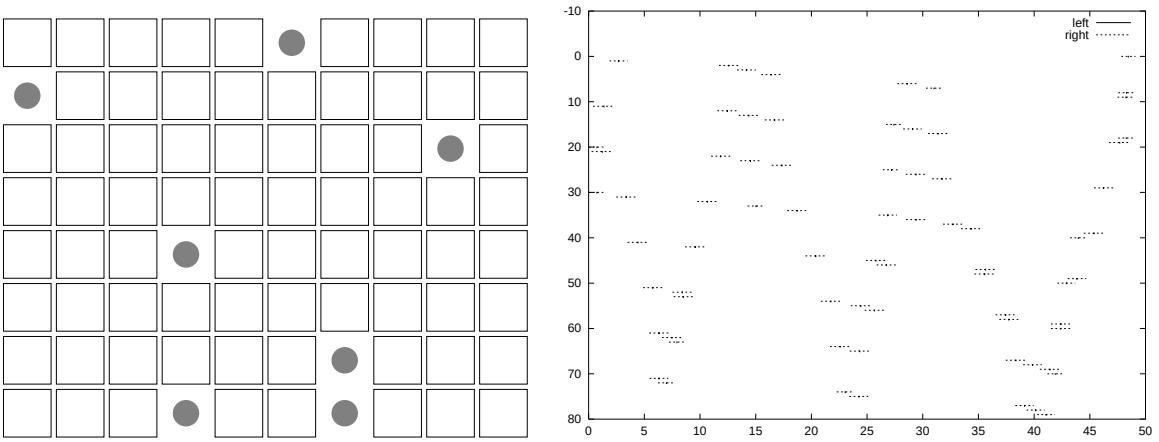
Figure 5.5: Network development with reduced activity levels for the left eye – 1: development with multiplicative and divisive presynaptic normalisation.



(a) $p_w = 0.005$



(b) $p_w = 0.002$



(c) $p_w = 0.000$

Figure 5.6: Network development with reduced activity levels for the left eye – 2: development with only divisive presynaptic normalisation, ignoring multiplicative normalisation.

	(a) D and M			(b) D only		
Left p_w	Left	Right	Dead	Left	Right	Dead
0.005	29	50	1	22	53	5
0.003	22	58	0	7	64	9
0.002	19	61	0	0	68	12
0.001	15	62	3	0	71	9
0.000	6	67	7	0	73	7

Table 5.4: Number of postsynaptic units responsive to the left, right or neither eye at the end of development under reduced activity conditions for the left eye. The value of p_w for the right eye was fixed at 0.020, and the value of p_w for the left eye varied as given in the table. (a): Results using both divisive and multiplicative presynaptic normalisation. (b): Results using only divisive presynaptic normalisation.

Figures 5.5 and 5.6 (especially when $p_w = 0.000$), each row of LGN units no longer completely covers the visual space — gaps are present within each row. The projection columns for these maps show that units in different rows respond to neighbouring parts, rather than the same parts, of the presynaptic sheet (Figure 5.7). This mapping is produced because the non-deprived eye innervates more rows of the LGN, allowing more freedom for neighbouring vertical units to represent neighbouring, rather than identical, parts of visual space.

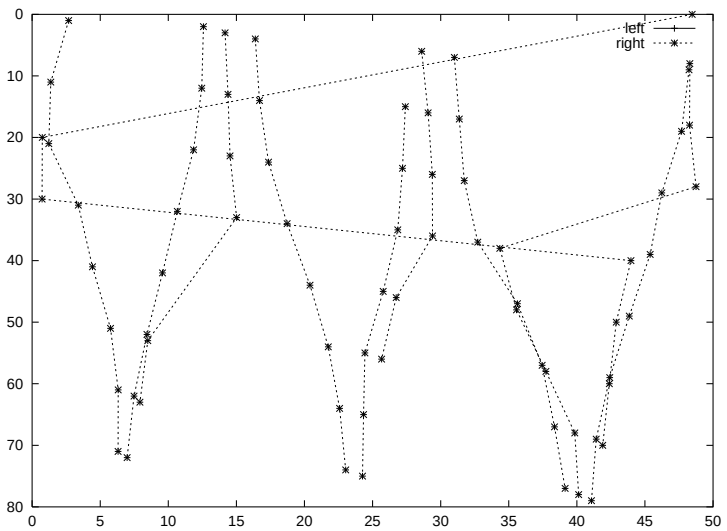


Figure 5.7: Final projection columns for a network with no waves generated within the left eye. Projection plots taken from network shown in Figure 5.6(c). One of the projection columns jumps from the far right of the visual field to the far left and then back to the right due to the periodic conditions in the LGN.

5.4 The effect of wave width upon development

Having considered the temporal properties of the waves, we now turn our attention to the spatial properties of retinal waves. The width of the wave controls the degree to which neighbouring presynaptic units within an eye are correlated. To investigate the effect of wave width σ_w upon the development of topography and ocularity, the LGN was reduced to a one-dimensional array of twenty postsynaptic units. This reduction in the postsynaptic sheet from a two-dimensional to a one-dimensional array allowed us to ignore the issue of projection columns and focus on the topography and ocularity of individual postsynaptic units within just one row of the LGN. It also shows that the set of weight modification rules can produce topographic and ocular refinement within networks of different sizes to the usual one-dimensional retina and two-dimensional geniculate.

For this network, apart from the changes in size of the postsynaptic sheet ($X_{\text{post}} = 20, Y_{\text{post}} = 1, N_{\text{post}} = 20$), all of the other parameters were the same as listed in Table 4.1. Initial weights were created with a topographic bias of $b = 20$. Figure 5.9 shows the final maps created using a range of values for σ_w. For values of σ_w smaller than 2.5, units have small receptive fields and are mostly monocular. As the wave width increases, so does the size of the receptive field. This is because the extent of the within-eye correlations widens with σ_w, as shown in a plot of the presynaptic correlations in Figure 5.8.

The dependence of wave width upon σ_w is summarised in Figure 5.10. Increasing σ_w also affects the overall topographic arrangement: when σ_w is small, the neighbouring postsynaptic units are organised into a smooth topographic mapping. As the wave width increases, although neighbouring units still tend to have similar receptive fields, the global topographic mapping is lost.

As shown by the ocularity plots in Figure 5.9, the wave width also affects the final ocularity of each postsynaptic unit. For small wave widths, most postsynaptic units become monocular. For wider waves, units tend to remain binocular (although the results when $\sigma_w = 4$ in Figure 5.9 are an exception). Although the waves are still uncorrelated in their activity patterns, the wider the waves, the more chance there is that waves will be present in both eyes at the same time. Given that the covariance rule will tend to connect active presynaptic units to the same postsynaptic unit, postsynaptic units will tend to receive inputs from both eyes.

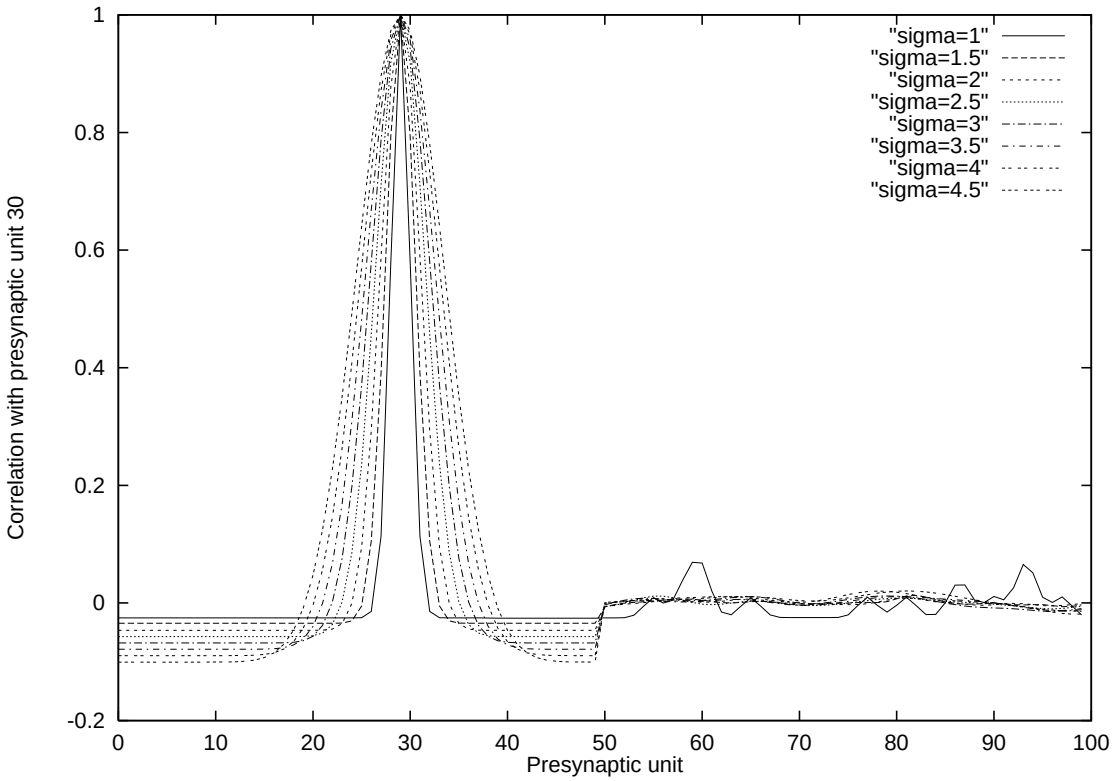


Figure 5.8: Correlations between presynaptic units as a function of wave width. The figure shows the correlation between presynaptic unit 30 and all of the other presynaptic units for various values of σ_w. As the wave width increases, the correlation function between neighbouring presynaptic units of the same eye gets broader, and distant presynaptic units get more anticorrelated. In contrast, the between-eye correlations remain uncorrelated regardless of the value of p_w. (The correlation between units is defined in Figure 5.2.)

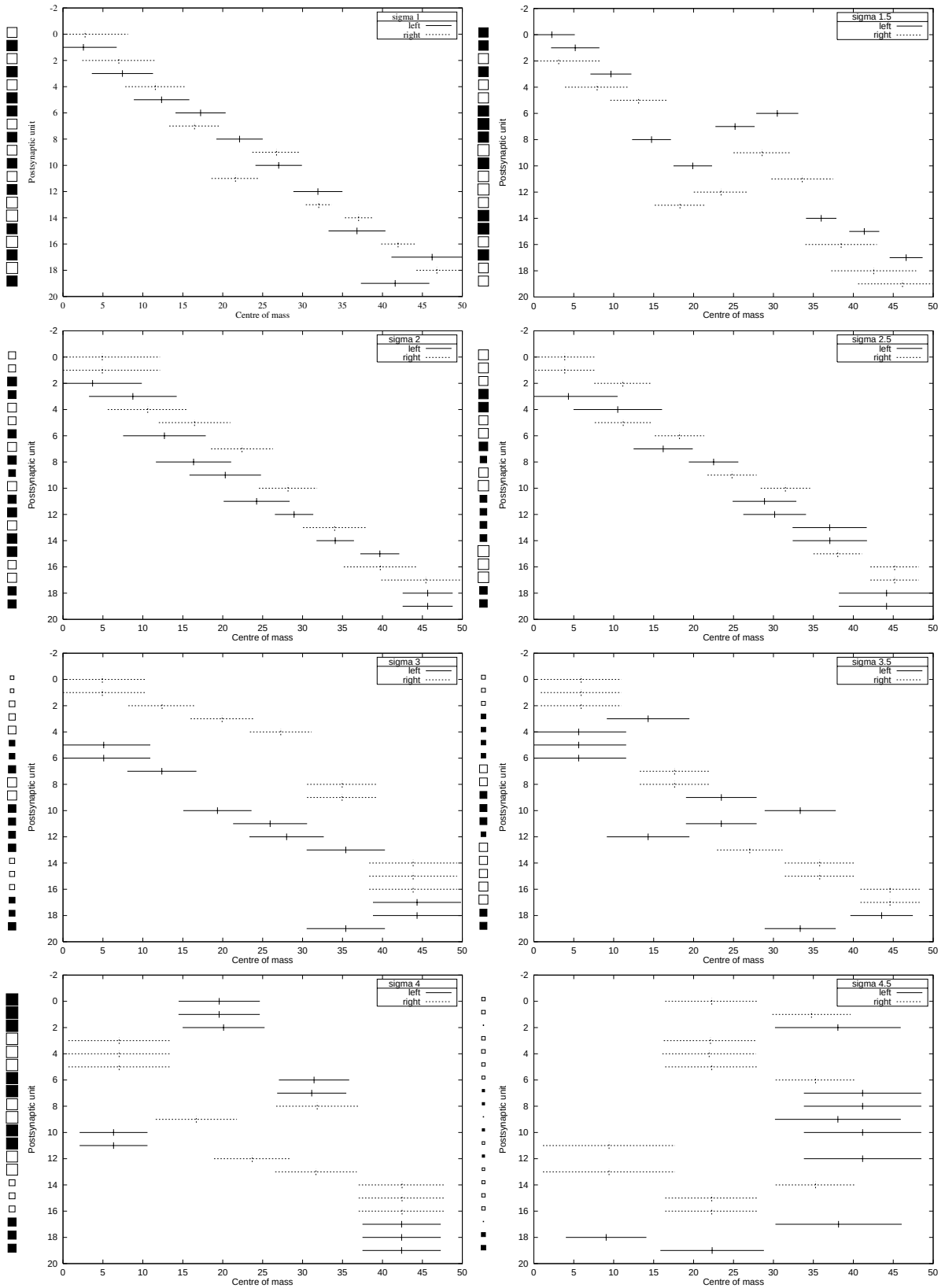


Figure 5.9: The effect of wave width upon topography and ocular dominance. The two retinae innervate a one-dimensional array of twenty postsynaptic units. Eight experiments were run using the same parameters except for σ_w which increased from 1.0 to 4.5 in steps of 0.5. The value of σ_w is given in the key of each plot. The ocularity plot for each network is given to the left of the topographic plot, and aligned with the postsynaptic unit in the topographic plot.

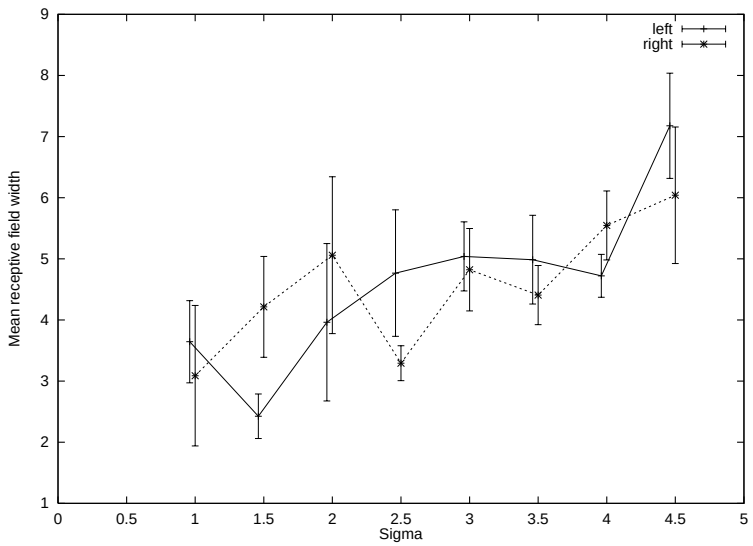


Figure 5.10: The effect of wave width upon receptive field width. Plots for the left eye are offset slightly to the left from the true value of sigma so that data points for both the left and right eyes can be clearly seen. Error bars indicate ± 1.0 standard deviation from the mean value.

5.5 The effect of polarity-specific wave patterns upon LGN sublamination

In the mature LGN, each relay cell receives input from only on- or off-centre retinal cells, but not both. Relay cells are therefore classified as either 'on-centre' or 'off-centre' according to the dominant polarity (on- or off-centre) of retinal inputs. Like-centred LGN cells also tend to cluster into sublamina, as described in Section 2.3.5. It has been suggested that development of these sublamina is driven by activity-dependent processes, although until recently no polarity-specific patterns amongst retinal waves had been discovered which could drive such segregation (Morgan & Thompson, 1993). The first signs of polarity-specific firing patterns in the waves were recently reported by Wong and Oakley (1996). At P9 in the ferret retina, both on- and off-centre cells participate in retinal waves, and cannot be distinguished by their activity patterns. By P16 however, an asynchrony in the wave patterns of on- and off-centre cells has developed: on-centre cells tend to fire less often than off-centre cells. This means that when on-centre cells are active, neighbouring off-centre cells are also active, but when off-centre cells are active, neighbouring on-centre cells are usually quiet (Wong & Oakley, 1996).

5.5.1 Representation of on- and off-centre retinal units

To investigate whether these polarity-specific patterns drive LGN sublamination, the model retina was modified to represent both on- and off-centre units. In the mature retina, the number of on- and off-centre retinal ganglion cells is roughly equal, although there are slightly more off cells (Wassle, Boycott, & Illing, 1981a; Wassle, Peichl, & Boycott, 1981b). The relative percentages of these cells in the developing retina though are still unknown because on- and off-centre units cannot be morphologically discriminated until the period of segregation in the inner plexiform layer (P9 onwards; see Table 2.3). Counts of on- and off-centre units in ferret retina during P9–21 suggest a roughly equal number of on- and off-centre units during the period of on- and off-centre

sublamination (Wong & Oakley, 1996). On- and off-centre cells independently tile the ganglion cell layer, with the nearest neighbouring unit nearly always a unit of the opposite sign (Wassle et al., 1981a, 1981b). For the model retina therefore, presynaptic units were alternately labelled on-centre and off-centre (starting with the leftmost unit as on-centre). This arrangement ensured an equal number of on- and off-centre units arranged such that the nearest neighbouring unit was always of the opposite polarity.

To mimic the polarity-specific firing of retinal cells, retinal inputs were generated using the same procedure as listed in Section 4.3 and then slightly modified. At each time step, after a wave has been positioned on a retina, a value p is chosen at random from the uniform distribution [0,1]. This value controls whether the activity of on- or off-centre units should be set to zero according to the rules in Table 5.5.

if $p \leq p_{\text{on}}$	set $x_i = 0 \; \forall i = 2, 4, \ldots, X_{\text{pre}}$	(keep on-centre activity)
else if $p \leq p_{\text{off}}$	set $x_i = 0 \; \forall i = 1, 3, \ldots, X_{\text{pre}} - 1$	(keep off-centre activity)
else	change nothing	(keep both on- and off-centre activity)

Table 5.5: Rules for generating polarity-specific waves.

This formulation allows us to vary the correlations between on- and off-centre units by varying the values of p_{on} and p_{off}. To mimic the late onset of polarity-specific waves, we only introduce this extra mechanism after a fixed number of epochs have elapsed (1200 epochs for the experiments reported here). We also allow the width of the waves to increase during the simulation so that neighbouring on- or off-centre units are correlated. This increase in wave width during development is consistent with wave recordings from the ferret retina (Wong et al., 1993, see also Section 5.6). (Narrow waves are not wide enough to produce sufficiently high correlations between neighbouring on- or off-centre units.) Table 5.6 lists the parameters used in the following simulations for generating polarity-specific waves along with details of the other simulation parameters.

Epoch	Action
0	Wave width set to $\sigma_w = 0.85$
400	Wave width increased to $\sigma_w = 2.00$
1200	Polarity-specific wave patterns introduced

Table 5.6: Parameters for the polarity-specific wave experiments. To allow the network more time to develop, t_r was set to 700 epochs. Values for p_{on} and p_{off} are given in the text. All remaining simulation parameters were the same as in Table 4.1, with initial weights the same as in Figure 4.10.

5.5.2 Measuring the degree of polarity dominance

The segregation of on- and off-centre inputs into postsynaptic units is assessed by taking a measurement of the fraction of input that comes from on-centre units. This measure is analogous to the measure of ocular dominance, and so will be called 'polarity dominance'. For a postsynaptic

unit j, the polarity dominance is denoted f_j, and is defined as:

$$f_j = \frac{f_j^{\text{on}}}{f_j^{\text{on}} + f_j^{\text{off}}} - \frac{1}{2}$$

$$f_j^{\text{on}} = \sum_{i=1}^{N_{\text{pre}}/2} w_{(2i-1)j} \qquad f_j^{\text{off}} = \sum_{i=1}^{N_{\text{pre}}/2} w_{(2i)j}$$

The sign of f_j shows whether the dominant polarity of input unit is on-centre $(+)$ or off-centre $(-)$, regardless of the eye of origin. The magnitude of this value indicates the degree to which the unit receives input from presynaptic units of the dominant polarity. For each postsynaptic unit, the value of f_j varies between $[-0.5 : +0.5]$. A value of $+0.5$ indicates that the unit receives all of its input from on-centred presynaptic units, and -0.5 indicates all the input comes from off-centred units. A value of 0.0 indicates that the unit receives equal input from on- and off-centre presynaptic units. To monitor the overall development of polarity dominance, we also take the mean and standard deviation of f_j, separately for all positive and negative values of f_j, at regular intervals during the simulation.

The value of f_j for each postsynaptic unit is visualised in the ocular dominance plot by drawing either a $+$ or $-$ sign centred over the top of the square indicating the ocularity status of the postsynaptic unit. The size of the $+$ or $-$ sign is scaled in proportion to the magnitude of f_j such that the maximum possible sign corresponds to a magnitude of 0.5.

5.5.3 Anticorrelated polarity-specific waves

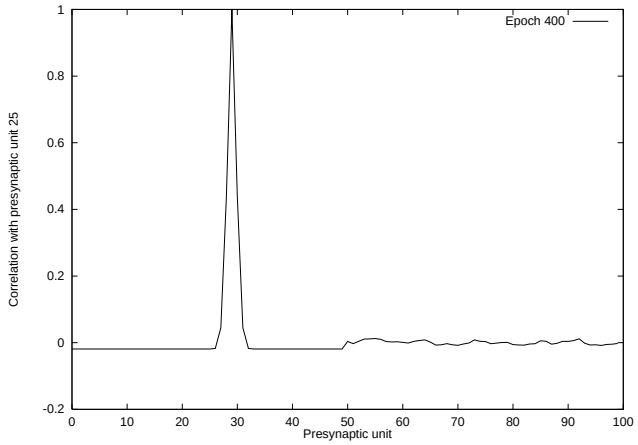
The first polarity segregation experiment was performed with $p_{\text{on}} = 0.5$ and $p_{\text{off}} = 1.0$. Choosing these values ensured that at each time step after polarity-specific waves were introduced, only on- or off-centre units (but not both at the same time) would be active, ensuring that units of opposite centre signs were anticorrelated.

The correlation matrices during the three key stages of wave development are shown in Figure 5.11. First, up until 400 epochs the presynaptic correlations are the same as seen in previous experiments. Second, after 400 epochs the waves are widened, which broadens the correlations amongst more distant presynaptic units. Third, after 1200 epochs the polarity-specific waves ensure that neighbouring on-centre units (separated by an off-centre unit) are more correlated than the nearest neighbouring off-centre units (and the same for neighbouring off-centre units).

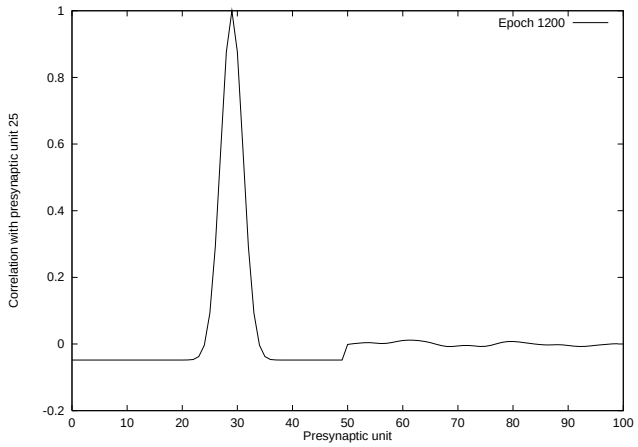
Figure 5.12 shows the development of topography, ocularity and polarity segregation at different stages during development. Up until 1200 epochs, the network develops in a typical manner, although the ipsilateral eye occupies slightly more of layer A than normal due to the higher value of σ_w used after 400 epochs. At this point, most postsynaptic units still receive an equal mixture of on- and off-centre inputs, as shown in the ocular dominance plots and the mean values of f_j in Figure 5.13. After 1200 epochs, segregation of on- and off-centre inputs gradually begins. The segregation is boosted once the growth rule finishes after 1400 epochs. By around 2500 epochs, most postsynaptic units receive input from presynaptic units of only one polarity. However, although individual postsynaptic units are responsive to either on- or off-centre units, there is no sublamination present: neighbouring postsynaptic units within the same row of the LGN do not always have the same polarity. Instead, each row of the LGN has a mixture of on- and off-centre

postsynaptic units.

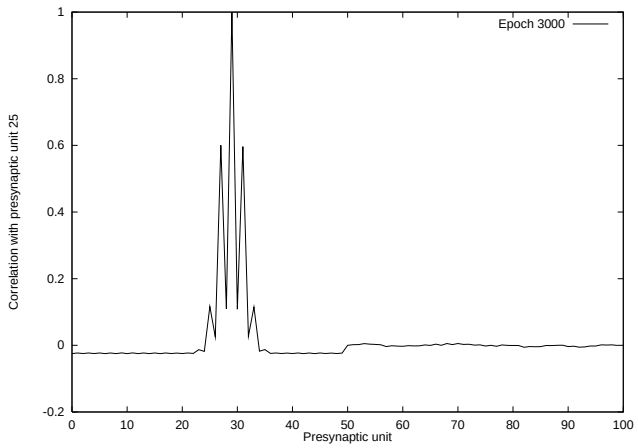
The introduction of polarity-specific waves did not radically affect the nature of the topographic mapping, as shown by the topographic plots in Figure 5.12. The final projection columns for the network, shown in Figure 5.14, are well organised, with only a few units in the wrong position. The average wave width at the end of development was slightly bigger than for the previous simulations (mean = 3.795, s.d. = 0.695, compared with mean = 2.496, s.d. = 0.737 for L4R8 of Figure 4.11 and mean = 2.922, s.d. = 0.894 for L2R4 of Figure 4.11). This increase in receptive field width was expected due to the increase in the width of the retinal waves after 400 epochs.



(a) Epoch 400

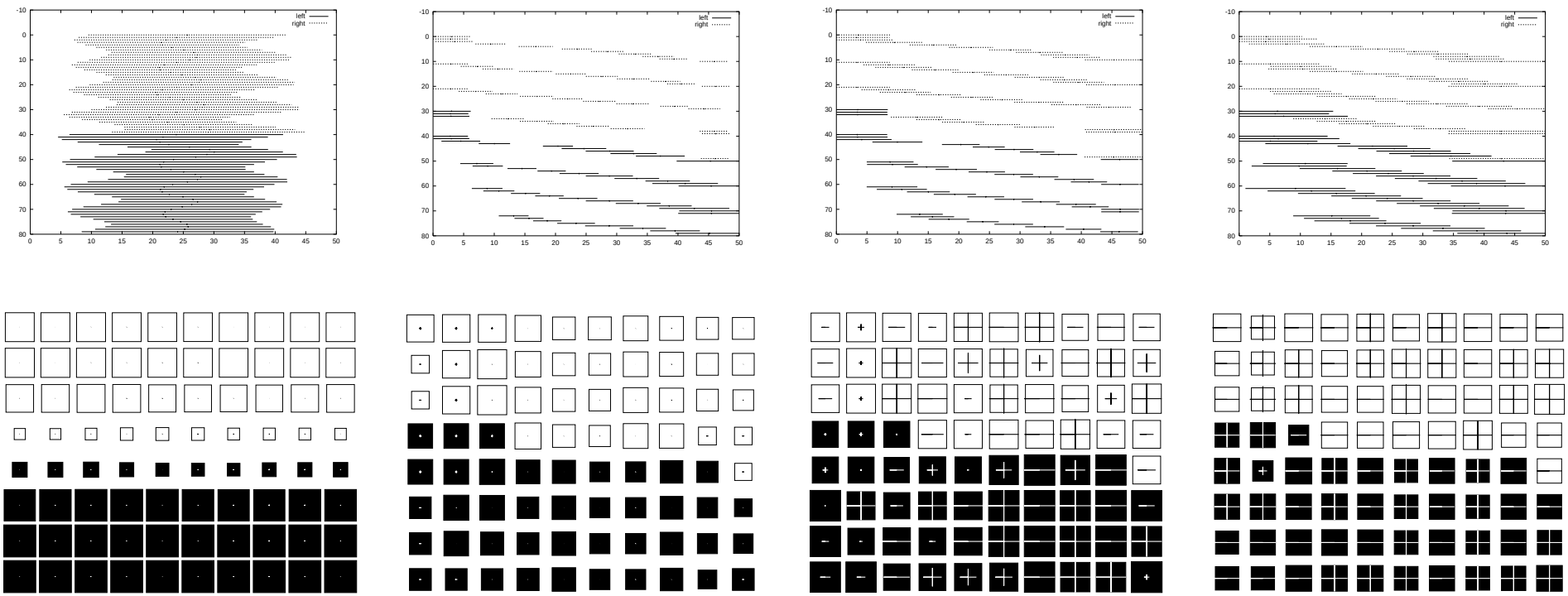


(b) Epoch 1200



(c) Epoch 3000

Figure 5.11: Slices of the input correlation matrix before and after the onset of polarity-specific wave patterns. Each correlation plot shows the correlation between presynaptic unit 25 (an on-centre unit) and all other presynaptic units. Conventions for correlations are the same as in Figure 5.2. ($p_{\text{on}} = 0.5, p_{\text{off}} = 1.0$.) (a): Initial correlations from the start of the simulation up to 400 epochs, just before waves are widened. (b): Correlations from epoch 1 up to epoch 1200, just before polarity-specific waves are introduced. (c): Correlations from epoch 1 to the end of the simulation at 3000 epochs.



(a) Epoch 100 (b) Epoch 1000 (c) Epoch 2000 (d) Epoch 3000

Figure 5.12: Time course of polarity dominance using anticorrelated on- and off-centre waves. The topography plot (above) and ocular dominance plot (below) is shown at four different points during development. (p_{on} =0.5,p_{off} =1.0).

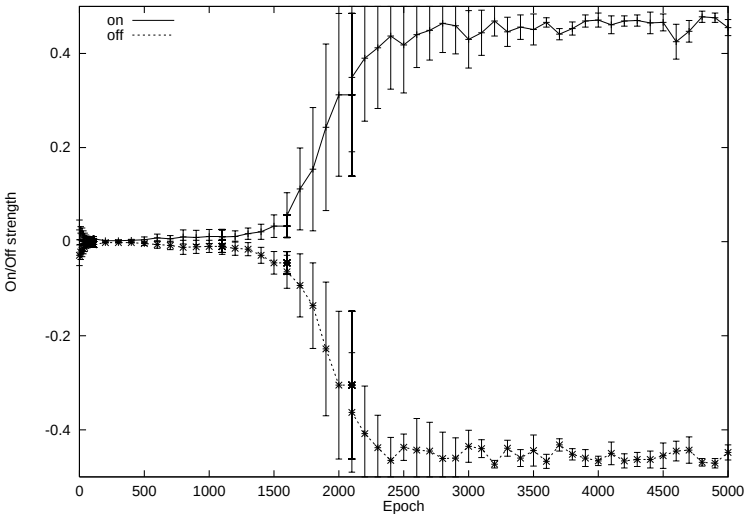


Figure 5.13: Development of polarity dominance in postsynaptic units. The mean value of f_j over all postsynaptic units is shown at regular intervals during development. Error bars indicate ± 1.0 standard deviation from the mean value. This experiment was continued after 3000 epochs for another 2000 epochs to ensure that the network had converged. It is clear that by around 3000 epochs, polarity segregation is almost complete.

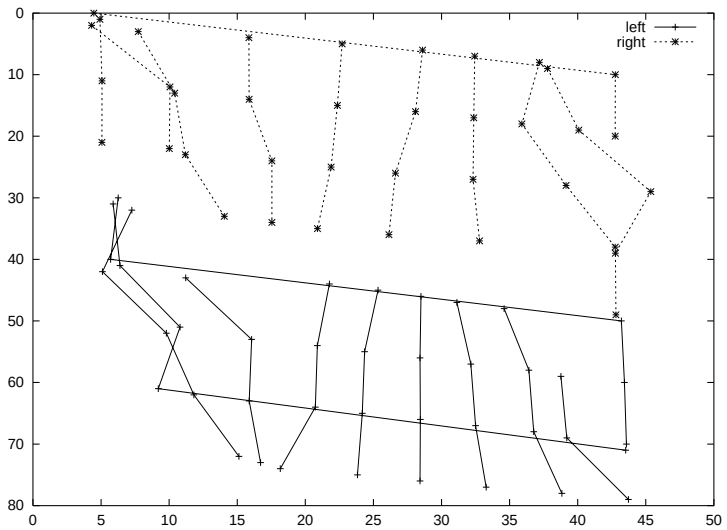


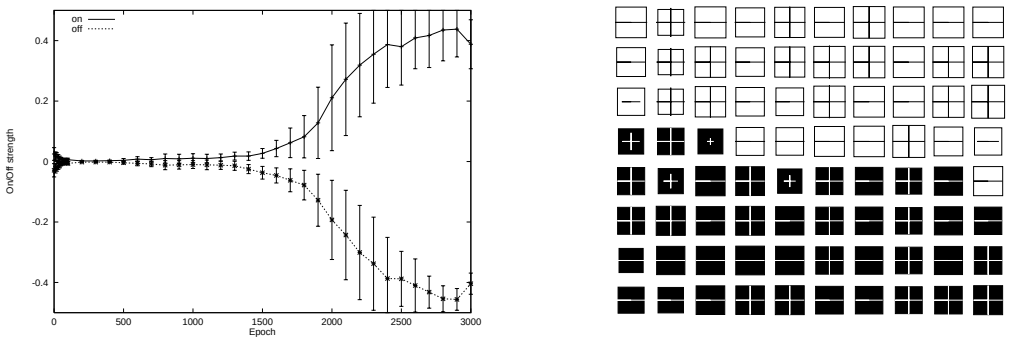
Figure 5.14: Final projection columns for the network with polarity-specific waves as input. The corresponding topography and ocular dominance plots for this experiment are shown in Figure 5.12.

5.5.4 Weakening the anticorrelations between on- and off-centre inputs

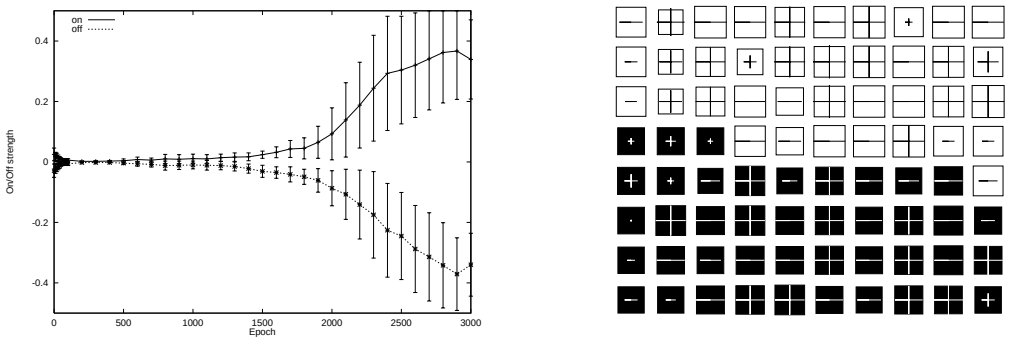
The parameters p_{on} and p_{off} control the nature of the correlations between the on- and off-centre units. In the previous section, the values $p_{\text{on}} = 0.5, p_{\text{off}} = 1.0$ were used. This ensured that once the polarity-specific waves had been introduced, an on-centre unit and off-centre unit would never be jointly active, maximising the chance that units of different polarity would not jointly innervate the same postsynaptic unit. However, such anticorrelations have not been found in the developing retina. Instead, as mentioned before, when on-centre cells are active, neighbouring off-centre cells also tend to be active, but not vice-versa (Wong & Oakley, 1996). These weaker anticorrelations can be modelled by reducing the value of p_{off} below 1.0 to allow some retinal waves to have both on- and off-centre activity (using the default rule from Table 5.5).

Since off-centre cells are active more often than on-centre cells during the period of polarity-specific waves, a set of experiments were performed with p_{on} reduced below 0.5, and p_{off} set to $p_{\text{on}} + 0.5$. This means that off-centre waves are present half of the time, with the rest of the time divided between on-centre waves and mixed on- and off-centre waves. As p_{on} is reduced to 0.0, only off-centre and mixed on- and off-centre waves are generated. This is the situation found in the developing retina (Wong & Oakley, 1996).

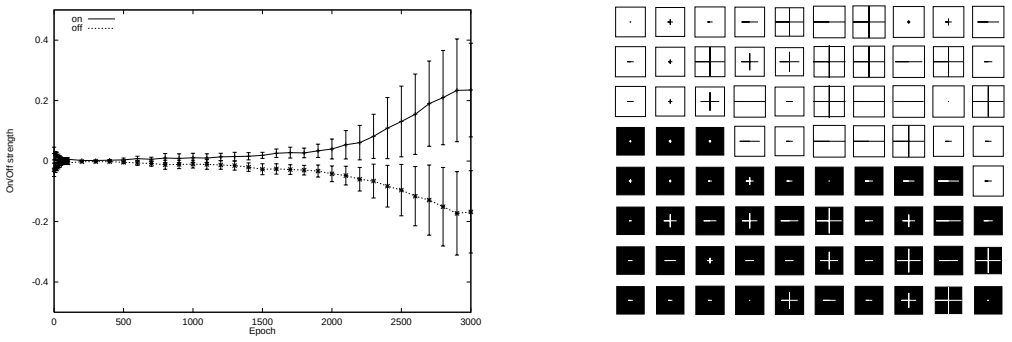
Figure 5.15 shows the results of development using $p_{\text{on}} = 0.48, 0.46, 0.44, 0.42$ with $p_{\text{off}} = p_{\text{on}} + 0.5$ in each case. The plots of polarity segregation for each experiment clearly show that as p_{on} decreases, the overall degree of polarity segregation is greatly reduced. The other network features, ocular dominance and topography, were unaffected. The existence of mixed on- and off-centre waves therefore specifically inhibits polarity segregation in this model.



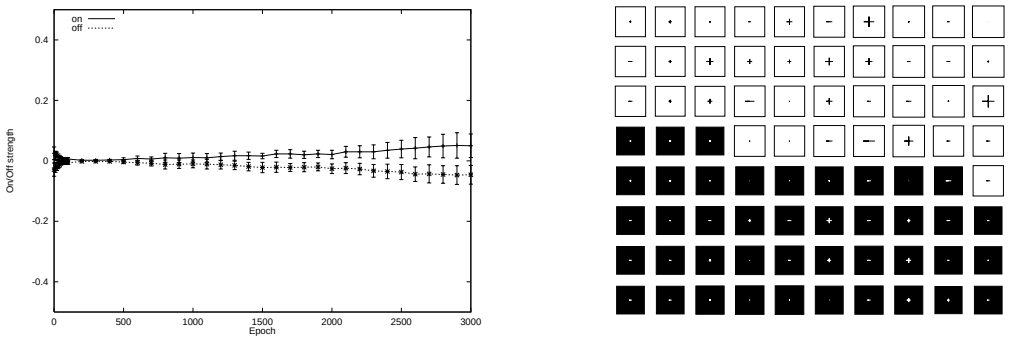
(a) $p_{\text{on}} = 0.48; p_{\text{off}} = 0.98$



(b) $p_{\text{on}} = 0.46; p_{\text{off}} = 0.96$



(c) $p_{\text{on}} = 0.44; p_{\text{off}} = 0.94$



(d) $p_{\text{on}} = 0.42; p_{\text{off}} = 0.92$

Figure 5.15: Polarity segregation for different correlations between on- and off-centre presynaptic units. For each value of p_{on} and p_{off} the mean value of f_j throughout development and the final ocular dominance plot are shown.

5.6 Comparison with biological data

To date, little quantitative data has been published on the spatio-temporal properties of the retinal waves. Instead, more attention has been paid to examining the mechanisms underlying the generation and propagation of the waves (Wong et al., 1995; Feller, Wellis, Stellwagen, Weblin, & Shatz, 1996). Despite the relative paucity of data, in this section we compare the limited biological data of the wave properties with those used in the model.

5.6.1 Temporal characteristics of retinal waves

Wong et al. (1993) provide average values for the burst duration and mean inter-burst interval for ferret retina. These values allow a broad comparison to be made between the temporal properties of the waves used in the simulations and the waves found in the natural retina. Table 5.7 summarises the results of this comparison.

Even when we take the maximum values for the overall probability of eye activity ($O_{\max}$), we see that the values of O are quite small. Table 5.1 shows how O varies with the parameter p_w from the computer simulations. From this, we can see that even when the waves are present most often on the retina at P21, this corresponds to a value for p_w much less than 0.020. It is important to note however that these calculations are only order of magnitude figures: for example, we do not compare the relative sizes of the retina and the speed at which the wave travels across the retina.

If the estimates are roughly correct, it would indicate that the activity-driven developmental mechanisms need to be robust enough to account for the long periods of silence between waves. We have shown here that it is possible for the covariance rule to be adapted to prevent weight adaptation when input and output are below threshold, to account for the long periods of silence between waves. Alternative mechanisms may also allow the network to cope with long periods of silence.

5.6.2 Spatial characteristics of retinal waves

The spatial characteristics of the waves have been analysed in some detail by examining the correlation index between pairs of ferret retinal cells (Wong et al., 1993). These correlation indexes reveal two features about the waves:

1. The correlation between pairs of retinal cells decreases with the distance between cells. This fits in with the model assumption that the retinal waves can be modelled as Gaussian functions of position from the centre of the wave. Providing an estimate of the width of Gaussians from the biological data is complicated, however, due to the arbitrary correspondence between the size of the natural and model retina.

2. During the first few postnatal weeks the waves tend to get wider, producing weaker correlations over a wider distance. This continues until around P30 when the broad waves eventually become constant maintained levels of activity, similar to that seen in the mature retina.

During development, the waves therefore get progressively wider until they eventually change into maintained background levels of activity. More details about the spatial width of retinal

Day	A	$I \pm sd$	$O_{\min}$	$O_{\max}$
P0	1–4	75 ± 6	0.012	0.055
P5	2–4	101 ± 20	0.016	0.047
P12	2–7	71 ± 20	0.022	0.121
P15	2–7	60 ± 9	0.028	0.121
P21	2–9	28 ± 3	0.061	0.265

The overall probability of eye activity, O, is given using:

$$O = \frac{A}{A+I}$$

A = Mean wave duration in seconds (active)

I = Mean interval between waves in seconds (inactive)

Minimum and maximum values for O were calculated by taking the lowest and highest values for O out of the four calculations using the two extreme values of A and the two values $I \pm sd$.

Table 5.7: Burst and inter-burst intervals at various postnatal stages in the ferret. The range of values for the overall probabilities $[O_{\min}, O_{\max}]$ can be compared to the values used in the simulations, shown in Table 5.1. Data taken from (Wong et al., 1993, p926).

waves and more detailed modelling of the retina are required before we can use the biological data to guide selection of model parameters such as wave width. As is shown in the work with the on- and off-centre units in Section 5.5, wide waves do not inhibit topographic refinement, as long as narrower waves are initially used to lay out the overall topography.

5.6.3 The nature of correlations between on- and off-centre retinal units

In Section 5.5 it was assumed that polarity-specific waves were either on-centre, off-centre or mixed. The results from the model showed that as the frequency of on-centre waves decreased, so did the degree of polarity segregation. The biological data does not support this result because segregation occurs in the absence of on-centre waves. Wong and Oakley (1996) note that during the period of polarity segregation, the on-centre cells fire for only around 25–35% of the time that off-centre cells fire, and that when on-centre cells fire, neighbouring off-centre cells nearly always also fire. Assuming that no on-centre waves are generated, this corresponds to simulation parameters $p_{\text{on}} = 0.0$ and $p_{\text{off}} = 0.66$. Given the failure of the network to develop polarity-specific units when $p_{\text{on}} = 0.42$ and $p_{\text{off}} = 0.92$ (Figure 5.15), it was not expected to see the network segregate using these more realistic values. An extra simulation was run using $p_{\text{on}} = 0.0, p_{\text{off}} = 0.66$, which produced similar results to those for $p_{\text{on}} = 0.42, p_{\text{off}} = 0.92$ (mean value of f_j for on-centre units $= 0.018$, s.d. $= 0.016$, n $= 36$; mean value of f_j for off-centre units $= -0.014$, s.d. $= 0.011$, n $= 44$).

Here the results from the network are therefore in disagreement with the biological data. The current model can only segregate on- and off-centre inputs when the inputs are anticorrelated. These anticorrelations are only present much later in life, when the retina is performing its normal function of converting light into electrical signals (Mastronarde, 1989). However, by this time, units in the LGN are already polarity-specific and are organised into sublamina. This rules out any

light-evoked anticorrelations between on- and off-centre cells driving polarity segregation.

5.7 Discussion

The formulation of the waves used in this thesis is relatively simple, but has allowed us to investigate the effect of the spatio-temporal wave properties upon development. This chapter has highlighted the following issues:

1. The probability of wave generation affects development. When the waves are frequently generated, the global retinotopic mapping breaks down although there is still some degree of local order. In contrast, the monocularity status of postsynaptic units are hardly affected. At the other end of the scale, when waves are not generated very often, the covariance rule causes most weights to converge to the same value, preventing both receptive field refinement and ocular segregation. A modified weight adaptation rule, the active-covariance rule, is presented which does not increase weights when both input and output activities are below threshold. This restores normal development in the model under conditions of sparse input. A simple comparison with biological data shows that such sparse inputs are likely to be present in the natural retina during the period of retinogeniculate development.

2. The model can account for monocular deprivation results if the presynaptic normalisation is weakened. This is achieved by allowing the sum of weights from a presynaptic unit to be less than, rather than strictly equal to, a target value. Without this weaker form of normalisation, presynaptic units remain connected to the postsynaptic sheet despite their lack of activity. It would be interesting to see how the projections columns generated by the network (shown in Figure 5.7) compare with the projection columns taken from a monocularly deprived animal.

3. Wave width affects overall topography and ocularity segregation. As the width of waves is increased, although units still refine their receptive fields, the overall size of receptive fields increases and the overall topographic mapping breaks down. Network development is robust to wide waves however, as long as narrow waves are initially used to set up the overall topographic mapping. A similar result has recently been found for retinal ganglion cells in the developing turtle retina (Sernagor & Grzywacz, 1996).

4. The model predicts segregation of on- and off-centre inputs, but only in the presence of anticorrelated on- and off-centre activity. The current model cannot segregate into polarity-specific units using input correlations similar to those recently discovered in the ferret retina (Wong & Oakley, 1996). In contrast, recent modelling work using a BCM rule has shown that a single postsynaptic unit can become responsive to only one polarity using these correlations (Lee & Wong, 1996).

5.7.1 New mechanisms introduced

This chapter has introduced two new mechanisms to allow the model to account for different aspects of development. Both of these are now discussed.

The active-covariance rule

The active-covariance rule specifies no weight change when both pre- and postsynaptic activities are below threshold (case four of the covariance rule). This rule allows the network to develop as normal when waves are rarely present on the retina. The effect of ignoring case four is not quite the same as preventing weight adaptation when both eyes are quiet: case four is normally invoked when both eyes are quiet, but is also invoked when at least one eye is active. Hence the fraction of time case four is invoked should always be greater than the value for O_0 — the fraction of time for which both eyes are quiet. (The remaining time case four is used is when there is some input to the network, but not enough to drive all postsynaptic units above threshold). Comparing the values of O_0 and the usage of case four in Tables 5.1 and 5.3, the case when both eyes are quiet accounts for almost all of the usage of case four of the covariance rule. The active-covariance rule therefore effectively prevents weight update in the absence of any presynaptic activity.

Case four of the covariance rule has attracted attention in other related research. First, Willshaw and Dayan (1990) noted that the covariance rule has "the biologically undesirable property that changes can occur when neither pre- nor postsynaptic neuron is active", and mention that this case just changes weights by a constant amount in the absence of any activity. This is a positive term however, and so it does not represent a decay term as might be expected. Second, Montague et al. (1991) have shown that the case is not required for development of ocular dominance columns, although in their case the network's inputs were anticorrelated and they did not compare the results of using the covariance rule with and without this case. (The inclusion of case four may inhibit ocular dominance in their model.)

Weakening presynaptic normalisation

In Section 4.2.1, the parameter r_s was used to weaken subtractive normalisation so that the network could develop using subtractive presynaptic normalisation. In this chapter, we have presented another way of weakening normalisation such that $\sum_{j=1}^{N_{\text{post}}} w_{ij} \leq T_{\text{pre}}$. (Here we have only weakened divisive normalisation, but subtractive normalisation can be similarly weakened by ignoring the case when normalisation should add a constant value to each unit.) With this constraint, the weights are allowed to be positioned within the constraint space, rather than strictly on the constraint surface. This weaker constraint allows the model to replicate monocular deprivation results showing that if an eye is not active, it will eventually disconnect from the postsynaptic sheet (all of the weights associated with the deprived eye will go to zero). In comparison, the standard form of normalisation prevents presynaptic units from disconnecting from the postsynaptic sheet, regardless of the activity levels of presynaptic units.

5.7.2 The polarity experiments

Unfortunately, the results from the polarity-specific wave experiments are at odds with the current biological data on two counts. First, postsynaptic units only become responsive to inputs of one polarity when the on- and off-centre inputs are anticorrelated. As these anticorrelations are weakened in correspondence with the biological data, polarity segregation is inhibited. Recent success at polarity segregation of one postsynaptic unit in the presence of weaker anticorrelations using the BCM rule (Lee & Wong, 1996) suggests that the covariance rule may not be suitable for modelling polarity development. Second, even when we assume that on- and off-centre in-

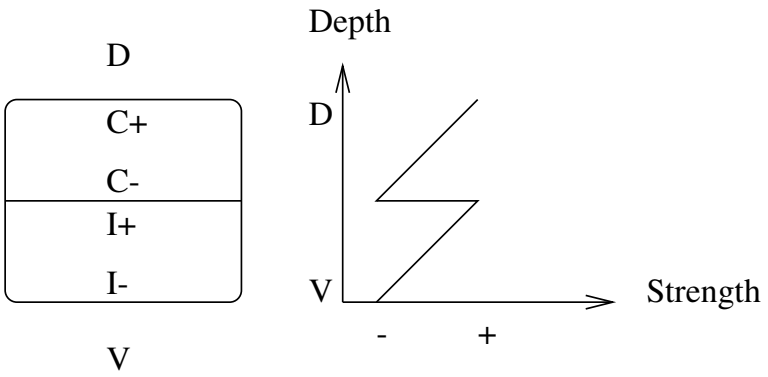
puts are anticorrelated, although postsynaptic units become responsive to one polarity, there is no consistent sublamination effect. This lack of sublayering is due to two reasons:

1. No systematic bias. In the ferret, the sublayering is consistent across animals, so that the on-centre units always cluster in the rostral regions of a layer, and the off-centre units into the caudal regions (see Figure 2.4). Similar clustering tendencies, although slightly weaker, are found in the cat. This would seem to indicate some form of systematic bias in the natural system, either in the initial arrangement of connections or through chemical gradients. However, no such bias has been suggested in the biological literature. It is possible to introduce a polarity bias into the model, although there are difficulties with putting the bias either in the initial weights or using chemical gradients.

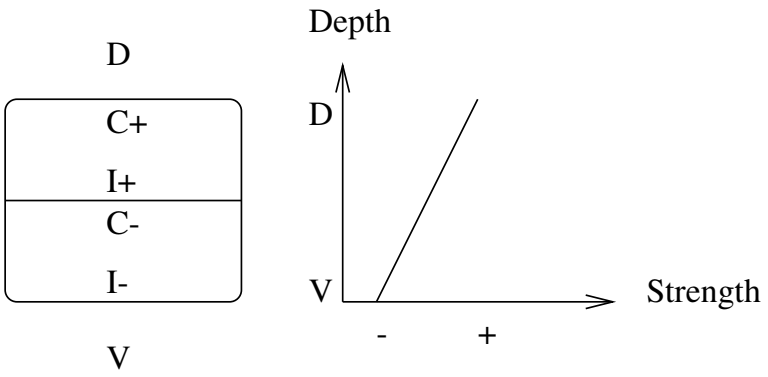
   Weights could be initialised so that certain postsynaptic units received more initial weight strength from either on- or off-centre presynaptic units. This can be objected to on both theoretical and practical grounds. First, assuming development begins before the period of polarity segregation (around P16 in the ferret), some retinal ganglion cells are still classified as mixed on-off cells since their dendritic trees have not segregated into one sublayer of the IPL. This segregation is activity-dependent, complicating prediction of a retinal cell's polarity (Bodnarenko & Chalupa, 1993; Bodnarenko, Jeyarasasingam, & Chalupa, 1995). It is therefore difficult to know beforehand the polarity of each retinal ganglion cell. Without knowing the polarity of a presynaptic unit it is difficult to know how to bias its weights in advance to help the layering. Second, even if we know the polarity of each presynaptic unit, any initial weight bias may be lost during the early stages of development before the onset of polarity-specific waves.

   Chemical gradients could also be used to bias on-centre units to connect to the top of each eye-specific layer. Figure 5.16 shows two possible biases that could be used. The first should ensure that eye-specific layers are preserved, but this entails that the chemical bias knows where the division between layer A and layer A1 occurs. The second bias is much simpler, but may have the side-effect of destroying the eye-specific layers if the polarity bias is too strong. Both weight bias and chemical gradients therefore have different drawbacks.

2. Nature of growth rule. The growth rule used here does not discriminate between on- and off-centre units. Any weight increases favouring an on-centre unit will be distributed to both neighbouring on- and off-centre units by the growth rule, and hence there is no significant segregation of polarity until the growth rule terminates. Once the growth rule stops, postsynaptic units can sharpen their responses to units of one polarity. There is no longer any direct interaction between neighbouring postsynaptic units however, and so neighbouring units cannot influence each other with respect to the polarity to which it becomes dominant.



(a) Bias including ocularity



(b) Bias without ocularity

Figure 5.16: Chemical gradients to bias the formation of on- and off-centre sublamina in the LGN. On the left a coronal slice of the cat LGN is shown, and on the right a possible chemical gradient, showing the strength of some chemical as a function of LGN depth. The strength of chemical indicates an affinity for on-centre (+) cells. (a): Chemical bias with ocularity information to prevent disruption of eye-specific layering. (b): A simpler chemical bias without ocularity information. This second bias will probably affect eye-specific segregation as a side-effect of polarity segregation, causing the LGN to be ordered according to polarity rather than ocularity. Key: C – contralateral eye input. I – ipsilateral eye input. D – dorsal. V – ventral.