

## Chapter 2

# Visual Deprivation

### 2.1 Introduction

There are two ways that a theory can be verified. One way is to spin out the consequences of the theory, and make predictions that can then be measured. Another is to directly measure some of the basic postulates of the theory. Mostly, this chapter uses the former approach, but recent advances have allowed us to use the latter approach in particular cases, as we have seen in the LTP/LTD discussion of the previous chapter. In order to understand the questions which can be asked about learning in the visual cortex we need to summarize the basic anatomy of the visual system, and describe the properties of each element in the system. This is done in the first part of the chapter. We then introduce a more sophisticated model, which has a rough correspondence with the anatomy of the visual system, and explore its consequences.

The goal of this chapter is to establish a model about which one can ask detailed questions. This allows us to compare precisely the theoretical consequences with experiment. We will use the experiments in visual deprivation as a foundation for this comparison, and ask about the dependence of the model on various parameters, trying to find a consistent set of parameters for the model. The dependence on the parameters allows us to predict testable behavior which can both distinguish between learning rules, and give some understanding about the underlying mechanisms.

### 2.2 The Visual System

The biological visual pathway is shown in Figure 2.1. Light from the visual field is projected onto the retina on the back of each eye. The retina is made up of an array of photoreceptors which change light into electrical signals. These electrical signals are then processed by the retinal circuitry and projected by the retinal ganglion cells onto the optic nerve. Signals from both eyes cross at the optic chiasm, and continue to the lateral geniculate nucleus (LGN). In the LGN the inputs from both eyes remain

separate, segregated into different layers. The cells in each layer of the LGN also maintain a map of the retinal organization: nearby cells in the LGN correspond to nearby cells on the retina (which in turn correspond to nearby points in the visual field). The LGN cells project then to the primary visual cortex, more specifically Layer IV of the visual cortex. This is the first place in the visual processing where inputs from *both* eyes come together.

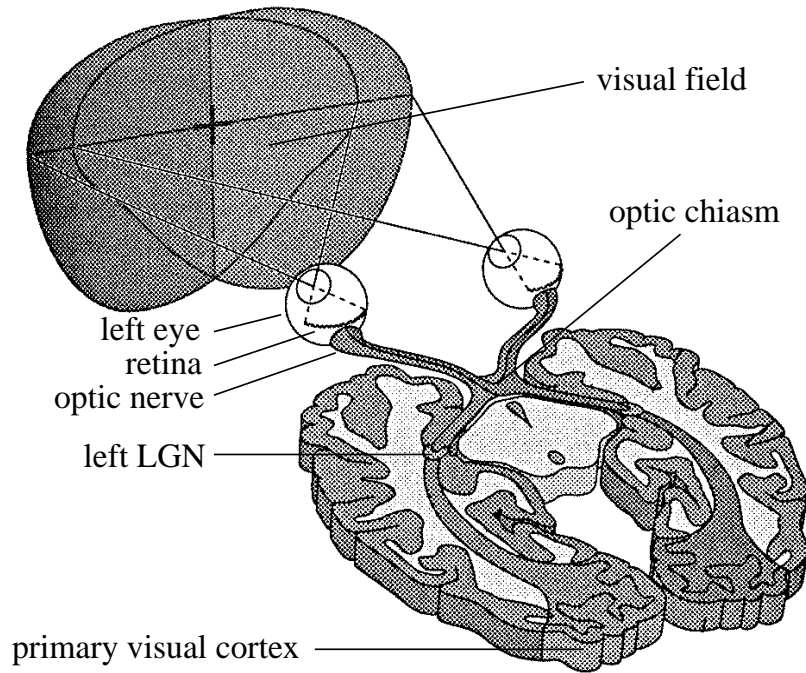


Figure 2.1: A horizontal slab through the brain exposing the visual pathway. (adapted from: Bear, Connors, Paradiso, *Neuroscience: Exploring the Brain*)

In every stage of visual processing, each cell has a *receptive field*. The receptive field is defined to be that area on the retina in which appropriate light stimulation changes the response of a cell. If the cell only responds to stimulus in a small part of the retina, then the receptive field is considered to be localized. If the cell responds to a bar of light of a particular orientation, then the receptive field can be considered *oriented*. The receptive field can be thought of as a simple function of two dimensions (see Figure 2.2). Since there are a discrete number of connections to a neuron, we sample this two dimensional function at those positions where the inputs lie. The value of the function at a particular position represents the effect of an input at that position on the activity of the cell. For example, if the function is positive at a particular position, then that input is effectively *excitatory*, or in other words, tends to increase the activity of the target cell. If the function is negative at a particular position, then that input is effectively *inhibitory*, or it tends to decrease the activity of the target cell.

The receptive field (RF) for an ON-center LGN cell is shown in Figure 2.2. The cell responds strongly to a small spot of light in that part of the visual field which lies on the center of the RF (thus, it is an ON-center cell). A spot of light in the surrounding area causes an inhibited response, which has the effect of making these cells unresponsive to large patches of uniform illumination. Both

retinal and LGN cells have this behavior. But what gives rise to the receptive field? The receptive field is influenced in several ways. The receptive field can be strongly influenced by those direct pathways between the cell and the photoreceptors. These pathways, however, can be altered by the efficacy of the synapses involved, so direct connections may not define the receptive field. Neighboring cells can also influence the receptive field, in a network effect. We ignore the network influences here, and focus on those influences caused by changes in synaptic efficacy.

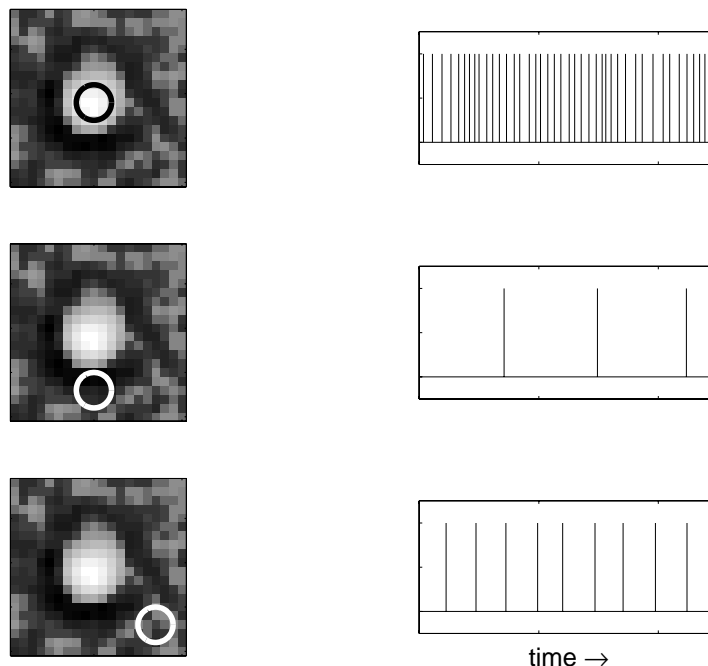


Figure 2.2: Example Receptive Field for an ON-center LGN cell. Shown are spots of light on different parts of the receptive field (left) and the resulting spike trains (right). Shining a spot of light in the center yields a strong response (top). A spot of light in the surround gives an inhibited response (middle), and outside of the receptive field, or at least beyond the surround, gives spontaneous activity (bottom). Copyright 1995, Izumi Ohzawa.

An input can be *functionally* disconnected if the synaptic efficacy, for example, is weak. A specific arrangement of synaptic efficacies can then account for the receptive field properties of a cell. In the case of ON-center ganglion cells, there is a strong positive weight (excitatory) from the photoreceptors in the center of the RF, and a strong negative weight (inhibitory) from the photoreceptors in the surround portion of the RF. The synaptic modification covered in the previous chapter could be a very convenient way for a cell to refine the structure which is partially in place from the anatomy. Therefore it is crucial to understand the response properties of these cells, as well as the anatomy, in order to uncover the possible changes in the functional connections of one cell to another.

It is known that retinal ganglion cells respond strongly to spots of light, as do LGN cells, and that the visual cortical cells receiving direct input from the LGN respond strongly to bars of light of a particular orientation (Hubel and Wiesel, 1959). Historically, once a stimulus was found to which a cell responded, experiments were done by varying the parameters of the stimulus. For example, once Hubel

and Wiesel discovered that visual cortical cells responded to bars of light of a particular orientation, then cells were characterized by the parameters of that stimulus, and the receptive field was then assumed. A cell might respond to a bar of a certain length at  $30^\circ$  angle, so a receptive field was assumed to be oriented at  $30^\circ$ , and have excitatory and inhibitory regions corresponding to the bright and dark parts of the bar stimulus. Though this may be correct, the receptive field is a two dimensional object (three dimensional if you count time, but we will leave discussions about that for later), so varying a few of parameters of an assumed stimulus may hide a lot of the structure.

To help deal with some of these inadequacies, a procedure called reverse correlation (Jones and Palmer, 1987) was introduced. This permits an unbiased measurement of the receptive field. In this procedure, one rapidly presents an image made of many small, uniformly distributed, random, bright and dark stimuli, and cross correlates between this random stimulus and the neuronal spike train elicited (see Figure 2.3). The cross correlation is done after a fixed delay, giving a picture of the receptive field without resorting to assumptions about the optimum stimulus to a neuron. One could also make this delay a parameter and obtain both the spatial and temporal structure of the receptive fields. Example receptive fields obtained using this method, for cells in the LGN and visual cortex are shown in Figure 2.4.

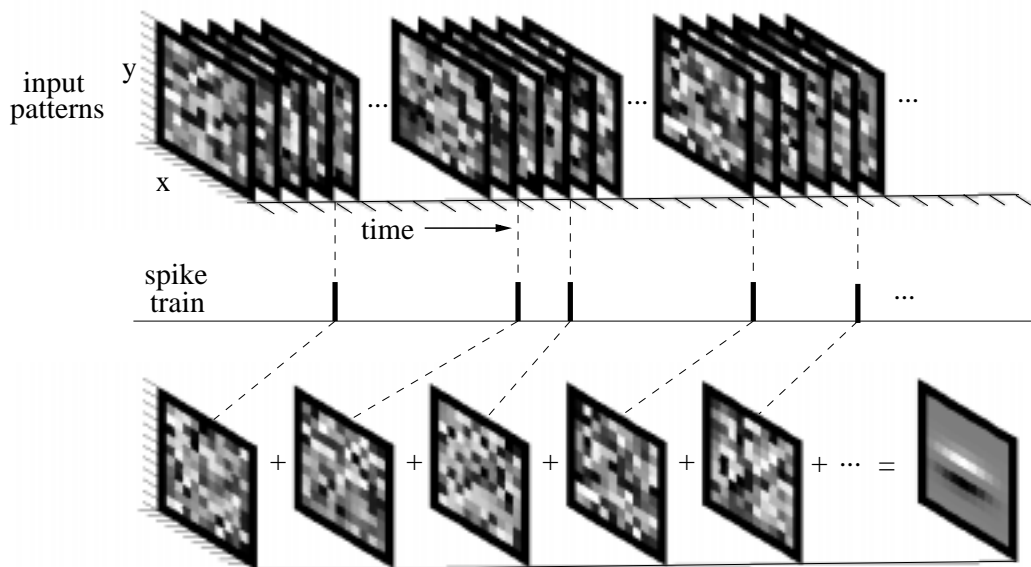


Figure 2.3: Reverse Correlation Example. Random stimulus is presented, and those patterns which give responses are added up to yield the receptive field.

Both retinal cells and LGN cells have a center-surround receptive field structure. In both areas, there are ON-center and OFF-center cells, which have excitatory or inhibitory centers, respectively, and the surrounds which are the opposite. There are also two types of cells, called X and Y cells, which have both ON- and OFF-center flavors, but have different properties. X cells have smaller receptive fields, and respond slowly to moving stimuli. Y cells have larger receptive fields, and respond rapidly to

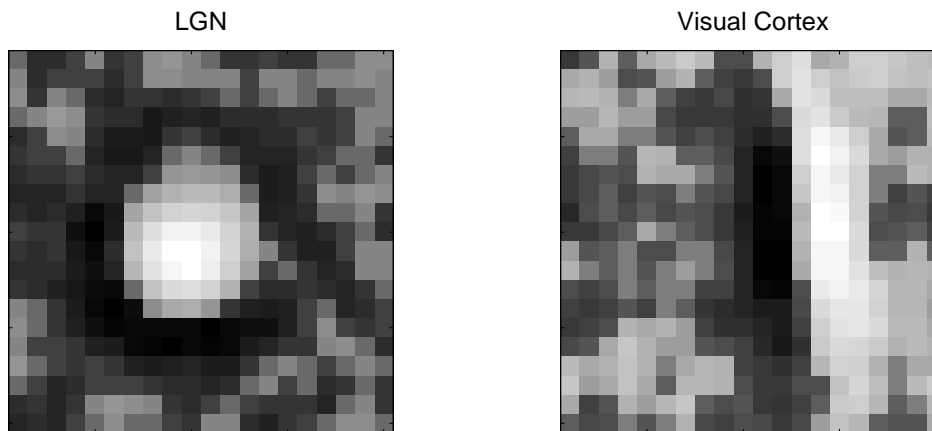


Figure 2.4: Example Receptive Fields for cells in the LGN and visual cortex obtained using reverse correlation. Copyright 1995-1997, Izumi Ohzawa.

moving stimuli. We will currently restrict our study to X cells because X cells have linear response to their input, and form the majority of the cells projecting to the visual cortex (Orban, 1984). A possible role for these different cells is explored in Chapter 4.

The LGN cells have similar response properties to retinal cells, and the organization of the LGN maintains the retinal map. This suggests that the LGN does little except to simply channel the retinal input. It is unclear, however, what other function the LGN serves in visual processing. It is certainly false that the LGN serves only to relay the information from the retina to the cortex, because more of the inputs into the LGN are *feedback-projections from the cortex* than from the retina! The role of the feedback is not yet known, and we do not include it in our models.

Visual cortical cells,<sup>1</sup> in Layer 4 of the cat, have oriented receptive fields, that is they respond to bars of light of a particular orientation. Hubel and Wiesel proposed that this could occur by an oriented arrangement of LGN inputs, each having center-surround responses (Figure 2.5). If several LGN cells, whose centers lie in a line, were to be functionally connected to a single cortical cell then that cortical cell would have a strong response to a bar oriented along the line of LGN receptive field centers. Recent experiments (Reid et al., 1996; Ferster et al., 1996) have given more quantitative support to this idea.

As well as being orientation selective, cortical cells tend to be *direction selective*, which means they respond stronger to an oriented bar moving in one direction than the opposite direction. This type of selectivity is clearly temporal, as opposed to orientation selectivity which is spatial, but it may arise from the similar mechanisms nonetheless. We explore this more in Section 4.3.

Most cortical cells are also *binocular*, which means that they respond to stimulus from either eye. There are a few cells which are *monocular* (respond to only one eye), but they are far fewer in number. The amount by which a cell responds to one eye or another is called *ocular dominance*.

Ocular dominance, orientation selectivity, and direction selectivity can be changed with visual

---

<sup>1</sup>I will be loose in my usage of the term “cortical cell” in this work, referring to simple cells in Layer 4 of the cat, unless I specify otherwise.

experience. The big question we want to answer is: What is the role of *visual experience* in the formation and maintenance of these properties? If we assume that changes in orientation selectivity and ocular dominance are due to *synaptic* changes, then we need to determine the rules by which the synapses change as a function of visual experience, in order to yield these properties. In the previous chapter we explored simple versions of these rules, but we will need to expand the models considerably to give a realistic representation of the biological system.

In the next section, we review the experimental results where the input environment is changed in specific ways to yield changes in ocular dominance and orientation selectivity. We then will present a model to generate an explanation (or perhaps multiple explanations) of the phenomena observed in the experiments. This will lead us into predictions from the models which can then be tested, a detailed exploration of the parameter space, and will give us a way to differentiate between the different learning rules.

## 2.3 Experimentally Modifying the Input Environment

There are some experiments that are fairly straightforward in both their implementation and their interpretation, and also give robust effects. On these we base most of our modeling. These include binocular and monocular deprivation, reverse suture, and strabismus. There are many variations, and different protocols, which are either not quite as straightforward to implement or interpret, or haven't been performed enough to be considered with as much weight as others. These experiments often seem very suggestive and can be understood, at least qualitatively, using the same arguments, but they are not the focus of the comparison we wish to make with experiment.

### 2.3.1 Normal Development

Around 1958, Hubel and Wiesel began their search for the receptive field properties in the visual cortex. They were motivated by earlier work by Kuffler on the retina, but saw a great deal of anatomical structure in the cortex and hoped that they would find something more than center-surround receptive fields there. They recorded the responses of small groups of cortical cells, a technique which was then quite new, and presented spots of light at different places on the retina. The apparatus consisted of different masks made of metal which one slid in front of a light source. After a lot of failed attempts to get responses, they noticed that the cell responded just a little more strongly when they were initially inserting, or removing, the mask. This fortuitous observation led them to recognize that the cells were responding to the *edge* of the mask, and it was shortly thereafter that they found that the edge had to be of a particular orientation.

They postulated that such an oriented, elongated, receptive field could be formed by the convergence of several LGN cells, with center-surround receptive fields lying along a particular direction

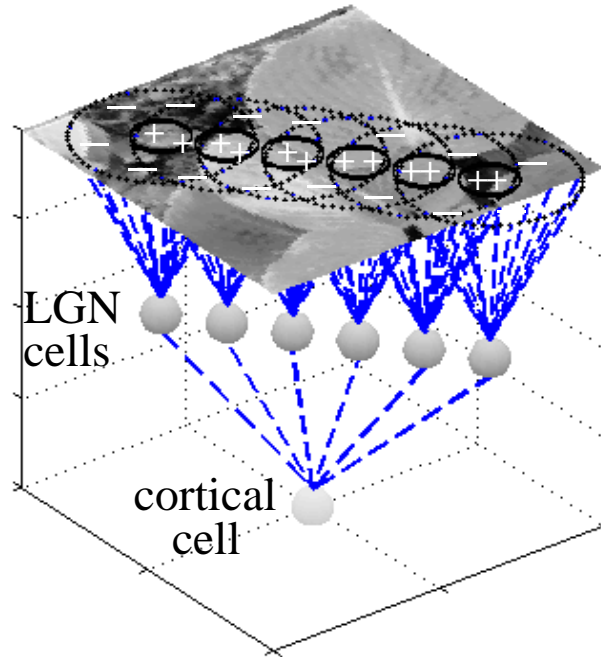


Figure 2.5: Possible construction of an oriented receptive field by the convergence of several LGN cells, with center-surround receptive fields (proposed by Hubel and Wiesel to explain orientation selective cortical cells). Shown is a layer of LGN cells projecting to a single cortical cell. If there are strong connections between LGN cells falling in a line, then a bar stimulus along that orientation will excite the centers of many LGN cells, and cause a larger response in the cortical cell.

(Figure 2.5). The question naturally arises, whether these receptive fields are predetermined (genetically), or whether they are affected by the environment in any way.

If they are affected by the environment, then we can formulate many questions as a consequence. What is required to get *oriented* receptive fields? How can the RFs be modified? What are the mechanisms behind the changes? We need now to outline the results from experiments where the receptive field properties were measured in different visual environments. A good review of these experiments, upon which this section is based, can be found in (Movshon and van Sluyters, 1981).

For the cat, the optics of the eye are cloudy until around 16 days after eye opening. After that, visual manipulation can affect the receptive field properties of cortical neurons until the kitten is several months old, the length of the *critical period*. This critical period marks the time when the neurons are plastic, or in other words, the time when significant synaptic modification can occur. Experiments using various visual environments must therefore occur within this time window to be effective.

It has been claimed that all of the properties of receptive fields in adult cats (binocularity, orientation and direction selectivity) are present in kittens upon eye opening (Hubel and Wiesel, 1963), only with small responses. This claim has been challenged by Pettigrew (1974), who claimed that most cells

lacked selectivity. Most other claims (Frégnac and Imbert, 1978; Blakemore and Van-Sluyters, 1975) fall between these two extremes. It is generally agreed that many cells in very young cortex lack selectivity, and a significant amount are not driven at all by visual input. It is also agreed that binocularity is roughly present, but in an immature form, in young cortex.

### 2.3.2 Binocular Deprivation

Raising animals either in total darkness, with binocular lid suture, or diffuse lenses covering both eyes has been shown to alter the development of the receptive field properties of visual cortical cells. Up to 3 weeks of such rearing (from eye opening), there is little difference between the properties in dark reared versus normally reared animals (Buisseret and Imbert, 1976; Frégnac and Imbert, 1978). This is at least partially due to the time it takes for the optics to clear. Long term deprivation results in a large percentage of unresponsive cells, most responsive cells remain binocularly activated, but most of the selective cells tend to be monocular (Blakemore and van Sluyters, 1974; Frégnac and Imbert, 1978; Leventhal and Hirsh, 1980). There is only anecdotal evidence for significant differences between the effects of total deprivation (dark rearing) and pattern deprivation (bilateral lid suture).

It has also been shown (Wiesel and Hubel, 1962; Freeman et al., 1981) that brief binocular deprivation, on the order of 3 or 4 days, causes deterioration of the receptive responses in normally reared cats. The cells, however, remain binocular and direction selective (reduced orientation selectivity). It is unclear what the exact time course of the loss of response, so it is more difficult to be conclusive about the mechanism involved. Chronic experiments need to be done to get a better idea of the time course of the deprivation.

### 2.3.3 Monocular Deprivation and Reverse Suture

There is a very striking difference between the effects of monocular and binocular deprivation. When one eye of a previously normally reared kitten is deprived of patterned input, cells in kitten's visual cortex change from mostly binocular to almost exclusively monocular: in as little as 24 hours most cells lose their response to stimulation through the deprived eye and can only be driven through the eye that remains open (Mioche and Singer, 1989; Wiesel and Hubel, 1963). This ocular dominance shift is not caused simply by the atrophy of neurons connected to the closed eye, because there is no measured decrease in the number of neurons which are visually responsive.

It appears that the lack of *patterned* visual input to the eye is required to attain these deprivation effects. Using a translucent contact lens, and keeping the total flux into both eyes the same, produces the same striking ocular dominance shift (Blakemore, 1976; Wiesel and Hubel, 1965). Merely reducing the total amount of light entering the eye, while maintaining pattern vision, does not produce a significant effect (Blakemore, 1976).

The very fast change in the ocular dominance of the cells during monocular deprivation (MD) is

likely to be due to the change of the efficacy of synapses from the closed eye. Another experiment, that supports this claim, uses a procedure called reverse suture (RS). In this procedure, there is an initial period of monocular deprivation: after the cortical neurons have become monocular, the deprived eye is opened and the other eye closed. In this situation the cortical neurons lose responsiveness to the newly closed eye, and become responsive to the newly opened eye (Blakemore and van Sluyters, 1974). At least 24 hours of RS is required before the responses to the deprived eye reappears (Mioche and Singer, 1989). The time it takes to lose the response in monocular deprivation, and to recover it after the reverse suture is probably too short to be entirely due to the addition or subtraction of new connections (Jacobson et al., 1985). The change in efficacy of connections, whatever the mechanism is, should be primarily responsible for these changes. The exact speed at which the deprivation effects occur is a function of where we are in the critical period. After five weeks of age, a constant deprivation period has less and less of an effect.

### 2.3.4 Strabismus

Hubel and Wiesel (1965) performed an experiment where the eyes of the kitten were artificially misaligned. After rearing with this condition, a drastic reduction in binocularity in the cortical cells was found, with most cells remaining responsive to one or the other eye. This result has been widely reproduced (van Sluyters, 1977; van Sluyters and Levitt, 1980; Blakemore, 1976). A common interpretation of this result is that the lack of correlation between the two eyes is responsible for the degradation of the binocular connections.

### 2.3.5 Stripe and Strobe Rearing

These experiments are not as straightforward in their implementation, nor have they been performed enough to give us as strong a confidence in their results as the preceding experiments. Nonetheless, they should be considered, because they seem quite suggestive and can give us some insights into the possible mechanisms for plasticity.

There have been numerous experiments in restricted environments to see the effects on cortical receptive fields. Examples include rearing in a striped (Blakemore and Cooper, 1970; Stryker et al., 1978) or spotted (Pettigrew and Freeman, 1973) environments, monocular deprivation of *particular* orientations (Cynander and Mitchell, 1977; Rauschecker and Singer, 1979). These experiments show particular biases in the receptive field formation. Rearing in a striped environment produces a predominance of cells of the stripe orientations. Using a particular lens which blurred a particular orientation, but correctly focussed the orthogonal orientation, Cynander and Mitchell were able to perform monocular deprivation of *particular* orientations, and Rauschecker and Singer performed a similar reverse suture experiment. Both groups measured orientation dependent ocular dominance shifts. This suggests that the deprivation effect is dependent on the cortical cells themselves, because only they would know about

orientation information.

Some experiments use rearing in restrictive motion environments to disrupt normal direction selectivity development. Rearing kittens in an environment illuminated by a low frequency (1 Hz) strobe has severe effects, causing an almost complete loss of both direction and orientation selectivity (Cynander et al., 1973). Rearing in a higher rate strobe light reduces the proportion of direction selective cells, leaving the spatial receptive field unaltered (Cynander and Chernenko, 1976).

Cats reared in an environment with motion restricted to one direction showed a bias in the direction selectivity (Cynander et al., 1975). Rearing in a moving striped environment produces a bias in both direction and orientation selectivity (Tretter et al., 1975).

### 2.3.6 Conclusions about Visual Environment Modification

It is clear that the cortex changes significantly when the visual environment changes. From the many experiments performed, it also seems clear that the neurons are modified by the *structure* in the environment (or lack thereof), and can be biased in very specific ways by the manipulation of the environment. Our initial goal will then be to model these different environments, both in an attempt to reproduce the experimental results but also to highlight those aspects of the model which give rise to these properties. This will both let us compare different models and make predictions for further experiments.

## 2.4 Modeling the Development Orientation Selectivity and Ocular Dominance

### 2.4.1 Normal Rearing (NR)

Figure 2.6 shows the architecture of our model and Figure 2.7 shows the environment used for our simulations. Patches are taken from 12 images of natural scenes and presented to left and right retinal cells. The eyes in these simulations see the same patch of the image. This assumption is easily relaxed, but for now we will deal with the simpler case of identical inputs to the two eyes. These cells have an ON-center, OFF-surround response, modeled here as a difference of Gaussians (DOG) filter. This is commonly used to model the processing done in the retina (Law and Cooper, 1994). The LGN in the model serves only to relay the signals from the retina to the cortex, so the inputs to the cortex are the retinally processed patches taken from the images.

The weights of the cortical neuron are then updated using a learning rule, such as BCM (Equation 1.4) or PCA (Equation 1.3). The receptive field for the neuron, then, can be visualized by making a gray-scale image of the weights. If white denotes strong responses and black denotes weak responses, then the pattern of this image shows the response of the cell to a spot of light at that position. Since

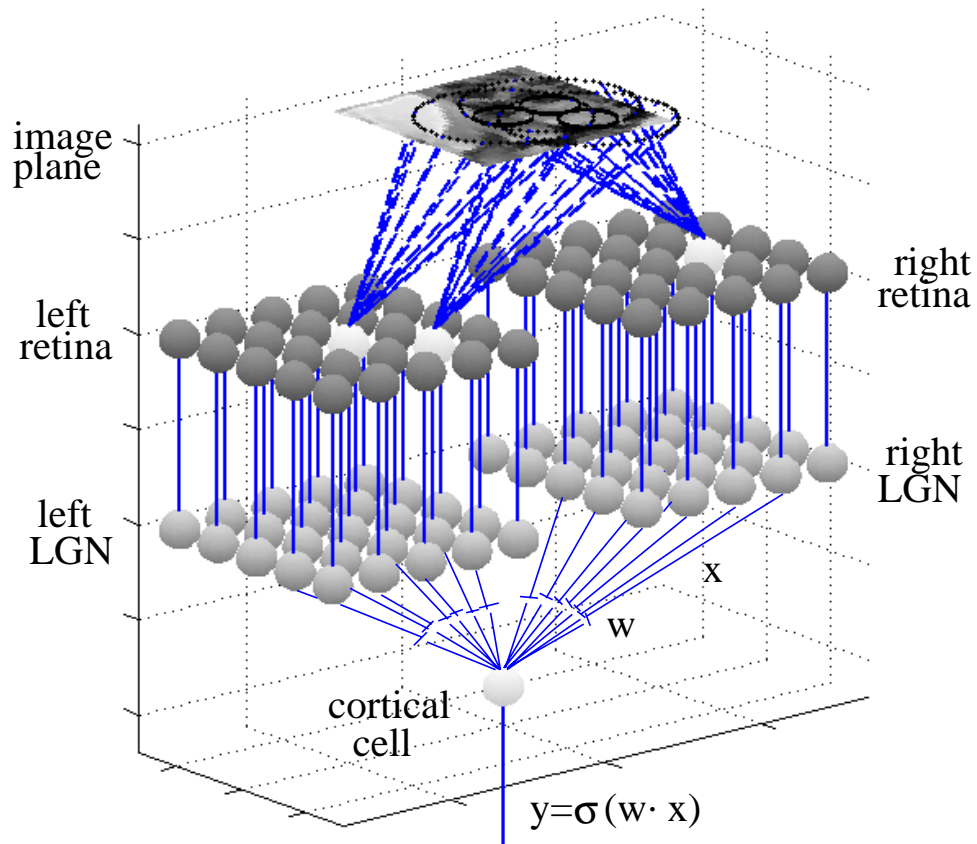


Figure 2.6: Model Architecture. Shown are the image plane (top), the left and right retinal cells, the left and right LGN cells, and the single cortical cell (bottom). In the image plane are drawn sample center-surround receptive fields, for the retinal cells highlighted. Nearby retinal cells see nearby points in the image plane. Retinal cells project directly to LGN cells, on a one-to-one basis. A circular patch of LGN cells projects to the single cortical cell. These projections form the input vector,  $\mathbf{x}$ , for which there is a corresponding weight vector,  $\mathbf{w}$ . The output of the cell is given simply as  $y = \sigma(\mathbf{w} \cdot \mathbf{x})$ .

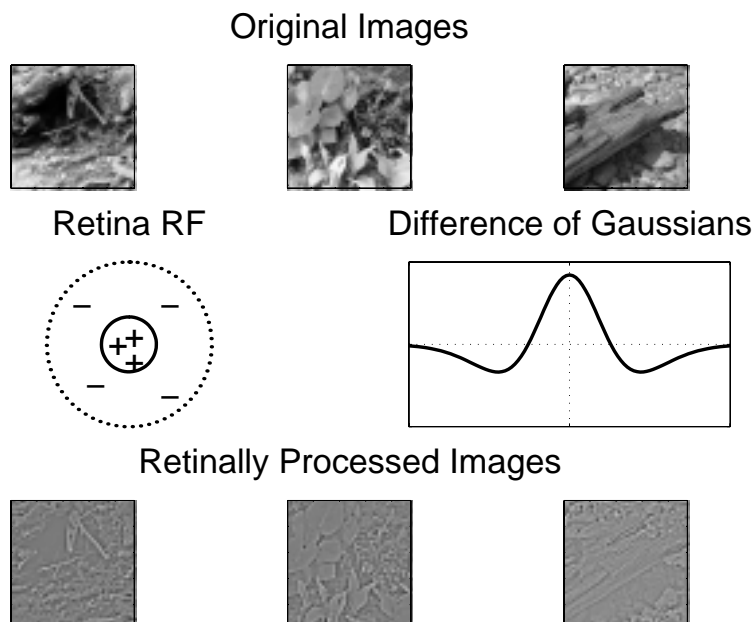


Figure 2.7: Input Environment. Shown are the original images (top) and the retinally processed images used as the actual inputs to the neuron (bottom). The images are processed with a difference of Gaussians (DOG) filter (center, right), which is used as a model of the receptive field properties of the retinal cells (center left).

we can have negative weights and inputs, weak responses are negative. This would correspond to responses *below spontaneous*. Example receptive fields for BCM and PCA, trained in this natural scene environment, are shown in Figure 2.8. We notice that they both achieve oriented receptive fields, just as those observed by Hubel and Wiesel.

Given this architecture and environment, and having introduced some learning rules, we are now in a position to ask some specific questions which relate directly to biology. What is the effect of replacing the natural scene input with deprived input, in models of monocular and binocular deprivation and reverse suture? (Section 2.4.2) How does the deprivation depend on the parameters in the model, and can we use the experimental deprivation results to more fully understand the roles of those parameters? (Section 2.5) To what in the environment is the neuron becoming selective when it forms an oriented receptive field? (Section 3.1) What is the effect of the retinal preprocessing? (Section 4.2) Is it necessary?

Since the deprivation results provide the framework for our comparison, we will introduce a model of deprivation, and then explore the role of the different model parameters on the learning.

## 2.4.2 Deprivation

Modeling deprivation requires an assumption about what the signals from a visually deprived eye would be. One might say that an eye deprived of visual input would send no signals. This would make the inputs to the neuron,  $\mathbf{x}$ , equal to zero in our model. This choice is clearly problematic for the BCM

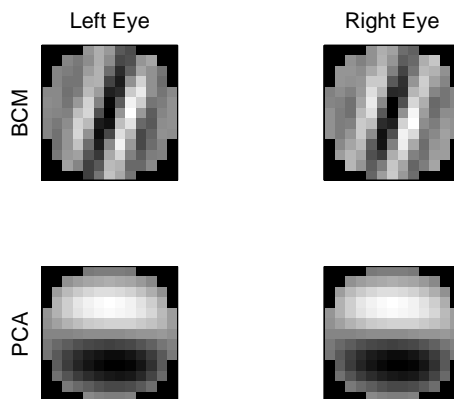


Figure 2.8: Example Receptive Fields from BCM and PCA trained in a natural scene environment. Shown are the left and right eye receptive fields for BCM (top) and PCA (bottom).

rule. From the modification equation

$$\dot{\mathbf{w}} = \phi(y, \theta)\mathbf{x} \quad (1.4)$$

we would conclude that *no* input onto a particular synapse would lead to *no* modification at all, and we would see no possible decrease in the responses from the closed eye. There is, however, some activity from the closed eye inputs in deprivation, even if the deprivation is done in total darkness. The neurons, both in the retina and in the LGN, have a spontaneous level of random firing when they have no input to them. We will refer to this random firing as “noise”, because it presumably carries no significant structure or correlations. The variance of the noise becomes a parameter in the model. We presume that this variance would be larger, say, for deprivation using a diffuse lens as opposed to an opaque eye patch, because the average light intensity would be higher with the diffuse lens. This is currently being investigated experimentally.

If we, then, use uncorrelated noise to correspond to the random firings of deprived eyes, we can straightforwardly model the different deprivation protocols of monocular deprivation (MD), binocular deprivation (BD), and reverse suture (RS). For BD, we follow normal rearing (NR) by presenting noise to both eyes, simulating the deprivation of visual experience to both eyes. MD, likewise, is modeled by following NR, with noise presented to one eye and the other eye is given patterned input from the natural scene environment. Reverse suture (RS) follows MD, but then the eye previously given noise is now presented patterned input, and the previously open eye is given noise. Examples of these simulations, using the PCA and BCM rules, are shown in Figures 2.9 and 2.10, respectively. Notice that both rules are, at least qualitatively, consistent with the ocular dominance shifts described in the experiments.

## PCA

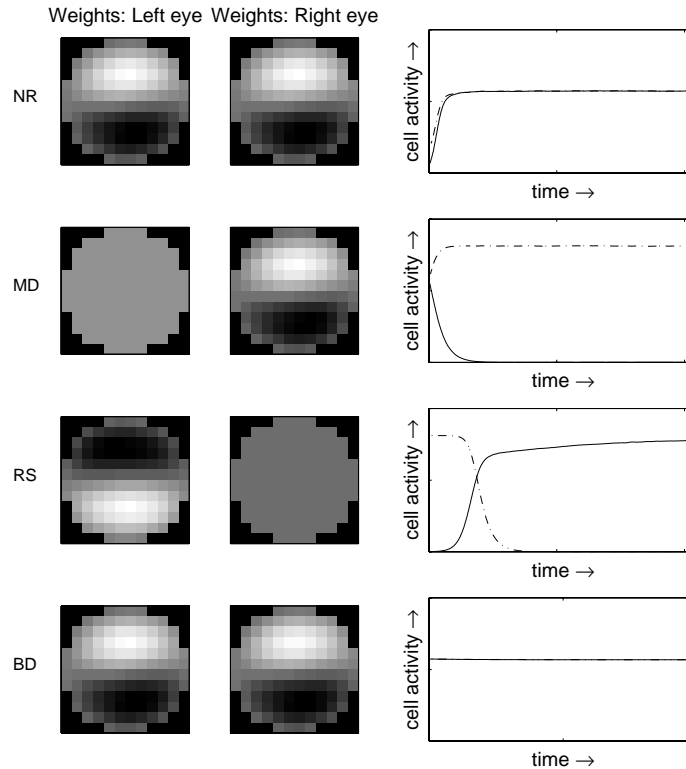


Figure 2.9: Example PCA Simulations. Left: Final weight configuration. Right: Maximum response to oriented stimuli, as a function of time. Simulations from top to bottom are as follows. Normal Rearing (NR): both eyes presented with patterned input. Monocular Deprivation (MD): following NR, one eye is presented with noisy input and the other with patterned input. Reverse Suture: following MD, the eye given noisy input is now given patterned input, and the other eye is given noisy input. Binocular Deprivation (BD): following NR, both eyes are given noisy input. **It is important to note that for PCA if Binocular Deprivation is run longer, selectivity will not be lost.**

## 2.5 Time Course of Deprivation: an exploration of parameter space

We require that the results of the modeling are consistent with *many* biological experiments, using *one set of parameters*. This is crucial to determining the validity of a model, because it provides the most strict tests, and highlights potential problems in the different models.

One straightforward way to compare the models to experiment is to compare the *relative* timing of the ocular dominance shifts during deprivation. For instance, we can measure from simulations the time for the closed eyes to lose responsiveness in binocular, monocular deprivation and reverse suture, and the time for the recovery of the newly open eye in reverse suture. Since we do not know the conversion between simulation time units (iterations) and real time units (seconds) we cannot directly compare the absolute timings of these simulations. We can compare *ratios* of these simulation times to

## BCM

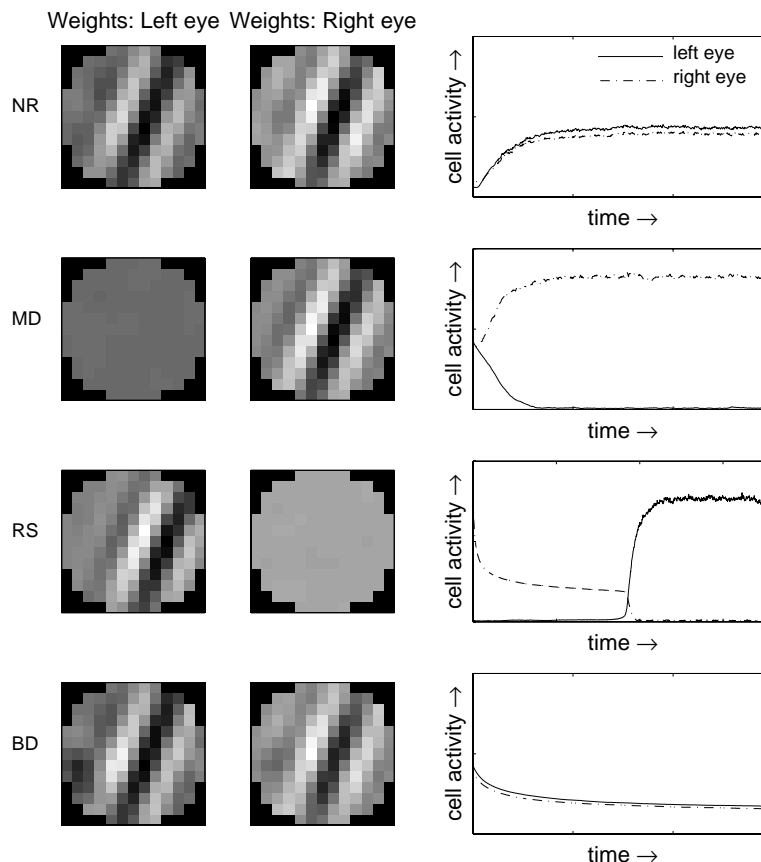


Figure 2.10: Example BCM Simulations. Left: Final weight configuration. Right: Maximum response to oriented stimuli, as a function of time. Simulations from top to bottom are as follows. Normal Rearing (NR): both eyes presented with patterned input. Monocular Deprivation (MD): following NR, one eye is presented with noisy input and the other with patterned input. Reverse Suture: following MD, the eye given noisy input is now given patterned input, and the other eye is given noisy input. Binocular Deprivation (BD): following NR, both eyes are given noisy input. **It is important to note that for BCM if Binocular Deprivation is run longer, selectivity will eventually be lost.**

the same ratios found in experiment, because ratios are not dependent on units.

Some of the questions we would like to answer are *Can the biological experiments in deprivation tell us about the validity of parameters in the theory?*, *Within this restricted parameter regime, how does the time course of deprivation depend on the parameters?*, and *Can we use this dependence to propose experimental measurements of the parameters?*

Since the time course of deprivation is dependent on the neuron parameters, such as the learning rate,  $\eta$ , memory constant (for BCM),  $\tau$ , and the closed eye noise variance,  $\sigma^2$  we can explore the parameter space to find a range of values which match experiment as closely as possible. This range (if it exists at all), as well as the parameter dependence, will depend critically on the modification function, so this approach may be used to distinguish between learning rules. Ideally, we'd like to determine some

of the dependence on the parameters through analysis, but that is only possible in limited cases. We can perform the analysis for PCA (Section 2.5.1), and in very limited cases for BCM (Section 3.6)

We measure the response  $Y(t)$  of the neurons using oriented stimuli. Of particular interest is the characteristic half-rise (half-fall) time for the growth (decay) of neuronal response, referred to as either  $t_{1/2}$  or simply  $\mathcal{T}$ . An example of this is shown in Figure 2.11.

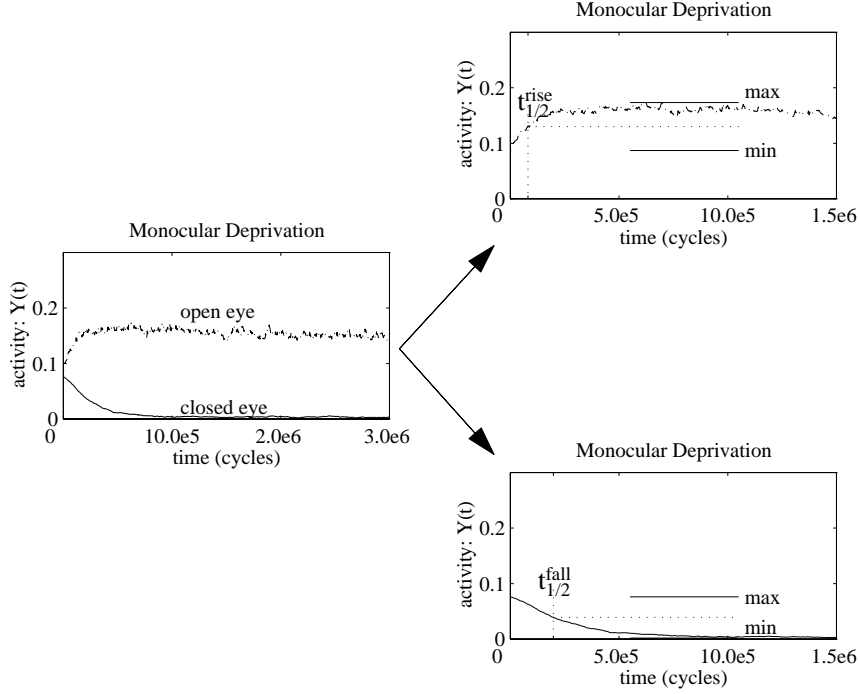


Figure 2.11: Example of a Response Half-Time Measurement. This is an illustration of the procedure for measuring the half-times. Though the example uses a BCM simulation, the specific numbers are not important. The time is measured for the neuron to either rise or fall, half way between its minimum and maximum responses.

Table 2.1 summarizes the experimental results from the deprivation literature. The exact results depend on when in the critical period the experiments were done, and the most appropriate experiments using chronic recording have not been done yet. The results shown in the table allow us to roughly estimate the values of the activity half-rise time,  $\mathcal{T}_{\text{rise}}$ , or half-fall time,  $\mathcal{T}_{\text{fall}}$ , whichever is appropriate for the particular experiment. The ratios of these times, from experiment, are simply

- $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}} \approx 1$
- $1 < \mathcal{T}_{\text{fall}}^{\text{BD}}/\mathcal{T}_{\text{fall}}^{\text{MD}} < 12$
- $2 < \mathcal{T}_{\text{rise}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}} < 16$
- $0.33 < \mathcal{T}_{\text{rise}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{BD}} < 16$

It is unclear whether the  $\mathcal{T}_{\text{rise}}^{\text{RS}}$  time is a good number to use for a comparison, because it depends on how long the preceding MD was performed. If, for instance, one performed MD for weeks and then

| Summary of Experimental Results |   |   |
|---------------------------------|---|---|
| Experiment                      | Reference   | Half-Time $\mathcal{T}$   |
| <b>Monocular Deprivation</b>    | <ul style="list-style-type: none"> <li>• OD changes were observed as <b>early as 6 h</b>(Freeman and Olson, 1982; Mioche and Singer, 1989)</li> <li>• <b>complete loss</b> of response to closed as early as <b>12 h</b>(Mioche and Singer, 1989)</li> <li>• moderate increase of response to the normal eye occasionally(Mioche and Singer, 1989)</li> </ul> | $\mathcal{T}_{\text{fall}}^{\text{MD}} \approx 6\text{-}12 \text{ h}$   |
| <b>Binocular Deprivation</b>    | <ul style="list-style-type: none"> <li>• cortical response reduced <b>within 3 d</b>(Freeman et al., 1981)</li> </ul>   | $\mathcal{T}_{\text{fall}}^{\text{BD}} < 3 \text{ d}$   |
| <b>Reverse Suture</b>           | <ul style="list-style-type: none"> <li>• the time course for the reduction of response to the newly deprived eye was <b>similar to monocular deprivation</b>(Mioche and Singer, 1989)</li> <li>• <b>At least 24 h</b> of reverse suture is required before the responses to the deprived eye reappear(Mioche and Singer, 1989)</li> </ul>                     | $\mathcal{T}_{\text{fall}}^{\text{RS}} \approx \mathcal{T}_{\text{fall}}^{\text{MD}}$<br><br>$\mathcal{T}_{\text{rise}}^{\text{RS}} \approx 1\text{-}4 \text{ d}$ |

Table 2.1: Summary of Experimental Results in Deprivation

followed with RS, then one would expect a slower recovery than RS following a day of MD. Therefore, we will only use the  $\mathcal{T}_{\text{rise}}^{\text{RS}}$  as a very rough test of consistency.

### 2.5.1 PCA: Parameter Dependence

As we saw in Chapter 1, the PCA rule is a lot easier to analyze than the BCM rule. Part of this ease, is due to the fact that the PCA rule depends only on two point correlations, and can thus be completely described by the correlation function of the inputs. Given the correlation function, and the initial conditions, we have a full time-domain solution for PCA(Wyatt and Elfadel, 1995). The detail of the following calculations is given in Appendix A.8.

Since we know that the weights for the PCA neuron become parallel to the eigenvector of the correlation function,  $\mathbf{C}$ , with the maximum eigenvalue (Section 1.3), we expand the initial weight vector in terms of the eigenvectors ( $\mathbf{v}_j$ ) of the natural scene correlation matrix

$$\mathbf{w}_0 \equiv \begin{pmatrix} \mathbf{w}_0^{\text{left}} \\ \mathbf{w}_0^{\text{right}} \end{pmatrix} = \sum_j \begin{pmatrix} a_j^l \mathbf{v}_j \\ a_j^r \mathbf{v}_j \end{pmatrix} \quad (2.1)$$

We then add the assumption that natural scenes are dominated by the first eigenvector of the correlation matrix,  $\mathbf{v}_1$  with eigenvalue  $\lambda_1$ , and take appropriate time limits for the different deprivation experiments. We then obtain the following equations for the time development of the weights.

$$\mathbf{w}^{\text{NR}}(t) = \frac{\sum_j \frac{1}{2} \begin{pmatrix} \mathbf{v}_j \left[ (a_j^l + a_j^r) e^{2\lambda_j t} + (a_j^l - a_j^r) \right] \\ \mathbf{v}_j \left[ (a_j^l + a_j^r) e^{2\lambda_j t} + (a_j^r - a_j^l) \right] \end{pmatrix}}{\left( \frac{1}{2} \sum_j \left[ (a_j^l + a_j^r)^2 e^{4\lambda_j t} + (a_j^l - a_j^r)^2 \right] + 1 - \sum_j \left[ (a_j^l)^2 + (a_j^r)^2 \right] \right)^{1/2}} \quad (2.2)$$

$$\mathbf{w}^{\text{MD}}(t) = \frac{\begin{pmatrix} e^{\lambda_1 t} \mathbf{v}_1 \\ e^{\sigma^2 t} \mathbf{v}_1 \end{pmatrix}}{(e^{2\lambda_1 t} + e^{2\sigma^2 t})^{1/2}} \quad (2.3)$$

$$\mathbf{w}^{\text{BD}}(t) = \sqrt{\frac{1}{2}} \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_1 \end{pmatrix} \quad (2.4)$$

$$\mathbf{w}^{\text{RS}}(t) = \frac{\begin{pmatrix} e^{\sigma^2 t} \mathbf{v}_1 \\ e^{\lambda_1 t} \epsilon \mathbf{v}_1 \end{pmatrix}}{(e^{2\sigma^2 t} + \epsilon e^{2\lambda_1 t})^{1/2}} \quad (2.5)$$

Comparing the MD and RS solutions (Equations 2.3 and 2.5) one sees that the times for the decay and recovery of neuronal activity must be identical for each regardless of noise levels and input statistics. This is consistent with experiment for the  $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$  ratio from experiment. It is not consistent with the  $\mathcal{T}_{\text{rise}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$  ratio, because it is observed that the recovery of the eye in RS takes longer than the loss of response to the closed eye in MD.

More striking, however, is Equation 2.4 which implies that during binocular deprivation, the weights perform a random walk about the normal reared state, and no significant loss of selectivity occurs. This is again inconsistent with the experiment results.

Also, looking at the exponentials in the MD and RS equations, we note that the *more noise* into the closed eye the *slower* the loss of response. This is important because, as it turns out, it is the opposite behavior from the BCM rule. This is critical, because it allows us to perform experiments to both test key assumptions about the interpretation of “noise” in the model, and also to distinguish between mechanisms predicted from these two learning rules. We will explore this in more detail in Chapter 3, where we can get an analytical understanding of this behavior for BCM. For now we explore, numerically, the properties of the BCM learning rule.

## 2.5.2 PCA: Summary of parameter dependence

The following observations were made

- for MD and RS, the times for the decay and recovery of neuronal activity are identical
- for BD, the weights perform a random walk about the normal reared state
- for MD and RS, the *more noise* into the closed eye, the *slower* the loss of response, which is qualitatively the opposite behavior from the BCM rule

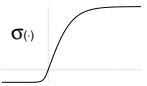
| Parameters for BCM Simulations |   |
|--------------------------------|---|
| Learning Rule                  | $\dot{\mathbf{w}} = \eta y(y - \theta)\mathbf{x}$   |
| Activation Rule                | $\dot{\theta} = \frac{1}{\tau}(y^2 - \theta)$   |
| cortical sigmoid               | $y = \sigma(\mathbf{w} \cdot \mathbf{x})$<br> $\begin{cases} \sigma(-\infty) = -1 \\ \sigma(+\infty) = 50 \end{cases}$ |
| Initial threshold              | $\theta_o = 0.73$   |
| Input mean                     | $\langle \mathbf{x} \rangle = 0$  |
| Input variance                 | $\text{var}(\mathbf{x}) = 1.0$  |
| RF Diameter                    | 13 pixels   |
| Retinal DOG ratio              | 3:1   |
| Learning rate                  | $\eta = 5 \cdot 10^{-7}, \dots, 5 \cdot 10^{-5}$  |
| Memory constant                | $\tau = 10, \dots, 3510$  |
| Noise Levels                   | uniform noise=<br>[-.25:.25],...,[-2.5,2.5]   |

Table 2.2: Parameters for BCM Simulations of Deprivation

### 2.5.3 BCM: Robustness to Parameters

In order to determine the dependence of the response half-times on model and input parameters we perform simulations over a range of those parameters and measure the time  $\mathcal{T}$  for each. These parameters are shown in Table 2.2. The parameter regime is initially chosen to give *stable* simulations, and then explored more finely to determine regions with consistent time ratios.

The main parameters we vary are the learning rate,  $\eta$ , the memory constant,  $\tau$ , and the noise level to the closed eye,  $\sigma^2$ . Other parameters which can significantly affect the timing include the RF size, the bottom value of the sigmoid on the output, the input variance and mean, and the form of the retinal preprocessing. The RF size simply increases the number of inputs, but not the overall statistics of the inputs (except for very large or very small RFs). This increase can be thought of as a simple increase in the learning rate. The input variance can be seen as a simple scaling of the learning rate and the threshold. The mean of the inputs, the form of the retinal preprocessing, and the lower limit of the sigmoid cannot be seen as a simple scaling of model parameters. We will see how some of these factors change some of the timing, but we will focus primarily on the three parameters  $\eta$ ,  $\tau$ , and  $\sigma^2$ . All of the discussion, unless otherwise specified, will have the parameter values from Table 2.2.

In order to determine the valid parameter regime, simulations had to be done at many points in the (learning rate)-(noise level)-(memory constant) space. Due to practical issues, the parameter space was not explored evenly: once a nearly valid range was found, parameters were tested in a finer grid. Therefore, on the results that follow, there are places where the data is more dense than in other places. Time values were averaged over several simulations.

### 2.5.4 BCM: Dependence on $\tau$

For each learning rate,  $\eta$ , there is a valid range for  $\tau$ . If  $\tau$  is too large, then oscillations occur (see Chapter 1) and the weights may not converge. Normal rearing is more sensitive to this than the various deprivation simulations. We would then choose the stability of normal rearing as the situation to set the upper bound on  $\tau$ . This upper bound, however, may be misleading because the stability of a simulation depends on the initial conditions, for which we have very little detailed knowledge in the case of normal rearing. If, for instance, there is a partially orientation selective receptive field *before* eye opening, the stability may be enhanced and we could use a larger value for  $\tau$ . We choose, then, to use the stability of the deprivation protocols to set the upper bound on  $\tau$ .

The lower bound for  $\tau$  is more difficult to set, because in many cases an instantaneously moving threshold ( $\tau \ll 1$ ) still converges. This, however, leads to rapid fluctuations of the threshold, which we do not believe is reasonable for the system to exhibit. When we convert the memory constant into the real time units of seconds (Section 2.5.7) we need to keep this in mind.

The dependence of the half-fall times,  $\mathcal{T}_{\text{fall}}^{\text{MD}}$ ,  $\mathcal{T}_{\text{fall}}^{\text{BD}}$ , and  $\mathcal{T}_{\text{fall}}^{\text{RS}}$  on the memory constant  $\tau$  within the stable range is shown in Figure 2.12. It is clear that there is little dependence of these half-times over a wide range of values of the memory constant. This is somewhat surprising, given that the memory constant determines the time scale over which the threshold averages the neural activity. What this suggests, from a mathematical point of view, is that there is limited sensitivity of the threshold to the averaging of the environment. Biologically, this suggests that the mechanisms behind the moving threshold need not have a constant time scale, as long as they are not too slow or too fast.

This may indicate a striking difference between the natural scene environment and the one dimensional oscillations results, where the memory constant played a more important role in the timing. This is not merely an effect of added dimensionality, because previous work modeling deprivation using higher dimensional inputs (Clothiaux et al., 1991) showed a strong dependence on the memory constant. There, the input environment was not taken from the natural scene images, but were high dimensional linearly independent vectors with added noise. We find here an indication that the *statistics* of the input environment has a dramatic effect on the behavior of the model. We will discuss this more in Chapter 3.

Another outcome of the lack of dependence on the memory constant,  $\tau$ , is that the deprivation experiments cannot give us any *detailed* information about this parameter. The range of valid  $\tau$  is determined completely by the stability of the simulations, and has no obvious measurable effect on the time course of deprivation. We can establish an upper bound, but that is all we can state with any degree of confidence. Thus, we have some freedom in establishing the mechanisms behind the moving threshold. The range that we do find is consistent with mechanisms such as gene expression and protein synthesis.

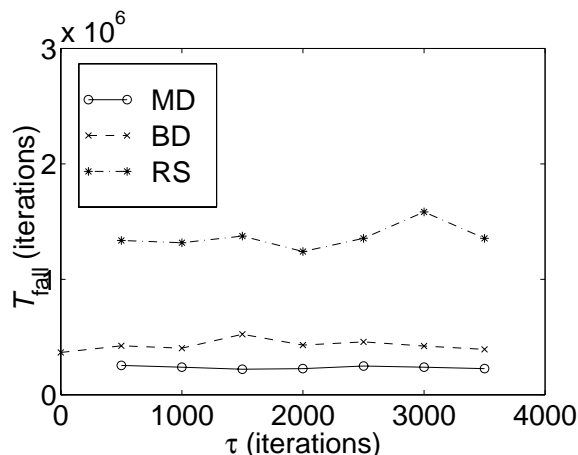


Figure 2.12: The dependence of the half-fall times,  $\mathcal{T}_{\text{fall}}^{\text{MD}}$ ,  $\mathcal{T}_{\text{fall}}^{\text{BD}}$ , and  $\mathcal{T}_{\text{fall}}^{\text{RS}}$  on the memory constant  $\tau$ .

### 2.5.5 BCM: Dependence on the noise

Figure 2.13 shows the dependence of the half-fall times,  $\mathcal{T}_{\text{fall}}^{\text{MD}}$ ,  $\mathcal{T}_{\text{fall}}^{\text{BD}}$ , and  $\mathcal{T}_{\text{fall}}^{\text{RS}}$  on the standard deviation of the noise,  $\sigma$ . There are two things to note here. First, the general behavior of the BCM learning rule on the noise into the closed eye. The *more noise* into the closed eye during any of the deprivation protocols, the *faster* the loss of response (smaller  $\mathcal{T}_{\text{fall}}$  time). We will explore this property in more detail in Chapter 3. The second thing to notice is that both reverse suture (RS) and binocular deprivation (BD) depend much more sharply on the noise than monocular deprivation (MD). We can understand why this might be the case for binocular deprivation, because noise is the only thing which is being presented to the cell. For RS we expect that, at first, it should have similar timings to BD. In both cases, the cell is essentially receiving noise into both eyes, perhaps with different variance. In BD we have noise entering both eyes, whereas in RS we have structured input entering an initially noisy receptive field and noise entering an initially oriented receptive field. Naturally we would expect a stronger dependence on the noise in these two cases.

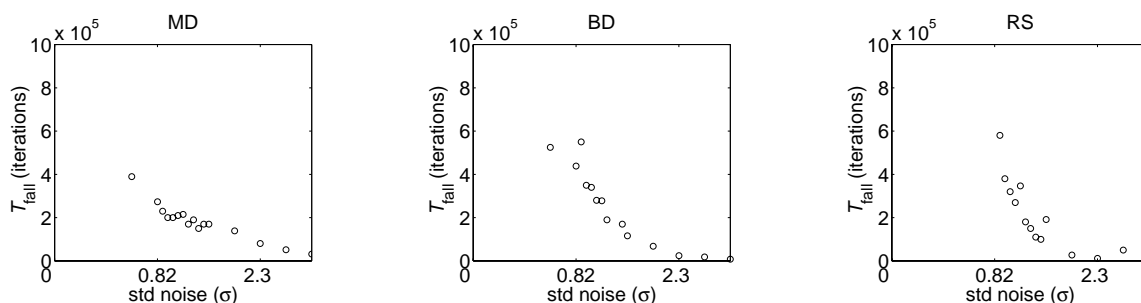


Figure 2.13: The dependence of the half-fall times,  $\mathcal{T}_{\text{fall}}^{\text{MD}}$ ,  $\mathcal{T}_{\text{fall}}^{\text{BD}}$ , and  $\mathcal{T}_{\text{fall}}^{\text{RS}}$  on the standard deviation of the noise,  $\sigma$ . ( $\eta = 5e - 6$ ,  $\tau = 1000$ ).

### 2.5.6 BCM: Finding a valid parameter regime

The basic approach of this section is shown schematically in Figure 2.14. Ranges of parameters are found which are consistent with the half-time ratios  $\mathcal{T}_{\text{fall}}^{\text{BD}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$ , and  $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$ . The overlap of these ranges specifies the valid parameter range, consistent with all three deprivation experiments. The time ratios involving  $\mathcal{T}_{\text{rise}}^{\text{RS}}$  are only used as a possible consistency check, and not to determine the valid range, because of problems discussed earlier.

Only the learning rate and the noise level are used, because the memory constant,  $\tau$ , has no appreciable effect on the half-times. The learning rate in the explored parameter regime had a lower value of  $\eta = 5 \cdot 10^{-7}$  due entirely to practical limitations: the simulations would take too long to run. It is likely that the valid parameter regime exists below this limit.

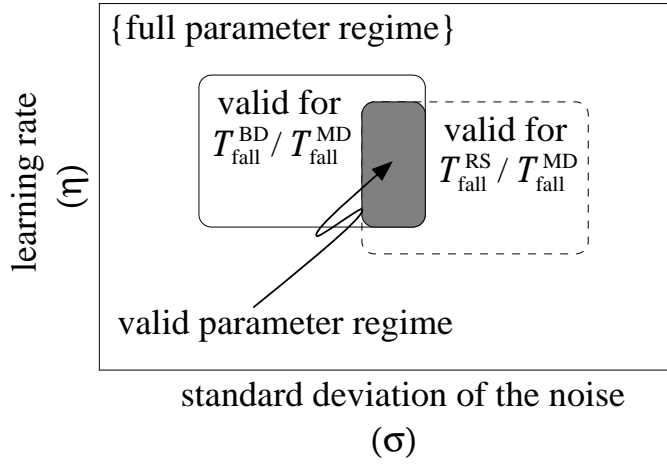


Figure 2.14: Schematic for finding a valid parameter regime. Areas in the parameter space are found which yield consistent half-time ratios,  $\mathcal{T}_{\text{fall}}^{\text{BD}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$ , and  $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$ . The time ratios involving  $\mathcal{T}_{\text{rise}}^{\text{RS}}$  are only used as a possible consistency check, and not to determine the valid range, because of problems discussed earlier. The memory constant,  $\tau$ , has no appreciable effect on the half-times so it is not used to determine the valid parameter regime.

Figure 2.15 shows the dependence of the half-fall time ratios  $\mathcal{T}_{\text{fall}}^{\text{BD}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$ , and  $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$  on the standard deviation of the noise,  $\sigma$ . For small noise,  $\mathcal{T}_{\text{fall}}^{\text{MD}}$  is faster than both  $\mathcal{T}_{\text{fall}}^{\text{BD}}$  and  $\mathcal{T}_{\text{fall}}^{\text{RS}}$ . This places an upper limit on the noise value of around  $\sigma \sim 1.4$ . Figure 2.16 shows the dependence on the learning rate,  $\eta$  for a *fixed* noise level. For smaller  $\eta$ , both ratios show a decrease, but the  $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$  ratio decrease is much larger: relative to MD, increasing the learning rate *speeds up* RS.

Finally, we can state that there is a parameter range around the parameter space point ( $\eta = 5 \cdot 10^{-6}, \sigma = 0.9$ ) which yields consistent time ratios compared to experiment. As a consistency check, Figure 2.17 shows the noise dependence on the ratios involving  $\mathcal{T}_{\text{rise}}^{\text{RS}}$ , which we stated before as being dubious. The results are still consistent, which gives us a little more confidence in them.

Finally, we can state that there is a range of parameters for BCM which over which the deprivation simulations are consistent with experiment. This range has an upper-bound on the learning rate at

$\eta = 4.5 \cdot 10^{-6}$ , and a possible lower bound at around  $\eta = 7 \cdot 10^{-6}$ . The lower bound may be extended using faster computers. For this range of learning rate, the upper bound on the value of  $\tau$  is determined to be  $\tau = 8500$  for stability with the deprivation experiments. As stated before, there is no strong lower bound on  $\tau$ . The value of the standard deviation of the noise is restricted to the range  $0.8 < \sigma < 1.4$ , in order to be consistent with experiment. This range is around the value 1, the standard deviation of the natural scene environment.

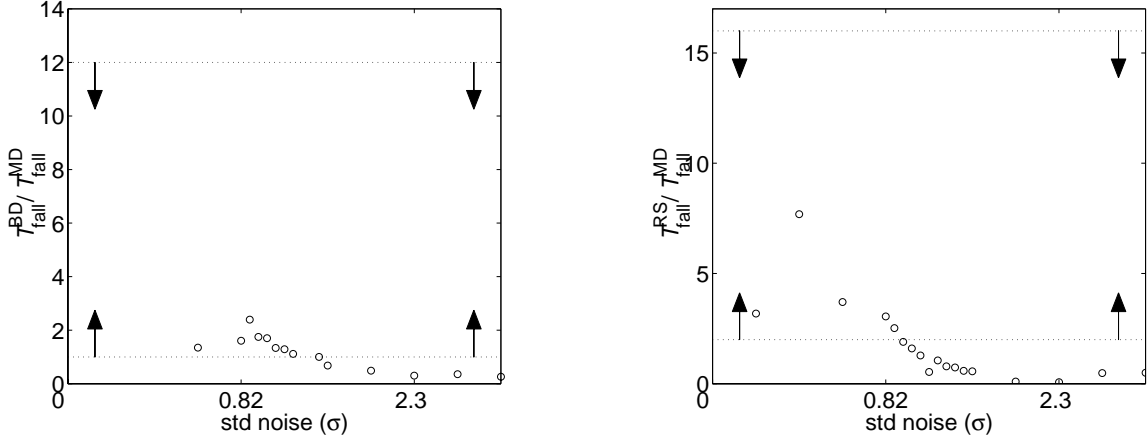


Figure 2.15: The dependence of the half-fall time ratios  $\mathcal{T}_{\text{fall}}^{\text{BD}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$ , and  $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$  on the standard deviation of the noise,  $\sigma$ . ( $\eta = 5e - 6$ ,  $\tau = 1000$ ). The valid ranges, from experiment, are shown with the arrows.

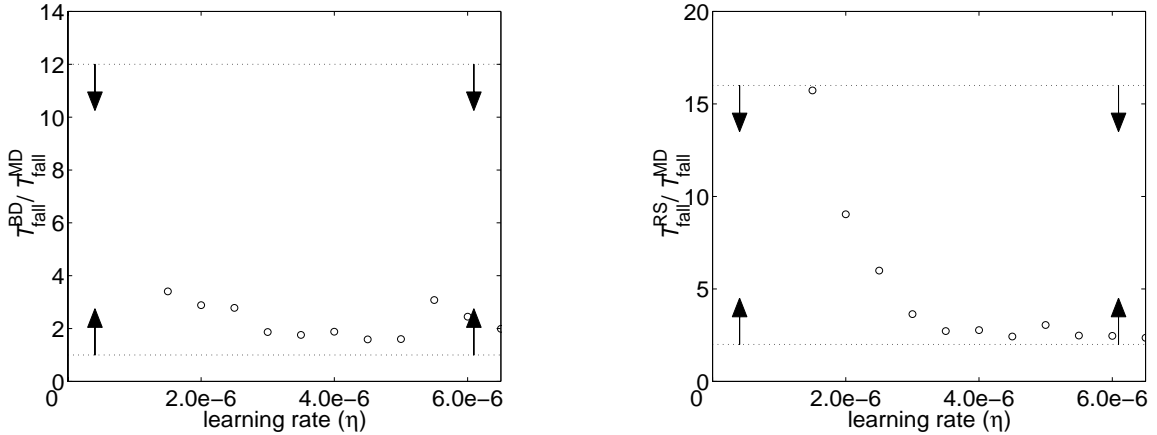


Figure 2.16: The dependence of the half-fall time ratios  $\mathcal{T}_{\text{fall}}^{\text{BD}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$ , and  $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$  on the learning rate,  $\eta$ . ( $\sigma = 0.8$ ,  $\tau = 1000$ ). The valid ranges, from experiment, are shown with the arrows.

### 2.5.7 BCM: Conclusions about parameter dependence

Given a valid parameter regime for the deprivation experiments, and half-times for those deprivation experiments within this range, we can obtain a conversion between the simulation time units (iterations)

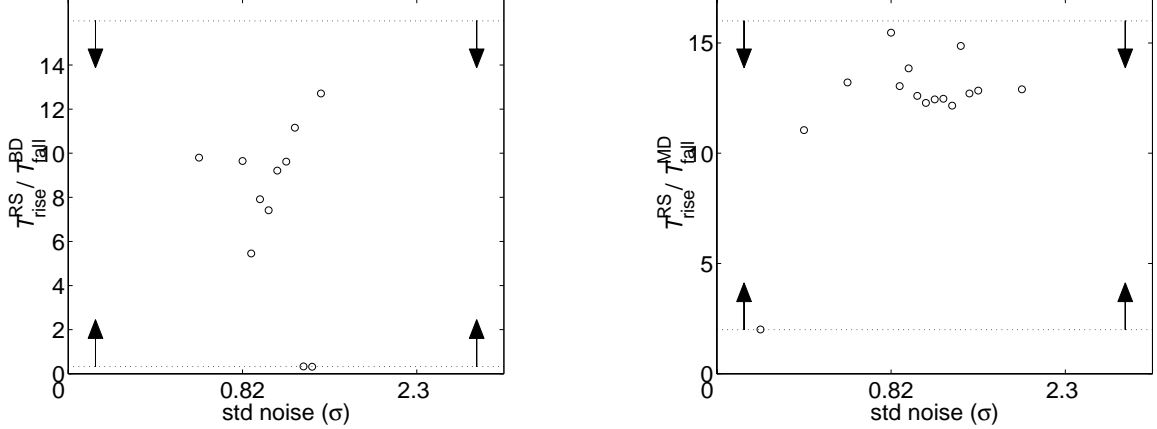


Figure 2.17: The dependence of the half-fall time ratios  $\mathcal{T}_{\text{rise}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{BD}}$ , and  $\mathcal{T}_{\text{rise}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$  on the standard deviation of the noise,  $\sigma$ . ( $\eta = 5e - 6$ ,  $\tau = 1000$ ). The valid ranges, from experiment, are shown with the arrows.

and the real time units (seconds). Monocular deprivation, within the parameter regime found, takes about  $2 \cdot 10^5$  iterations. In experiment, the half-time is around 12 hours. This makes one iteration about a 0.2 sec.

We then note that for monocular deprivation (and other deprivation protocols), using a noise standard deviation of  $\sigma = 0.8$ , the maximum consistent memory constant is as high as  $\tau \sim 8500$  iterations, which is around 30 minutes. We are not certain what mechanisms govern the motion of the threshold, but this time scale is consistent with protein synthesis which can occur on the order of minutes (Alberts et al., 1994). The upper limit set by normal rearing is  $\tau \sim 3500$  iterations, or 13 minutes, but it may be altered by the initial conditions which are not known precisely. The lower bound on  $\tau$  is quite small, on the order of a few iterations, which is equivalent to a second or more. Basically,  $\tau$  is on the order of minutes or a few seconds, but not on the order of milliseconds or hours. This places constraints on the possible mechanisms involved.

## 2.6 Summary of Comparison between BCM and PCA

Table 2.3 contains a summary of the comparison between the BCM and PCA rules. The first item of the table, the noise dependence of deprivation, is explored more in the next chapter. The other two items, the behavior under binocular deprivation and the relative timing of monocular deprivation and reverse suture, demonstrate how BCM is more consistent with experiment.

|                 | <b>BCM</b>   | <b>PCA</b>   | <b>Experiment</b>   |
|-----------------|--|--|---|
| <b>MD,BD,RS</b> | <i>more noise</i> into the closed eye the <i>faster</i> the loss of response | <i>more noise</i> into the closed eye the <i>slower</i> the loss of response | in MD, <i>more noise</i> into the closed eye the <i>faster</i> the loss of response (Rittenhouse et al., 1998)(see Section 3.5.1) |
| <b>BD</b>       | selectivity is lost, and responses are reduced                               | weights perform a random walk about the normal reared state                  | BD for 3 or 4 days, causes deterioration of the responses(Wiesel and Hubel, 1962; Freeman et al., 1981)                           |
| <b>MD,RS</b>    | times are different for the loss and recovery of response                    | times are identical for the loss and recovery of response                    | time for RS recovery of response longer than MD, while time for RS loss of response about the same as MD(Mioche and Singer, 1989) |

Table 2.3: Summary of the Comparison between BCM and PCA using the deprivation protocols monocular deprivation (MD), binocular deprivation (BD), and reverse suture (RS)

## 2.7 Other Models of Orientation Selectivity and Ocular Dominance

There are many different models of orientation selectivity and ocular dominance. There are some learning rules, such as the ICA rule of Bell and Sejnowski (1997) and the sparse coding algorithm of Olshausen and Field (1995), which have been used with natural scene inputs to produce oriented receptive fields. We do not include these in our comparison because the learning is not based on the activity and weights of a *single neuron*, and thus detracts from our immediate goal of comparing rules with the same single cell architecture. A previous model(Clothetaux et al., 1991) of deprivation, using the BCM rule, has been mentioned in Section 2.5.4. The largest difference from that work is our use of a more realistic input environment, but many of the qualitative results are the same, so we do not repeat them here.

Another large class of models has been proposed, referred to as *correlation-based models*(Linsker, 1986; MacKay and Miller, 1994; Miller, 1992; Miller et al., 1989; Erwin and Miller, 1995). They are correlation-based because they all share a common basis in the Hebb rule, which leads to finding eigenvectors of the correlation matrix (Section 1.3). Though some of the rules use networks of neurons, their behavior can be mostly understood from the single cell perspective. These models can be understood with arguments similar to those used for the PCA rule.

The correlation-based models all have a common set of assumptions. They are

- The environment is determined by a correlation function,  $\mathbf{C}$ . The form of the correlation function is specified by the modeler, and becomes a prediction for experiment. Sometimes this correlation function is broken up into separate components, like left eye and right eye correlation functions, or ON channel and OFF channel correlation functions. The correlation function is almost always assumed, for simplicity, to be radially symmetric and translationally invariant. This makes the correlation function only a function of the distance between spatial points in the environment,  $\mathbf{C}(|\mathbf{r} - \mathbf{r}'|)$ .
- The weights are explicitly bounded. The most common method is to use “sticky” bounds, where a weight growing past a bound is frozen at the value of the bound. Sometimes the weight bounds are set to be both negative and positive(Linsker, 1986), but it is more common to set zero as a lower bound, and some value,  $\mathbf{w}_{\max}$ , as an upper bound(MacKay and Miller, 1994; Miller, 1994).
- The weights change with a (possibly modified) Hebb rule. The local stochastic version of the Hebb rule, as we know, is

$$\frac{d\mathbf{w}}{dt} = y\mathbf{x} \quad (2.6)$$

with an averaged version (see Section 1.3) written as

$$\frac{d\mathbf{w}}{dt} = \mathbf{C}\mathbf{w} \quad (2.7)$$

This makes the learning correlation-based because, as we’ve seen, Hebb rules (like the PCA rule) have fixed points parallel to the eigenvectors of the correlation matrix. The modifications to the rule, as well as the sticky bounds, may cause the weights to become a mixture of the dominant eigenvectors(Erwin and Miller, 1995). Although it is convenient to use the averaged version of the learning rule, we must always make sure we return to the stochastic version in order to see if it is biologically plausible. A modification we make in the averaged version may not have a biologically realistic stochastic equivalent.

- A competition mechanism is explicitly added, in order to match the observed competition in such processes as monocular deprivation and others. This competition mechanism, in the form of a constraint, takes on two common forms: multiplicative and subtractive(MacKay and Miller, 1994). These can be written (in the averaged form)

$$\frac{d\mathbf{w}}{dt} = \mathbf{C}\mathbf{w} - \underbrace{\gamma(\mathbf{w})\mathbf{w}}_{\text{multiplicative}} - \underbrace{\epsilon(\mathbf{w})\mathbf{n}}_{\text{subtractive}} \quad (2.8)$$

where  $\mathbf{n}$  is a ones vector,  $(1, 1, \dots, 1)^T$ . In this terminology, the PCA rule has a multiplicative constraint, where  $\gamma(\mathbf{w}) = y^2$ . The use of these labels is not that informative, and almost misleading, because the functions  $\gamma(\mathbf{w})$  and  $\epsilon(\mathbf{w})$  can depend not only on the weight, but on the input and output values as well. To avoid confusion, whenever we use a constraint in describing one of these rules, we will make sure to rely more on what the constraint *does*, rather than what it is called.

### 2.7.1 A Toy Model

Most of the work done with the correlation-based models is done on networks of neurons. The single cell receptive fields are limited to a particular area of the visual field with an arbor function, and the network interaction is given by a lateral interaction function. While both of these are biologically plausible, they complicate the picture and make it more difficult to see what the neurons are actually doing. To make this easier, we present a toy model which captures many of the aspects of more complicated models.

Since the correlation function is often taken to be radially symmetric and translationally invariant, it is a function in one dimension: the distance between points in the visual field. The arbor function, restricting the receptive field to an local area in space, imposes a two dimensional structure on the correlation function as we shall see. Initially, however, we will take a one dimensional arbor function, for instructional purposes. Because this is a fully one dimensional problem, an oriented receptive field has no real meaning. There is, however, an analogue to the oriented receptive field in one dimension.

Let us say that our receptive field is some function of the position in one dimension,<sup>2</sup>  $\mathbf{w}(x)$ . This is a continuous generalization of the discrete weight vectors we usually use. Let's say that the arbor function is given by

$$A(x) = \begin{cases} 1 & \text{if } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.9)$$

which simply means that the visual field is restricted between the positions  $x = -1$  and  $x = 1$ . Let us further assume that the correlation function<sup>3</sup> is given by

$$\mathbf{C}(x - x') = 1 + q \cos(\pi(x - x')) \quad (2.10)$$

This correlation function is shown in Figure 2.18. The correlation-based rules find the dominant solution of the eigenvalue equation

$$\int_{-\infty}^{+\infty} A(x') \mathbf{C}(x - x') \mathbf{w}(x') dx' = \lambda \mathbf{w}(x') \quad (2.11)$$

$$= \int_{-1}^{+1} [1 + q \cos(\pi(x - x'))] \mathbf{w}(x') dx' \quad (2.12)$$

We can expand the  $\cos(\cdot)$  in the correlation function,

$$\mathbf{C}(x - x') = 1 + q \cos(\pi(x - x')) = 1 + q \cos(\pi x) \cos(\pi x') + q \sin(\pi x) \sin(\pi x')$$

expand the weights in a Fourier series,

$$\mathbf{w}(x) = a_o + \sum_{l=1}^{\infty} a_l \cos(\pi l x) + \sum_{l=1}^{\infty} b_l \sin(\pi l x)$$

---

<sup>2</sup>We are using the variable  $x$  here to denote position in one-dimension, which is not to be mistaken for the input of the neuron. Since we are dealing with equations averaged over the environment, we need never refer to the input directly, so no confusion should arise.

<sup>3</sup>or pseudo-correlation function

use the orthogonality of sines and cosines,

$$\begin{aligned}\int_{-1}^{+1} \sin(k\pi x') \sin(n\pi x') dx' &= \delta_{kn} \\ \int_{-1}^{+1} \cos(k\pi x') \cos(n\pi x') dx' &= \delta_{kn}\end{aligned}$$

and obtain the following equation

$$2a_o + qa_1 \cos(\pi x) + qb_1 \sin(\pi x) = \lambda \left( a_o + \sum_{l=1}^{\infty} a_l \cos(\pi l x) + \sum_{l=1}^{\infty} b_l \sin(\pi l x) \right) \quad (2.13)$$

For  $l > 1$ , the coefficients  $a_l$  and  $b_l$  are all zero. There are, thus, three solutions:  $\mathbf{w}(x) = \text{const}$ , with  $\lambda = 2$ ,  $\mathbf{w}(x) = \sin(\pi x)$  and  $\mathbf{w}(x) = \cos(\pi x)$ , both with  $\lambda = q$ .

It is clear that, if  $q$  is changed (which is the same as changing the visual environment), the dominant solution can switch between the DC solution ( $\mathbf{w}(x) = \text{const}$ ) to one of the AC solutions ( $\mathbf{w}(x) = \cos(\pi x)$  or  $\sin(\pi x)$ ), as shown in Figure 2.18. If we change the shape, or form, of the arbor function, the dominant eigenvectors can change as well. Introducing a network, with lateral inhibition, will also change which eigenvector of the correlation matrix is dominant. The AC solutions are the analog of oriented receptive fields, because they are the only ones in this model with structure. With the correlation based rules, the correlation function is chosen so that, say, orientation selective receptive fields are among the *possible* dominant eigenvectors and the rest of the details (subtractive vs. multiplicative constraints, arbor function shape, lateral inhibition, etc.) are chosen to make those oriented receptive fields the dominant ones.

## 2.7.2 Correlation-based Model of Orientation Selectivity

A correlation-based model of orientation selectivity has been proposed, using the competition between ON- and OFF-center inputs (Miller, 1994). As before, we write down an equation for the change in the weights, using a correlation function that is radially symmetric and translationally invariant. This time the correlation function is broken into ON and OFF channels.

$$\frac{d}{dt} \begin{pmatrix} \mathbf{w}^{\text{ON}} \\ \mathbf{w}^{\text{OFF}} \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{\text{ON,ON}} & \mathbf{C}^{\text{ON,OFF}} \\ \mathbf{C}^{\text{OFF,ON}} & \mathbf{C}^{\text{OFF,OFF}} \end{pmatrix} \begin{pmatrix} \mathbf{w}^{\text{ON}} \\ \mathbf{w}^{\text{OFF}} \end{pmatrix} \quad (2.14)$$

It is then assumed that the two same-channel correlation functions,  $\mathbf{C}^{\text{ON,ON}}$  and  $\mathbf{C}^{\text{OFF,OFF}}$ , are equal and the two opposite-channel correlation functions are equal. This leaves us with

$$\frac{d}{dt} \begin{pmatrix} \mathbf{w}^{\text{ON}} \\ \mathbf{w}^{\text{OFF}} \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{\text{same}} & \mathbf{C}^{\text{opposite}} \\ \mathbf{C}^{\text{opposite}} & \mathbf{C}^{\text{same}} \end{pmatrix} \begin{pmatrix} \mathbf{w}^{\text{ON}} \\ \mathbf{w}^{\text{OFF}} \end{pmatrix} \quad (2.15)$$

The correlation functions are chosen as difference of Gaussians, as follows

$$\mathbf{C}^{\text{same}}(r) = \exp(-r^2/\sigma^2) - \frac{1}{9} \exp(-r^2/(9\sigma^2)) \quad (2.16)$$

$$\mathbf{C}^{\text{opposite}}(r) = -0.5\mathbf{C}^{\text{same}}(r) \quad (2.17)$$

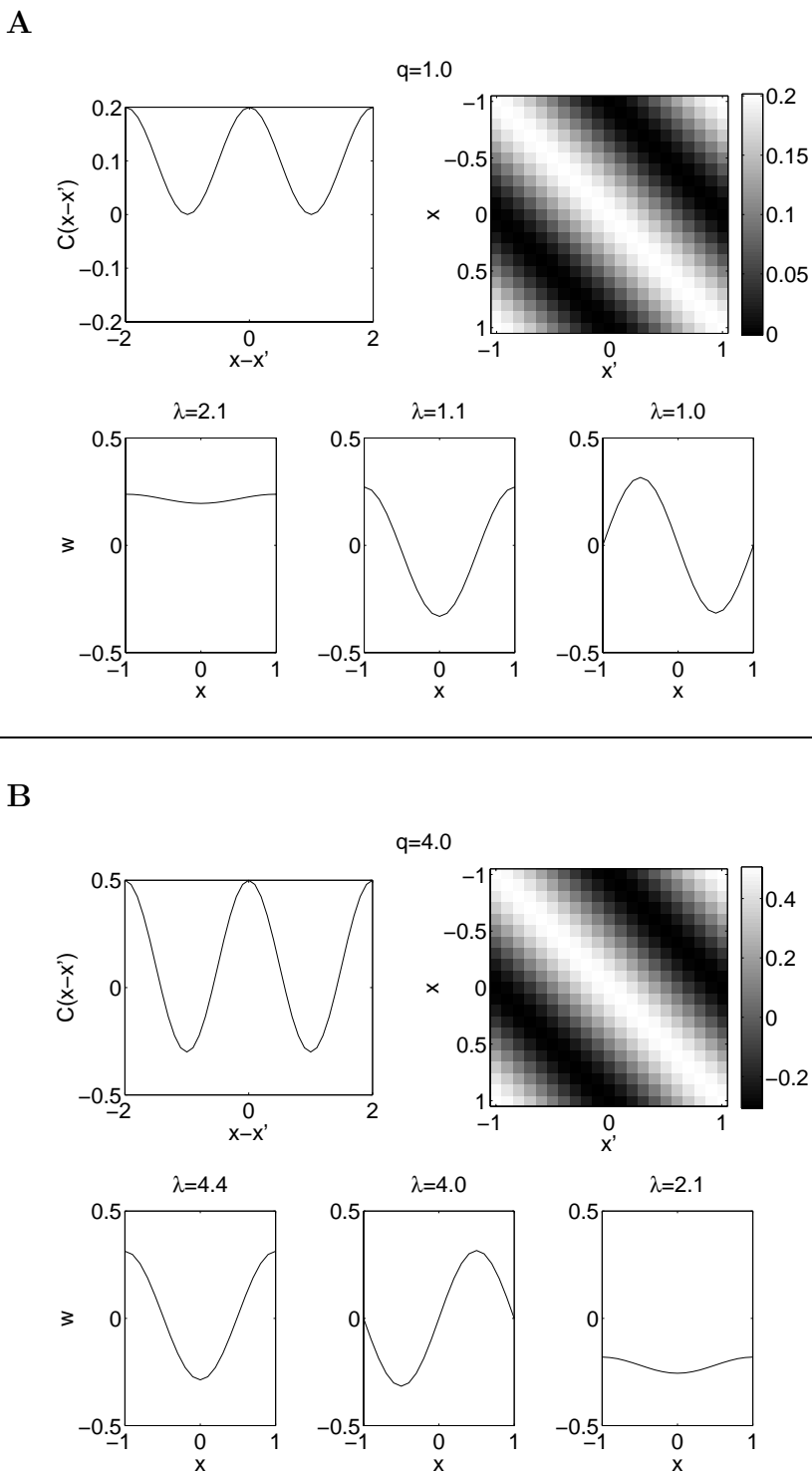


Figure 2.18: Toy model correlation functions and eigenvectors. Shown is the toy correlation function,  $C(x - x') = 1 + q \cos(\pi(x - x'))$ , as a one dimensional function (upper left, both **A** and **B**) and as a two dimensional function (upper right, both **A** and **B**). The three eigenvectors are shown in the lower plots of **A** and **B**, with the corresponding eigenvalues,  $\lambda$ . Plots in **A** are for  $q = 1$ , where the DC solution is dominant. Plots in **B** are for  $q = 4$ , where the AC solution is dominant. Small deviations from the expected values of  $\lambda$ , and the small non-DC parts of the DC solution are merely numerical in origin, and should be ignored.

This means that there are short range correlations, and long range anti-correlations within each eye and anti-correlations between the eyes. The anti-correlations between the eyes is certainly not biologically plausible for normal vision, where the eyes see nearly the same thing. The models are often proposed to explain the pre-eye-opening formation of orientation selectivity. It is unknown what the correlation functions are in this case, but we predict from this model that the anti-correlations are *necessary*.

We can diagonalize the correlation matrix, and transform to the diagonal basis, which is simply

$$\mathbf{w}^S \equiv \mathbf{w}^{\text{ON}} + \mathbf{w}^{\text{OFF}} \quad (2.18)$$

$$\mathbf{w}^D \equiv \mathbf{w}^{\text{ON}} - \mathbf{w}^{\text{OFF}} \quad (2.19)$$

leaving us with

$$\frac{d}{dt} \begin{pmatrix} \mathbf{w}^S \\ \mathbf{w}^D \end{pmatrix} = \begin{pmatrix} \mathbf{C}^S & 0 \\ 0 & \mathbf{C}^D \end{pmatrix} \begin{pmatrix} \mathbf{w}^S \\ \mathbf{w}^D \end{pmatrix} \quad (2.20)$$

We also need to include an arbor function, which restricts the range of the receptive field to some radius. Without all of the complications presented in Miller (1994), we can calculate numerically the eigenvectors of  $\mathbf{C}^S$  and  $\mathbf{C}^D$ , the sum and difference correlation matrix. This is shown in Figure 2.19

Note that the eigenvalues for  $\mathbf{C}^D$  are much larger than for  $\mathbf{C}^S$ , and the eigenvectors are oriented. This means that the full receptive field is both oriented, and opposite for the ON and OFF channels, which is consistent with experiment (Reid and Alonso, 1995). Also notice that the eigenvalues for the oriented solutions are *very* close to the non-oriented solutions. Small changes in the arbor function, lateral connections, shape of the correlation function, etc. can make the dominant solution not oriented.

### 2.7.3 Correlation-based Model of Monocular Deprivation

A model of monocular deprivation was proposed (Miller and Stryker, 1990) in a similar way to the model of orientation selectivity presented previously. The correlation functions for ocular dominance were given (as opposed to orientation selectivity), and monocular deprivation was modeled using a closed eye correlation function which was a simple *scaling* of the normal open eye correlation function. This implies that monocular deprivation is simply a *variance change*, which is inconsistent with experiment (Blakemore, 1976; Wiesel and Hubel, 1965). Since this model ignored experimental data from the outset, we do not consider it further.

### 2.7.4 Problems with Correlation-based Models

There are many problems with the correlation-based models. Some of them are fundamental problems, and some of them are problems with *particular* implementations. We summarize a few here.

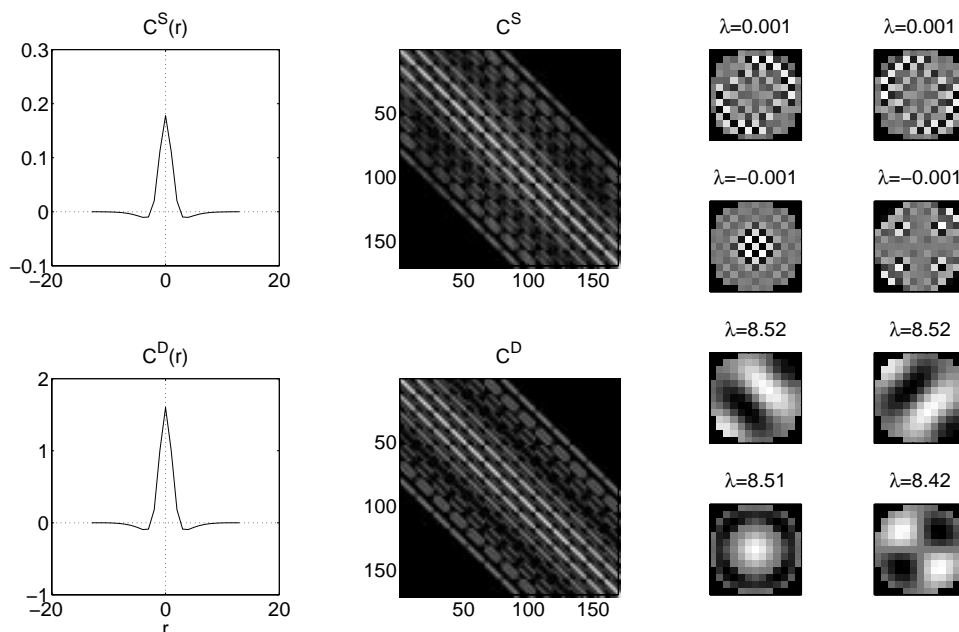


Figure 2.19: Correlation-based model for orientation selectivity. Shown are the sum and difference correlation functions,  $\mathbf{C}^S$  and  $\mathbf{C}^D$  respectively, for a model of ON and OFF channel inputs, as a function of the distance (upper and lower left plots), and in a matrix form (upper and lower center graphs) where the four dimensions have been collapsed onto two. The banded structure comes both from the arbor function and the correlation function restricting the extent of the receptive field. The four receptive fields on the upper right are the dominant eigenvectors of  $\mathbf{C}^S$  with labeled eigenvalues. The four receptive fields on the lower right are the dominant eigenvectors of  $\mathbf{C}^D$ , also with labeled eigenvalues.

- In these models, the natural environment is determined by only second order statistics. This has been shown to be incomplete(Olshausen and Field, 1996b) at best. Also, since monocular deprivation has been shown to depend on *patterned* input, not just variance, then two point correlations are simply not enough.
- As shown in Section 2.5.1, PCA-like rules behave very unrealistically under binocular deprivation conditions, and somewhat unrealistically under reverse suture conditions.
- The orientation selectivity found in these models is not very robust, as demonstrated in the previous section. Changes in the upper weight cutoff, lateral connections, arbor function, etc. have significant effects.
- The subtractive constraints that have been used in many of the results cited, require a possibly non-local calculation, as we discussed earlier (Section 1.1.4). The subtractive constraint consists of subtracting off the total weight change at each iteration, thereby maintaining a constant total weight (or weight resources, as it is often phrased). The problem with this is that, essentially, the sum of the weights needs to be calculated *every iteration* in order to enforce such a constraint. This places severe restrictions on the types of processes possible. The sum of the weights would have to be calculated (which means that the weights would have to be coded in such a way that their sum is available to the cell), and this sum would have to be communicated to all the synapses, despite the large distances between them, all on the order of milliseconds. Though this is *possible*, we do not think it is likely and the burden of proof lies firmly with the ones proposing such a constraint.
- We find in the next chapter that the noise dependence we saw for PCA is inconsistent with experiment(Rittenhouse et al., 1998).

These problems, and others, make us lean away from strictly correlation-based models.

## 2.8 Conclusions

We have presented a model of the initial visual system, which captures many of the qualitative features that we observe from experiment. We have further used the results from experiment to quantitatively compare to the model, in the form of ratios of times for the loss or gain of neuronal response during deprivation protocols. The initial comparison has pointed out some inconsistencies with predictions from the PCA rule. It has also pointed out qualitatively different noise dependence for the PCA and BCM rules, which we will explore in more detail in the next chapter.

Our goal of obtaining a parameter regime consistent with *all* of the deprivation experiments, using the BCM rule in the natural scene environment, was successful. The restriction on the memory constant,  $\tau$ , which gives the time scale over which the threshold is averaged, was shown to be primarily

due to stability, and not from an effect on the time course of deprivation. The time scales found for this parameter are consistent with some of the mechanisms, such as gene expression, thought to be behind the moving threshold.