Chapter 2

Neural Encoding II: Reverse Correlation and Visual Receptive Fields

2.1 Introduction

The spike-triggered average stimulus introduced in chapter 1 is a standard way of characterizing the selectivity of a neuron. In this chapter, we show how spike-triggered averages and reverse-correlation techniques can be used to construct estimates of firing rates evoked by arbitrary time-dependent stimuli. Firing rates calculated directly from reverse-correlation functions provide only a linear estimate of the response of a neuron, but we also present in this chapter various methods for including nonlinear effects such as firing thresholds.

retina
LGN
V1, area 17

Spike-triggered averages and reverse-correlation techniques have been used extensively to study properties of visually responsive neurons in the retina (retinal ganglion cells), lateral geniculate nucleus (LGN), and primary visual cortex (V1, or area 17 in the cat). At these early stages of visual processing, the responses of some neurons (simple cells in primary visual cortex, for example) can be described quite accurately using this approach. Other neurons (complex cells in primary visual cortex, for example) can be described by extending the formalism. Reverse-correlation techniques have also been applied to responses of neurons in visual areas V2, area 18, and MT, but they generally fail to capture the more complex and nonlinear features typical of responses at later stages of the visual system. Descriptions of visual responses based on reverse correlation are approximate, and they do not explain how visual responses arise from the synaptic, cellular, and network properties of retinal, LGN, and cortical circuits. Nevertheless, they provide a important framework for character-

izing response selectivities, a reference point for identifying and characterizing novel effects, and a basis for building mechanistic models, some of which are discussed at the end of this chapter and in chapter 7.

2.2 Estimating Firing Rates

In chapter 1, we discussed a simple model in which firing rates were estimated as instantaneous functions of the stimulus, using response tuning curves. The activity of a neuron at time t typically depends on the behavior of the stimulus over a period of time starting a few hundred milliseconds prior to t and ending perhaps tens of milliseconds before t. Reverse correlation methods can be used to construct a more accurate model that includes the effects of the stimulus over such an extended period of time. The basic problem is to construct an estimate $r_{est}(t)$ of the firing rate $r(t)$ evoked by a stimulus $s(t)$. The simplest way to construct an estimate is to assume that the firing rate at any given time can be expressed as a weighted sum of the values taken by the stimulus at earlier times. Since time is a continuous variable this 'sum' actually takes the form of an integral, and we write

firing rate estimate $r_{est}(t)$

$$r_{\rm est}(t) = r_0 + \int_0^\infty d\tau\, D(\tau) s(t-\tau)\,. \tag{2.1}$$

The term r_0 accounts for any background firing that may occur when $s = 0$. $D(\tau)$ is a weighting factor that determines how strongly, and with what sign, the value of the stimulus at time $t-\tau$ affects the firing rate at time t. Note that the integral in equation 2.1 is a linear filter of the same form as the expressions used to compute $r_{\rm approx}(t)$ in chapter 1.

As discussed in chapter 1, sensory systems tend to adapt to the absolute intensity of a stimulus. It is easier to account for the responses to fluctuations of a stimulus around some mean background level than it is is to account for adaptation processes. We therefore assume throughout this chapter that the stimulus parameter $s(t)$ has been defined with its mean value subtracted out. This means that the time integral of $s(t)$ over the duration of a trial is zero.

We have provided a heuristic justification for the terms in equation 2.1 but, more formally, they correspond to the first two terms in a systematic expansion of the response in powers of the stimulus. Such an expansion is the functional equivalent of the Taylor series expansion used to generate power series approximations of functions, and it is called the Volterra expansion. For the case we are considering, it takes the form

Volterra expansion

$$\begin{aligned} r_{\rm est}(t) \;=\; & r_0 + \int d\tau\, D(\tau) s(t-\tau) + \int d\tau_1 d\tau_2\, D_2(\tau_1,\tau_2) s(t-\tau_1) s(t-\tau_2) + \\ & \int d\tau_1 d\tau_2 d\tau_3\, D_3(\tau_1,\tau_2,\tau_3) s(t-\tau_1) s(t-\tau_2) s(t-\tau_3) + \ldots\,. \end{aligned} \tag{2.2}$$

This series was rearranged by Wiener to make the terms easier to compute. The first two terms of the Volterra and Wiener expansions are identical and are given by the two expressions on the right side of equation 2.1. For this reason, D is called the first Wiener kernel, the linear kernel, or, when higher-order terms (terms involving more than one factor of the stimulus) are not being considered, simply the kernel.

Wiener expansion

Wiener kernel

To construct an estimate of the firing rate based on an expression of the form 2.1, we choose the kernel D to minimize the squared difference between the estimated response to a stimulus and the actual measured response averaged over time,

$$E = \frac{1}{T}\int_0^T dt\ (\mathrm{r}_{\mathrm{est}}(t) - \mathrm{r}(t))^2\ . \tag{2.3}$$

This expression can be minimized by setting its derivative with respect to the function D to zero (see appendix A). The result is that D satisfies an equation involving two quantities introduced in chapter 1, the firing rate-stimulus correlation function, $Q_{\mathrm{rs}}(\tau) = \int dt\, r(t)\ s(t+\tau)/T$, and the stimulus autocorrelation function, $Q_{ss}(\tau) = \int dt\, s(t)\ s(t+\tau)/T$,

optimal kernel

$$\int_0^\infty d\tau'\, Q_{ss}(\tau - \tau')D(\tau') = Q_{\mathrm{rs}}(-\tau)\,. \tag{2.4}$$

The method we are describing is called reverse correlation because the firing rate–stimulus correlation function is evaluated at $-\tau$ in this equation.

Equation 2.4 can be solved most easily if the stimulus is white noise, although it can be solved in the general case as well (see appendix A). For a white-noise stimulus $Q_{ss}(\tau) = \sigma_s^2\delta(\tau)$ (see chapter 1), so the left side of equation 2.4 is

$$\sigma_s^2\int_0^\infty d\tau'\,\delta(\tau - \tau')D(\tau') = \sigma_s^2 D(\tau)\,. \tag{2.5}$$

As a result, the kernel that provides the best linear estimate of the firing rate is

white-noise kernel

$$D(\tau) = \frac{Q_{\mathrm{rs}}(-\tau)}{\sigma_s^2} = \frac{\langle r\rangle C(\tau)}{\sigma_s^2} \tag{2.6}$$

where $C(\tau)$ is the spike-triggered average stimulus, and $\langle r\rangle$ is the average firing rate of the neuron. For the second equality, we have used the relation $Q_{\mathrm{rs}}(-\tau) = \langle r\rangle C(\tau)$ from chapter 1. Based on this result, the standard method used to determine the optimal kernel is to measure the spike-triggered average stimulus in response to a white-noise stimulus.

In chapter 1, we introduce the H1 neuron of the fly visual system, which responds to moving images. Figure 2.1 shows a prediction of the firing rate of this neuron obtained from a linear filter. The velocity of the moving image is plotted in 2.1A, and two typical responses are shown in 2.1B.

The firing rate predicted from a linear estimator, as discussed above, and the firing rate computed from the data by binning and counting spikes are compared in Figure 2.1C. The agreement is good in regions where the measured rate varies slowly but the estimate fails to capture high-frequency fluctuations of the firing rate, presumably because of nonlinear effects not captured by the linear kernel. Some such effects can be described by a static nonlinear function, as discussed below. Others may require including higher-order terms in a Volterra or Wiener expansion.

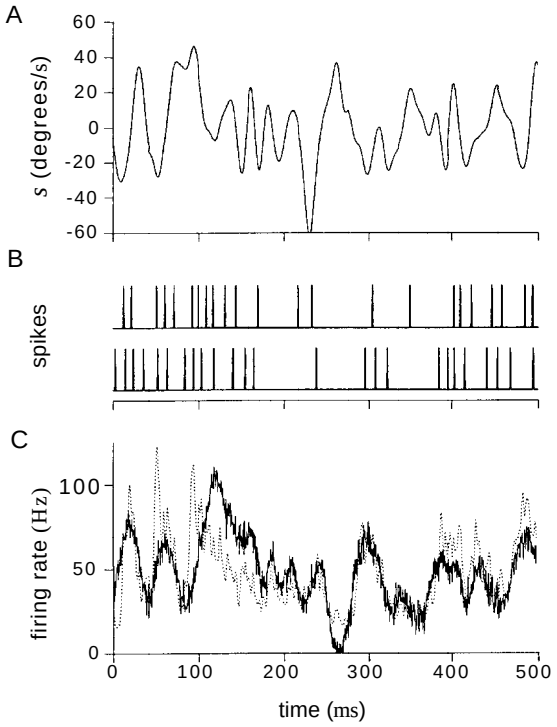

Figure 2.1: Prediction of the firing rate for an H1 neuron responding to a moving visual image. A) The velocity of the image used to stimulate the neuron. B) Two of the 100 spike sequences used in this experiment. C) Comparison of the measured and computed firing rates. The dashed line is the firing rate extracted directly from the spike trains. The solid line is an estimate of the firing rate constructed by linearly filtering the stimulus with an optimal kernel. (Adapted from Rieke et al., 1997.)

The Most Effective Stimulus

Neuronal selectivity is often characterized by describing stimuli that evoke maximal responses. The reverse-correlation approach provides a justification for this procedure by relating the optimal kernel for firing rate estimation to the stimulus predicted to evoke the maximum firing rate, subject

to a constraint. A constraint is essential because the linear estimate 2.1 is unbounded. The constraint we use is that the time integral of the square of the stimulus over the duration of the trial is held fixed. We call this integral the stimulus energy. The stimulus for which equation 2.1 predicts the maximum response at some fixed time subject to this constraint, is computed in appendix B. The result is that the stimulus producing the maximum response is proportional to the optimal linear kernel, or equivalently to the white-noise spike-triggered average stimulus. This is an important result because in cases where a white-noise analysis has not been done, we may still have some idea what stimulus produces the maximum response.

The maximum stimulus analysis provides an intuitive interpretation of the linear estimate of equation 2.1. At fixed stimulus energy, the integral in 2.1 measures the overlap between the actual stimulus and the most effective stimulus. In other words, it indicates how well the actual stimulus matches the most effective stimulus. Mismatches between these two reduce the value of the integral and result in lower predictions for the firing rate.

Static Nonlinearities

The optimal kernel produces an estimate of the firing rate that is a linear function of the stimulus. Neurons and nervous systems are nonlinear, so a linear estimate is only an approximation, albeit a useful one. The linear prediction has two obvious problems: there is nothing to prevent the predicted firing rate from becoming negative, and the predicted rate does not saturate, but instead increases without bound as the magnitude of the stimulus increases. One way to deal with these and some of the other deficiencies of a linear prediction is to write the firing rate as a background rate plus a nonlinear function of the linearly filtered stimulus. We use L to represent the linear term we have been discussing thus far,

$$L(t) = \int_0^\infty d\tau D(\tau) s(t-\tau)\,. \tag{2.7}$$

$r_{est}(t)$ *with static nonlinearity*

The modification is to replace the linear prediction $r_{est}(t) = r_0 + L(t)$ by the generalization

$$r_{est}(t) = r_0 + F(L(t)) \tag{2.8}$$

where F is an arbitrary function. F is called a static nonlinearity to stress that it is a function of the linear filter value evaluated instantaneously at the time of the rate estimation. If F is appropriately bounded from above and below, the estimated firing rate will never be negative or unrealistically large.

F can be extracted from data by means of the graphical procedure illustrated in figure 2.2A. First, a linear estimate of the firing rate is computed

using the optimal kernel defined by equation 2.4. Next a plot is made of the pairs of points $(L(t), r(t))$ at various times and for various stimuli, where $r(t)$ is the actual rate extracted from the data. There will be a certain amount of scatter in this plot due to the inaccuracy of the estimation. If the scatter is not too large, however, the points should fall along a curve, and this curve is a plot of the function $F(L)$. It can be extracted by fitting a function to the points on the scatter plot. The function F typically con-

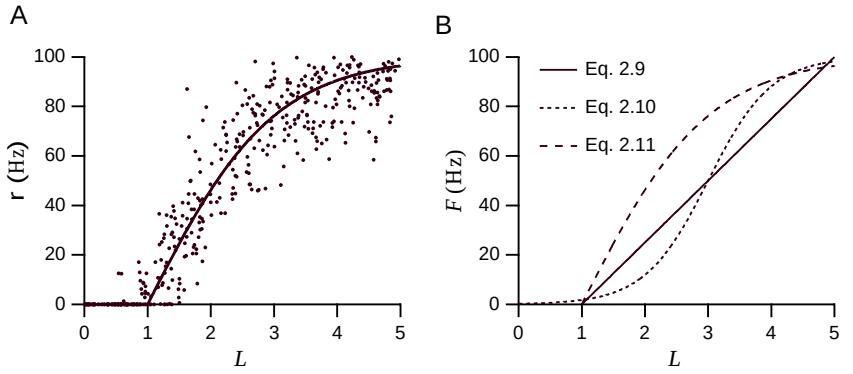

Figure 2.2: A) A graphical procedure for determining static nonlinearities. The linear estimate L and the actual firing rate r are plotted (solid points) and fit by the function $F(L)$ (solid line). B) Different static nonlinearities used in estimating neural responses. L is dimensionless, and equations 2.9, 2.10, and 2.10 have been used with $G = 25$ Hz, $L_0 = 1$, $L_{1/2} = 3$, $r_{\max} = 100$ Hz, $g_1 = 2$, and $g_2 = 1/2$.

tains constants used to set the firing rate to realistic values. These give us the freedom to normalize $D(\tau)$ in some convenient way, correcting for the arbitrary normalization by adjusting the parameters within F.

threshold function

Static nonlinearities are used to introduce both firing thresholds and saturation into estimates of neural responses. Thresholds can be described by writing

$$F(L) = G[L - L_0]_+ \tag{2.9}$$

where L_0 is the threshold value that L must attain before firing begins. Above the threshold, the firing rate is a linear function of L, with G acting as the constant of proportionality. Half-wave rectification is a special case of this with $L_0 = 0$. That this function does not saturate is not a problem if large stimulus values are avoided. If needed, a saturating nonlinearity can be included in F, and a sigmoidal function is often used for this purpose,

rectification

sigmoid function

$$F(L) = \frac{r_{\max}}{1 + \exp\left(g_1(L_{1/2} - L)\right)}\,. \tag{2.10}$$

Here $r_{\max}$ is the maximum possible firing rate, $L_{1/2}$ is the value of L for which F achieves half of this maximal value, and g_1 determines how rapidly the firing rate increases as a function of L. Another choice that

combines a hard threshold with saturation uses a rectified hyperbolic tangent function,

$$F(L) = r_{max}[\tanh\left(g_2(L - L_0)\right)]_+ \tag{2.11}$$

where r_{max} and g_2 play similar roles as in equation 2.10, and L_0 is the threshold. Figure 2.2B shows the different nonlinear functions that we have discussed.

Although the static nonlinearity can be any function, the estimate of equation 2.8 is still restrictive because it allows for no dependence on weighted autocorrelations of the stimulus or other higher-order terms in the Volterra series. Furthermore, once the static nonlinearity is introduced, the linear kernel derived from equation 2.4 is no longer optimal because it was chosen to minimize the squared error of the linear estimate $r_{est}(t) = L(t)$, not the estimate with the static nonlinearity $r_{est}(t) = F(L(t))$. A theorem due to Bussgang (see appendix C) suggests that equation 2.6 will provide a reasonable kernel, even in the presence of a static nonlinearity, if the white noise stimulus used is Gaussian.

In some cases, the linear term of the Volterra series fails to predict the response even when static nonlinearities are included. Systematic improvements can be attempted by including more terms in the Volterra or Wiener series, but in practice it is quite difficult to go beyond the first few terms. The accuracy with which the first term, or first few terms, in a Volterra series can predict the responses of a neuron can sometimes be improved by replacing the parameter s in equation 2.7 by an appropriately chosen function of s, so that

$$L(t) = \int_0^\infty d\tau D(\tau) f(s(t-\tau)) \,. \tag{2.12}$$

A reasonable choice for this function is the response tuning curve. With this choice, the linear prediction is equal to the response tuning curve, $L = f(s)$, for static stimuli provided that the integral of the kernel D is equal to one. For time-dependent stimuli, we can think of equation 2.12 as a dynamic extension of the response tuning curve.

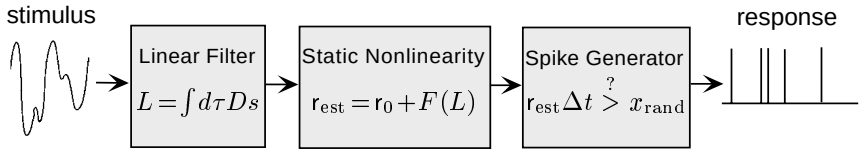

Figure 2.3: Simulating spiking responses to stimuli. The integral of the stimulus s times the optimal kernel D is first computed. The estimated firing rate is the background rate r_0 plus a nonlinear function of the output of the linear filter calculation. Finally, the estimated firing rate is used to drive a Poisson process that generates spikes.

A model of the spike trains evoked by a stimulus can be constructed by using the firing rate estimate of equation 2.8 to drive a Poisson spike generator (see chapter 1). Figure 2.3 shows the structure of such a model with a linear filter, a static nonlinearity, and a stochastic spike-generator. In the figure, spikes are shown being generated by comparing the spiking probability $r(t)\Delta t$ to a random number, although the other methods discussed in chapter 1 could be used instead. Also, the linear filter acts directly on the stimulus s in figure 2.3, but it could act instead on some function $f(s)$ such as the response tuning curve.

2.3 Introduction to the Early Visual System

Before discussing how reverse correlation methods are applied to visually responsive neurons, we review the basic anatomy and physiology of the early stages of the visual system. The conversion of a light stimulus into an electrical signal and ultimately an action potential sequence occurs in the retina. Figure 2.4A is an anatomical diagram showing the five principal cell types of the retina, and figure 2.4B is a rough circuit diagram. In the retina, light is first converted into an electrical signal by a phototransduction cascade within rod and cone photoreceptor cells. Figure 2.4B shows intracellular recordings made in neurons of the retina of a mudpuppy (an amphibian). The stimulus used for these recordings was a flash of light falling primarily in the region of the photoreceptor at the left of figure 2.4B. The rod cells, especially the one on the left side of figure 2.4B, are hyperpolarized by the light flash. This electrical signal is passed along to bipolar and horizontal cells through synaptic connections. Note that in one of the bipolar cells, the signal has been inverted leading to depolarization. These smoothly changing membrane potentials provide a graded representation of the light intensity during the flash. This form of coding is adequate for signaling within the retina, where distances are small. However, it is inadequate for the task of conveying information from the retina to the brain.

retinal ganglion cells

The output neurons of the retina are the retinal ganglion cells whose axons form the optic nerve. As seen in figure 2.4B, the subthreshold potentials of the two retinal ganglion cells shown are similar to those of the bipolar cells immediately above them in the figure, but now with superimposed action potentials. The two retinal ganglion cells shown in the figure have different responses and transmit different sequences of action potentials.

ON and OFF responses

G_2 fires while the light is on, and G_1 fires when it turns off. These are called ON and OFF responses, respectively. The optic nerve conducts the output spike trains of retinal ganglion cells to the lateral geniculate nucleus of the thalamus, which acts as a relay station between the retina and primary visual cortex (figure 2.5). Prior to arriving at the LGN, some retinal ganglion cell axons cross the midline at the optic chiasm. This allow the left and right sides of the visual fields from both eyes to be represented on the

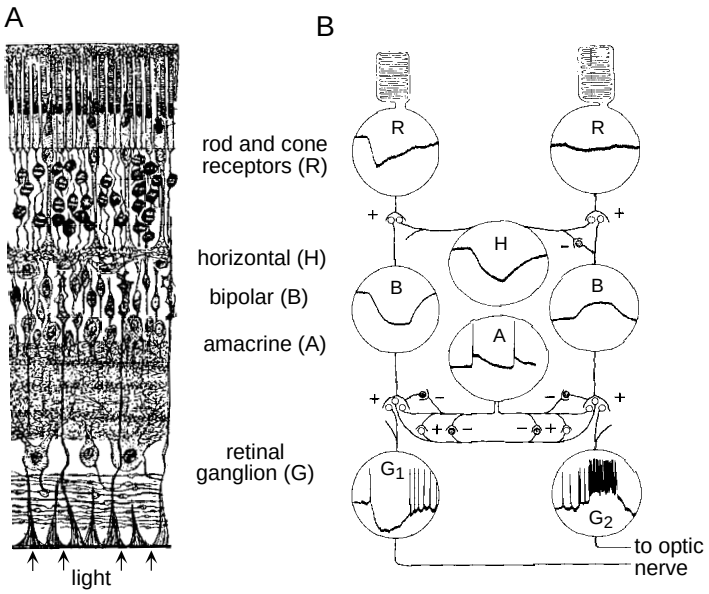

Figure 2.4: A) An anatomical diagram of the circuitry of the retina of a dog. Cell types are identified at right. In the intact eye, illumination is, counter-intuitively, from the bottom of this figure. B) Intracellular recordings from retinal neurons of the mudpuppy responding to flash of light lasting for one second. In the column of cells on the left side of the diagram, the resulting hyperpolarizations are about 4 mV in the rod and retinal ganglion cells, and 8 mV in the bipolar cell. Pluses and minuses represent excitatory and inhibitory synapses respectively. (A adapted from Nicholls et al., 1992; drawing from Cajal, 1911. B data from Werblin and Dowling 1969; figure adapted from Dowling, 1992.)

right and left sides of the brain respectively (figure 2.5).

Neurons in the retina, LGN, and primary visual cortex respond to light stimuli in restricted regions of the visual field called their receptive fields. Patterns of illumination outside the receptive field of a given neuron cannot generate a response directly, although they can significantly affect responses to stimuli within the receptive field. We do not consider such effects, although they are a current focus of experimental and theoretical interest. In the monkey, cortical receptive fields range in size from around a tenth of a degree near the fovea to several degrees in the periphery. Within the receptive fields, there are regions where illumination higher than the background light intensity enhances firing, and other regions where lower illumination enhances firing. The spatial arrangement of these regions determines the selectivity of the neuron to different inputs. The term receptive field is often generalized to refer not only to the overall region where light affects neuronal firing, but also to the spatial and temporal structure within this region.

Visually responsive neurons in the retina, LGN, and primary visual cortex

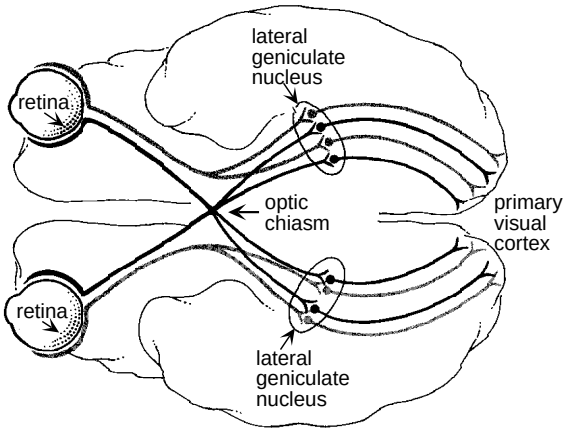

Figure 2.5: Pathway from the retina through the lateral geniculate nucleus (LGN) of the thalamus to the primary visual cortex in the human brain. (Adapted from Nicholls et al., 1992.)

simple and complex cells

are divided into two classes depending on whether or not the contributions from different locations within the visual field sum linearly, as assumed in equation 2.24. X-cells in the cat retina and LGN, P-cells in the monkey retina and LGN, and simple cells in primary visual cortex appear to satisfy this assumption. Other neurons, such as Y cells in the cat retina and LGN, M cells in the monkey retina and LGN, and complex cells in primary visual cortex, do not show linear summation across the spatial receptive field and nonlinearities must be included in descriptions of their responses. We do this for complex cells later in this chapter.

A first step in studying the selectivity of any neuron is to identify the types of stimuli that evoke strong responses. Retinal ganglion cells and LGN neurons have similar selectivities and respond best to circular spots of light surrounded by darkness or dark spots surrounded by light. In primary visual cortex, many neurons respond best to elongated light or dark bars or to boundaries between light and dark regions. Gratings with alternating light and dark bands are effective and frequently used stimuli for these neurons.

Many visually responsive neurons react strongly to sudden transitions in the level of image illumination, a temporal analog of their responsiveness to light-dark spatial boundaries. Static images are not very effective at evoking visual responses. In awake animals, images are constantly kept in motion across the retina by eye movements. In experiments in which the eyes are fixed, moving light bars and gratings, or gratings undergoing periodic light-dark reversals (called counterphase gratings) are used as more effective stimuli than static images. Some neurons in primary visual cortex are directionally selective; they respond more strongly to stimuli moving in one direction than in the other.

To streamline the discussion in this chapter, we consider only greyscale images, although the methods presented can be extended to include color. We also restrict the discussion to two-dimensional visual images, ignoring how visual responses depend on viewing distance and encode depth. In discussing the response properties of retinal, LGN, and V1 neurons, we do not follow the path of the visual signal, nor the historical order of experimentation, but, instead, begin with primary visual cortex and then move back to the LGN and retina. The emphasis is on properties of individual neurons, so we do not discuss encoding by populations of visually responsive neurons. For V1, this has been analyzed in terms of wavelets, a scheme for decomposing images into component pieces, as discussed in chapter 10.

The Retinotopic Map

A striking feature of most visual areas in the brain, including primary visual cortex, is that the visual world is mapped onto the cortical surface in a topographic manner. This means that neighboring points in a visual image evoke activity in neighboring regions of visual cortex. The retinotopic map refers to the transformation from the coordinates of the visual world to the corresponding locations on the cortical surface.

Objects located a fixed distance from one eye lie on a sphere. Locations on this sphere can be represented using the same longitude and latitude angles used for the surface of the earth. Typically, the 'north pole' for this spherical coordinate system is located at the fixation point, the image point that focuses onto the fovea or center of the retina. In this system of coordinates (figure 2.6), the latitude coordinate is called the eccentricity, ϵ, and the longitude coordinate, measured from the horizontal meridian, is called the azimuth a. In primary visual cortex, the visual world is split in half, with the region $-90^\circ \leq a \leq +90^\circ$ for ϵ from 0° to about 70° (for both eyes) represented on the left side of the brain, and the reflection of this region about the vertical meridian represented on the right side of the brain.

eccentricity ϵ
azimuth a

In most experiments, images are displayed on a flat screen (called a tangent screen) that does not coincide exactly with the sphere discussed in the previous paragraph. However, if the screen is not too large the difference is negligible, and the eccentricity and azimuth angles approximately coincide with polar coordinates on the screen (figure 2.6A). Ordinary Cartesian coordinates can also be used to identify points on the screen (figure 2.6). The eccentricity ϵ and the x and y coordinates of the Cartesian system are based on measuring distances on the screen. However, it is customary to divide these measured distances by the distance from the eye to the screen and to multiply the result by $180^\circ/\pi$ so that these coordinates are ultimately expressed in units of degrees. This makes sense because it is the angular not the absolute size and location of an image that is typically relevant for studies of the visual system.

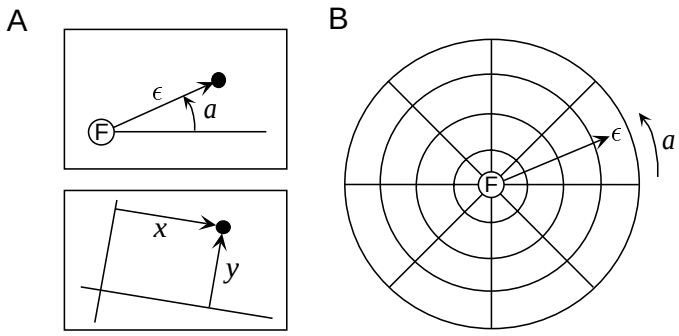

Figure 2.6: A) Two coordinate systems used to parameterize image location. Each rectangle represents a tangent screen, and the filled circle is the location of a particular image point on the screen. The upper panel shows polar coordinates. The origin of the coordinate system is the fixation point F, the eccentricity ϵ is proportional to the radial distance from the fixation point to the image point, and a is the angle between the radial line from F to the image point and the horizontal axis. The lower panel shows Cartesian coordinates. The location of the origin for these coordinates and the orientation of the axes are arbitrary. They are usual chosen to center and align the coordinate system with respect to a particular receptive field being studied. B) A bullseye pattern of radial lines of constant azimuth, and circles of constant eccentricity. The center of this pattern at zero eccentricity is the fixation point F. Such a pattern was used to generated the image in figure 2.7A.

Figure 2.7A shows a dramatic illustration of the retinotopic map in the primary visual cortex of a monkey. The pattern on the cortex seen in figure 2.7A was produced by imaging a radioactive analog of glucose that was taken up by active neurons while a monkey viewed a visual image consisting of concentric circles and radial lines, similar to the pattern in figure 2.6B. The vertical lines correspond to the circles in the image, and the roughly horizontal lines are due to the activity evoked by the radial lines. The fovea is represented at the left-most pole of this piece of cortex and eccentricity increases toward the right. Azimuthal angles are positive in the lower half of the piece of cortex shown, and negative in the upper half.

cortical magnification factor

Figure 2.7B is an approximate mathematical description of the map illustrated in figure 2.7A. To construct this map we assume that eccentricity is mapped onto the horizontal coordinate X of the cortical sheet, and a is mapped onto its Y coordinate. The equations for X and Y as functions of ϵ and a can be obtained through knowledge of a quantity called the cortical magnification factor, $M(\epsilon)$. This determines the distance across a flattened sheet of cortex separating the activity evoked by two nearby image points. Suppose that the two image points in question have eccentricities ϵ and $\epsilon + \Delta\epsilon$ but the same value of the azimuthal coordinate a. The angular distance between these two points is $\Delta\epsilon$. The distance separating the activity evoked by these two image points on the cortex is ΔX. By the definition of the cortical magnification factor, these two quantities satisfy

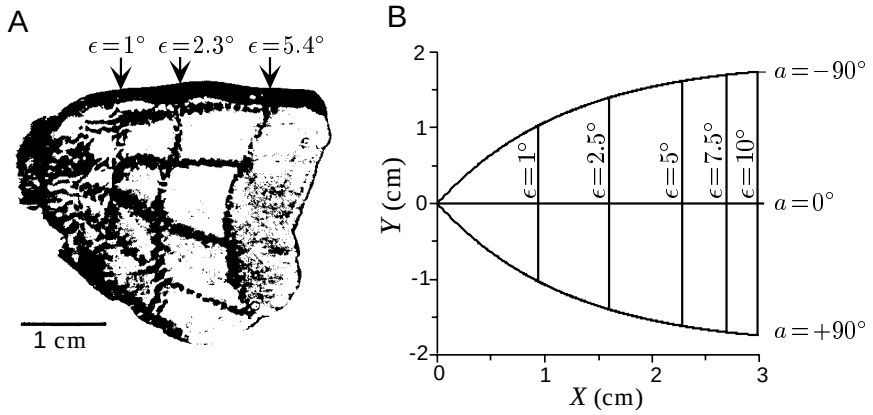

Figure 2.7: A) An autoradiograph of the posterior region of the primary visual cortex from the left side of a macaque monkey brain. The pattern is a radioactive trace of the activity evoked by an image like that in figure 2.6B. The vertical lines correspond to circles at eccentricities of 1°, 2.3°, and 5.4°, and the horizontal lines (from top to bottom) represent radial lines in the visual image at a values of $-90°$, $-45°$, 0°, 45°, and 90°. Only the part of cortex corresponding to the central region of the visual field on one side is shown. B) The mathematical map from the visual coordinates ϵ and a to the cortical coordinates X and Y described by equations 2.15 and 2.17. (A adapted from Tootell et al., 1982.)

$\Delta X = M(\epsilon)\Delta\epsilon$ or, taking the limit as ΔX and $\Delta\epsilon$ go to zero,

$$\frac{dX}{d\epsilon} = M(\epsilon)\,. \tag{2.13}$$

The cortical magnification factor for the macaque monkey, obtained from results such as figure 2.7A is approximately

$$M(\epsilon) = \frac{\lambda}{\epsilon_0 + \epsilon}\,. \tag{2.14}$$

with $\lambda \approx 12$ mm and $\epsilon_0 \approx 1°$. Integrating equation 2.13 and defining $X = 0$ to be the point representing $\epsilon = 0$, we find

$$X = \lambda \ln(1 + \epsilon/\epsilon_0)\,. \tag{2.15}$$

We can apply the same cortical amplification factor to points with the same eccentricity but different a values. The angular distance between two points at eccentricity ϵ with an azimuthal angle difference of Δa is $\Delta a \epsilon \pi/180°$. In this expression, the factor of ϵ corrects for the increase of arc length as a function of eccentricity, and the factor of $\pi/180°$ converts ϵ from degrees to radians. The separation on the cortex, ΔY, corresponding to these points has a magnitude given by the cortical amplification times this distance. Taking the limit $\Delta a \to 0$, we find that we find that

$$\frac{dY}{da} = -\frac{\epsilon\pi}{180°} M(\epsilon)\,. \tag{2.16}$$

The minus sign in this relationship appears because the visual field is inverted on the cortex. Solving equation 2.16 gives

$$Y = -\frac{\lambda \epsilon a \pi}{(\epsilon_0 + \epsilon) 180^\circ}. \tag{2.17}$$

Figure 2.7B shows that these coordinates agree fairly well with the map in figure 2.7A.

complex logarithmic map

For eccentricities appreciably greater than 1°, equations 2.15 and 2.17 reduce to $X \approx \lambda \ln(\epsilon/\epsilon_0)$ and $Y \approx -\lambda \pi a/180^\circ$. These two formulae can be combined by defining the complex numbers $Z = X + iY$ and $z = (\epsilon/\epsilon_0) \exp(-i\pi a/180^\circ)$ (with i equal to the square root of -1) and writing $Z = \lambda \ln(z)$. For this reason, the cortical map is sometimes called a complex logarithmic map (see Schwartz, 1977). For an image scaled radially by a factor γ, eccentricities change according to $\epsilon \to \gamma\epsilon$ while a is unaffected. Scaling of the eccentricity produces a shift $X \to X + \lambda \ln(\gamma)$ over the range of values where the simple logarithmic form of the map is valid. The logarithmic transformation thus causes images that are scaled radially outward on the retina to be represented at locations on the cortex translated in the X direction. For smaller eccentricities, the map we have derived is only approximate even in the complete form given by equations 2.15 and 2.17. This is because the cortical magnification factor is not really isotropic as we have assumed in this derivation, and a complete description requires accounting for the curvature of the cortical surface.

Visual Stimuli

Earlier in this chapter, we used the function $s(t)$ to characterize a time-dependent stimulus. The description of visual stimuli is more complex. Greyscale images appearing on a two-dimensional surface, such as a video monitor, can be described by giving the luminance, or light intensity, at each point on the screen. These pixel locations are parameterized by Cartesian coordinates x and y, as in the lower panel of figure 2.6A. However, pixel-by-pixel light intensities are not a useful way of parameterizing a visual image for the purposes of characterizing neuronal responses. This is because visually responsive neurons, like many sensory neurons, adapt to the overall level of screen illumination. To avoid dealing with adaptation effects, we describe the stimulus by a function $s(x, y, t)$ that is proportional to the difference between the luminance at the point (x, y) at time t and the average or background level of luminance. Often $s(x, y, t)$ is also divided by the background luminance level, making it dimensionless. The resulting quantity is called the contrast.

counterphase sinusoidal grating

During recordings, visual neurons are usually stimulated by images that vary over both space and time. A commonly used stimulus, the counterphase sinusoidal grating, is described by

$$s(x, y, t) = A \cos\left(Kx \cos\Theta + Ky \sin\Theta - \Phi\right) \cos(\omega t). \tag{2.18}$$

Figure 2.8 shows a cartoon of a similar grating (a spatial square-wave is drawn rather than a sinusoid) and illustrates the significance of the parameters K, Θ, Φ, and ω. K and ω are the spatial and temporal frequencies of the grating (these are angular frequencies), Θ is its orientation, Φ its spatial phase, and A its contrast amplitude. This stimulus oscillates in both space and time. At any fixed time, it oscillates in the direction perpendicular to the orientation angle Θ as a function of position, with wavelength $2\pi/K$ (figure 2.8A). At any fixed position, it oscillates in time with period $2\pi/\omega$ (figure 2.8B). For convenience, Θ is measured relative to the y axis rather than the x axis so that a stimulus with $\Theta = 0$, varies in the x, but not in the y, direction. Φ determines the spatial location of the light and dark stripes of the grating. Changing Φ by an amount $\Delta\Phi$ shifts the grating in the direction perpendicular to its orientation by a fraction $\Delta\Phi/2\pi$ of its wavelength. The contrast amplitude A controls the maximum degree of difference between light and dark areas. Because x and y are measured in degrees, K has the rather unusual units of radians per degree and $K/2\pi$ is typically reported in units of cycles per degree. Φ has units of radians, ω is in radians per s, and $\omega/2\pi$ is in Hz.

spatial frequency K
frequency ω
orientation Θ
spatial phase Φ
amplitude A

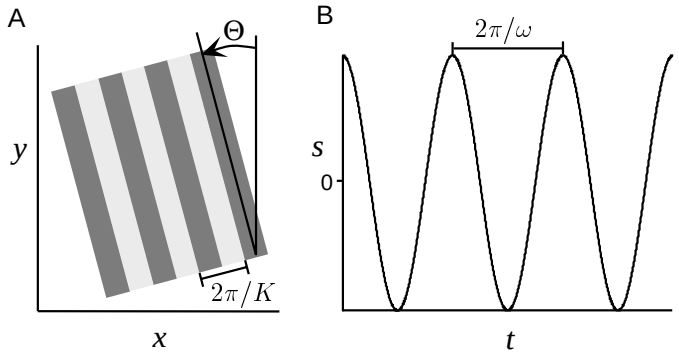

Figure 2.8: A counterphase grating. A) A portion of a square-wave grating analogous to the sinusoidal grating of equation 2.18. The lighter stripes are regions where $s > 0$, and $s < 0$ within the darker stripes. K determines the wavelength of the grating and Θ its orientation. Changing its spatial phase Φ shifts the entire light-dark pattern in the direction perpendicular to the stripes. B) The light-dark intensity at any point of the spatial grating oscillates sinusoidally in time with period $2\pi/\omega$.

Experiments that consider reverse correlation and spike-triggered averages use various types of random and white-noise stimuli in addition to bars and gratings. A white-noise stimulus, in this case, is one that is uncorrelated in both space and time so that

white-noise image

$$\frac{1}{T}\int_0^T dt\, s(x,y,t)s(x',y',t+\tau) = \sigma_s^2\delta(\tau)\delta(x-x')\delta(y-y')\,. \qquad (2.19)$$

Of course, in practice a discrete approximation of such a stimulus must be used by dividing the image space into pixels and time into small bins. In

addition, more structured random sets of images (randomly oriented bars, for example) are sometime used to enhance the responses obtained during stimulation.

The Nyquist Frequency

Many factors limit the maximal spatial frequency that can be resolved by the visual system, but one interesting effect arises from the size and spacing of individual photoreceptors on the retina. The region of the retina with the highest resolution is the fovea at the center of the visual field. Within the macaque or human fovea, cone photoreceptors are densely packed in a regular array. Along any direction in the visual field, a regular array of tightly packed photoreceptors of size Δx samples points at locations $m\Delta x$ for $m = 1, 2, \ldots$. The (angular) frequency that defines the resolution of such an array is called the Nyquist frequency and is given by

Nyquist frequency

$$K_{\rm nyq} = \frac{\pi}{\Delta x}\,. \tag{2.20}$$

To understand the significance of the Nyquist frequency, consider sampling two cosine gratings with spatial frequencies of K and $2K_{\rm nyq} - K$, with $K < K_{\rm nyq}$. These are described by $s = \cos(Kx)$ and $s = \cos((2K_{\rm nyq} - K)x)$. At the sampled points, these functions are identical because $\cos((2K_{\rm nyq} - K)m\Delta x) = \cos(2\pi m - Km\Delta x) = \cos(-Km\Delta x) = \cos(Km\Delta x)$ by the periodicity and evenness of the cosine function (see figure 2.9). As a result, these two gratings cannot be distinguished by examining them only at the sampled points. Any two spatial frequenices $K < K_{\rm nyq}$ and $2K_{\rm nyq} - K$ can be confused with each other in this way, a phenomenon known as aliasing. Conversely, if an image is constructed solely of frequencies less than $K_{\rm nyq}$, it can be reconstructed perfectly from the finite set of samples provided by the array. There are 120 cones per degree at the fovea of the macaque retina which makes $K_{\rm nyq}/(2\pi) = 1/(2\Delta x) = 60$ cycles per degree. In this result, we have divided the right side of equation 2.20, which gives $K_{\rm nyq}$ in units of radians per degree, by 2π to convert the answer to cycles per degree.

2.4 Reverse Correlation Methods - Simple Cells

The spike-triggered average for visual stimuli is defined, as in chapter 1, as the average over trials of stimuli evaluated at times $t_i - \tau$, where t_i for $i = 1, 2, \ldots, n$ are the spike times. Because the light intensity of a visual image depends on location as well as time, the spike-triggered average stimulus is a function of three variables,

$$C(x, y, \tau) = \frac{1}{\langle n \rangle}\left\langle \sum_{i=1}^{n} s(x, y, t_i - \tau) \right\rangle . \tag{2.21}$$

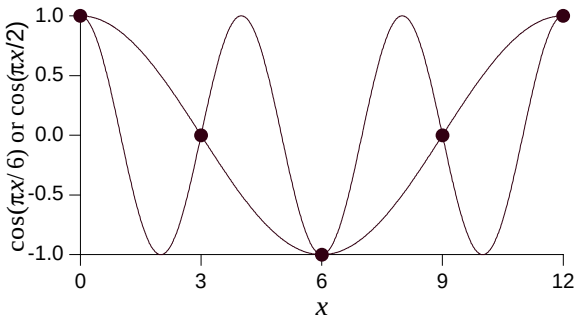

Figure 2.9: Aliasing and the Nyquist frequency. The two curves are the functions $\cos(\pi x/6)$ and $\cos(\pi x/2)$ plotted against x, and the dots show points sampled with a spacing of $\Delta x = 3$. The Nyquist frequency in this case is $\pi/3$, and the two cosine curves match at the sampled points because their spatial frequencies satisfy $2\pi/3 - \pi/6 = \pi/2$.

Here, as in chapter 1, the brackets denote trial averaging, and we have used the approximation $1/n \approx 1/\langle n \rangle$. $C(x, y, \tau)$ is the average value of the visual stimulus at the point (x, y) a time τ before a spike was fired. Similarly, we can define the correlation function between the firing rate at time t and the stimulus at time $t + \tau$, for trials of duration T, as

$$Q_{rs}(x, y, \tau) = \frac{1}{T} \int_0^T dt\, r(t) s(x, y, t + \tau)\,. \tag{2.22}$$

The spike-triggered average is related to the reverse correlation function, as discussed in chapter 1, by

$$C(x, y, \tau) = \frac{Q_{rs}(x, y, -\tau)}{\langle r \rangle}\,, \tag{2.23}$$

where $\langle r \rangle$ is, as usual, the average firing rate over the entire trial, $\langle r \rangle = \langle n \rangle / T$.

To estimate the firing rate of a neuron in response to a particular image, we add a function of the output of a linear filter of the stimulus to the background firing rate r_0, as in equation 2.8, $r_{est}(t) = r_0 + F(L(t))$. As in equation 2.7, the linear estimate $L(t)$ is obtained by integrating over the past history of the stimulus with a kernel acting as the weighting function. Because visual stimuli depend on spatial location, we must decide how contributions from different image locations are to be combined to determine $L(t)$. The simplest assumption is that the contributions from different spatial points add linearly so that $L(t)$ is obtained by integrating over all x and y values,

linear response estimate

$$L(t) = \int_0^\infty d\tau \iint dx dy\, D(x, y, \tau) s(x, y, t - \tau)\,. \tag{2.24}$$

The kernel $D(x, y, \tau)$ determines how strongly, and with what sign, the visual stimulus at the point (x, y) and at time $t - \tau$ affects the firing

rate of the neuron at time t. As in equation 2.6, the optimal kernel is given in terms of the firing rate-stimulus correlation function, or the spike-triggered average, for a white-noise stimulus with variance parameter σ_s^2 by

$$D(x, y, \tau) = \frac{Q_{rs}(x, y, -\tau)}{\sigma_s^2} = \frac{\langle r \rangle C(x, y, \tau)}{\sigma_s^2} \,. \tag{2.25}$$

space-time receptive field

The kernel $D(x, y, \tau)$ defines the space-time receptive field of a neuron. Because $D(x, y, \tau)$ is a function of three variables, it can be difficult to measure and visualize. For some neurons, the kernel can be written as a product of two functions, one that describes the spatial receptive field and the other the temporal receptive field,

$$D(x, y, \tau) = D_{\rm s}(x, y) D_{\rm t}(\tau) \,. \tag{2.26}$$

separable receptive field

nonseparable receptive field

Such neurons are said to have separable space-time receptive fields. Separability requires that the spatial structure of the receptive field does not change over time except by an overall multiplicative factor. When $D(x, y, \tau)$ cannot be written as the product of two terms, the neuron is said to have a nonseparable space-time receptive field. Given the freedom in equation 2.8 to set the scale of D (by suitably adjusting the function F), we typically normalize $D_{\rm s}$ so that its integral is one, and use a similar rule for the components from which $D_{\rm t}$ is constructed. We begin our analysis by studying first the spatial and then the temporal components of a separable space-time receptive field and then proceed to the nonseparable case. For simplicity, we ignore the possibility that cells can have slightly different receptive fields for the two eyes, which underlies the disparity tuning considered in chapter 1.

Spatial Receptive Fields

Figures 2.10A and C show the spatial structure of spike-triggered average stimuli for two simple cells in the primary visual cortex of a cat (area 17) with approximately separable space-time receptive fields. These receptive fields are elongated in one direction, and there are some regions within the receptive field where $D_{\rm s}$ is positive, called ON regions, and others where it is negative, called OFF regions. The integral of the linear kernel times the stimulus can be visualized by noting how the OFF (black) and ON (white) regions overlap the image (see figure 2.11) . The response of a neuron is enhanced if ON regions are illuminated ($s > 0$) or if OFF regions are darkened ($s < 0$) relative to the background level of illumination. Conversely, they are suppressed by darkening ON regions or illuminating OFF regions. As a result, the neurons of figures 2.10A and C respond most vigorously to light-dark edges positioned along the border between the ON and OFF regions and oriented parallel to this border and to the elongated direction of the receptive fields (figure 2.11). Figures 2.10 and 2.11 show

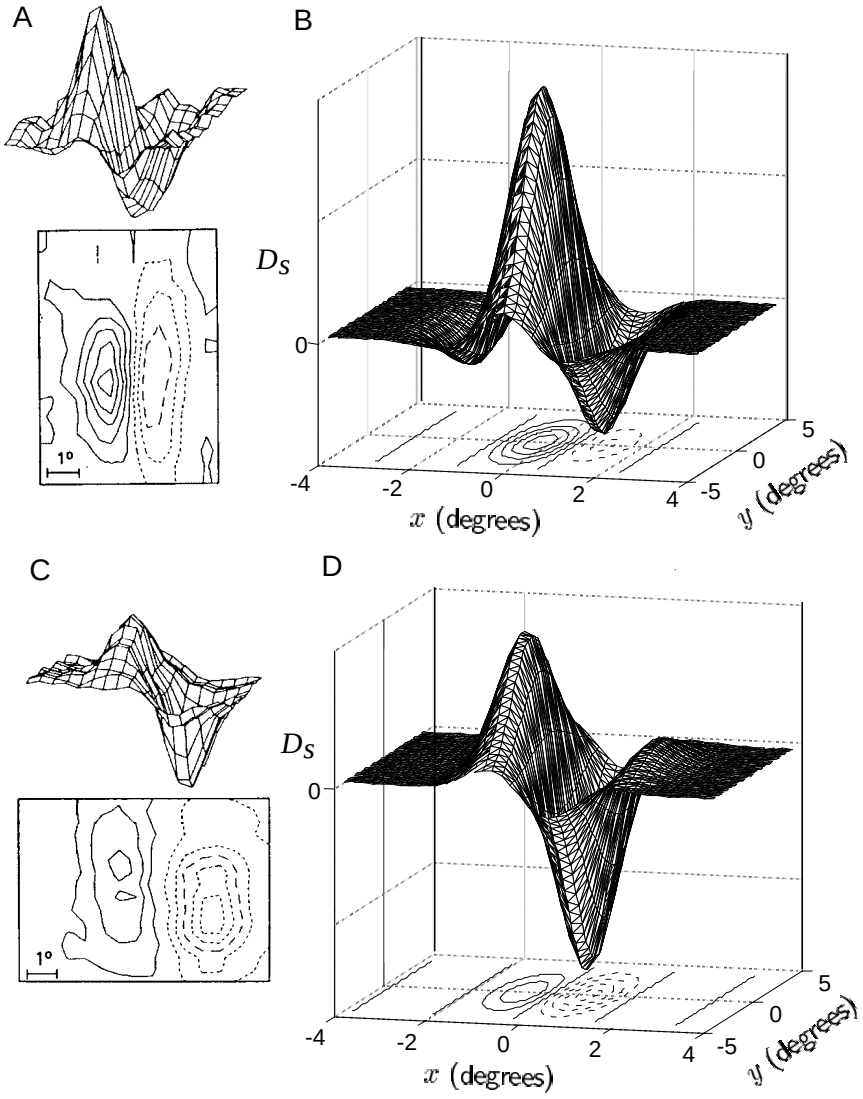

Figure 2.10: Spatial receptive field structure of simple cells. A) and C) Spatial structure of the receptive fields of two neurons in cat primary visual cortex determined by averaging stimuli between 50 ms and 100 ms prior to an action potential. The upper plots are three-dimensional representations, with the horizontal dimensions acting as the x-y plane and the vertical dimension indicating the magnitude and sign of $D_s(x, y)$. The lower contour plots represent the x-y plane. Regions with solid contour curves are ON areas where $D_s(x, y) > 0$ and regions with dashed contours show OFF areas where $D_s(x, y) < 0$. B) and D) Gabor functions of the form 2.27 with $\sigma_x = 1^\circ$, $\sigma_y = 2^\circ$, $1/k = 0.56^\circ$, and $\phi = 1 - \pi/2$ (B) or $\phi = 1 - \pi$ (D) chosen to match the receptive fields in A and C. (A and C adapted from Jones and Palmer, 1987a.)

receptive fields with two major subregions. Simple cells are found with from one to five subregions. Along with the ON-OFF patterns we have seen, another typical arrangement is a three-lobed receptive field with an OFF-ON-OFF or ON-OFF-ON subregions, as seen in figure 2.17B.

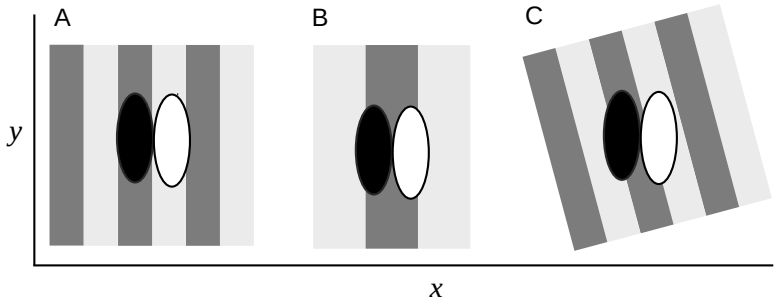

Figure 2.11: Grating stimuli superimposed on spatial receptive fields similar to those shown in figure 2.10. The receptive field is shown as two oval regions, one dark to represent an OFF area where $D_s < 0$ and one white to denote an ON region where $D_s > 0$. A) A grating with the spatial wavelength, orientation, and spatial phase shown produces a high firing rate because a dark band completely overlaps the OFF area of the receptive field and a light band overlaps the ON area. B) The grating shown is non-optimal due to a mismatch in both the spatial phase and frequency, so that the ON and OFF regions each overlap both light and dark stripes. C) The grating shown is at a non-optimal orientation because each region of the receptive field overlaps both light and dark stripes.

Gabor function

A mathematical approximation of the spatial receptive field of a simple cell is provided by a Gabor function, which is a product of a Gaussian function and a sinusoidal function. Gabor functions are by no means the only functions used to fit spatial receptive fields of simple cells. For example, gradients of Gaussians are sometimes used. However, we will stick to Gabor functions, and to simplify the notation, we choose the coordinates x and y so that the borders between the ON and OFF regions are parallel to the y axis. We also place the origin of the coordinates at the center of the receptive field. With these choices, we can approximate the observed receptive field structures using the Gabor function

$$D_s(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{x^2}{2\sigma_x^2}-\frac{y^2}{2\sigma_y^2}\right)\cos(kx-\phi)\,. \qquad (2.27)$$

rf size σ_x, σ_y
preferred spatial frequency k
preferred spatial phase ϕ

The parameters in this function determine the properties of the spatial receptive field: σ_x and σ_y determine its extent in the x and y directions respectively; k, the preferred spatial frequency, determines the spacing of light and dark bars that produce the maximum response (the preferred spatial wavelength is $2\pi/k$); and ϕ is the preferred spatial phase which determines where the ON-OFF boundaries fall within the receptive field. For this spatial receptive field, the sinusoidal grating of the form 2.18 that produces the maximum response for a fixed value of A has $K = k$, $\Phi = \phi$, and $\Theta = 0$.

Figures 2.10B and D, show Gabor functions chosen specifically to match the data in figures 2.10A and C. Figure 2.12 shows x and y plots of a variety of Gabor functions with different parameter values. As seen in figure 2.12, Gabor functions can have various types of symmetry, and variable numbers of significant oscillations (or subregions) within the Gaussian envelope. The number of subregions within the receptive field is determined by the product $k\sigma_x$ and is typically expressed in terms of a quantity known as the bandwidth b. The bandwidth is defined as $b = \log_2(K_+/K_-)$ where $K_+ > k$ and $K_- < k$ are the spatial frequencies of gratings that produce one half the response amplitude of a grating with $K = k$. High bandwidths correspond to low values of $k\sigma_x$, meaning that the receptive field has few subregions and poor spatial frequency selectivity. Neurons with more subfields are more selective to spatial frequency, and they have smaller bandwidths and larger values of $k\sigma_x$.

bandwidth

The bandwidth is the width of the spatial frequency tuning curve measured in octaves. The spatial frequency tuning curve as a function of K for a Gabor receptive field with preferred spatial frequency k and receptive field width σ_x is proportional to $\exp(-\sigma_x^2(k-K)^2/2)$ (see equation 2.34 below). The values of K_+ and K_- needed to compute the bandwidth are thus determined by the condition $\exp(-\sigma_x^2(k-K_\pm)^2/2) = 1/2$. Solving this equation gives $K_\pm = k \pm (2\ln(2))^{1/2}/\sigma_x$ from which we obtain

$$b = \log_2\left(\frac{k\sigma_x + \sqrt{2\ln(2)}}{k\sigma_x - \sqrt{2\ln(2)}}\right) \quad \text{or} \quad k\sigma_x = \sqrt{2\ln(2)}\,\frac{2^b+1}{2^b-1}. \tag{2.28}$$

Bandwidth is only defined if $k\sigma_x > \sqrt{2\ln(2)}$, but this is usually the case for V1 neurons. For V1 neurons, bandwidths range from about 0.5 to 2.5 corresponding to $k\sigma_x$ between 1.7 and 6.9.

The response characterized by equation 2.27 is maximal if light-dark edges are parallel to the y axis, so the preferred orientation angle is zero. An arbitrary preferred orientation θ can be generated by rotating the coordinates, making the substitutions $x \to x\cos(\theta) + y\sin(\theta)$ and $y \to y\cos(\theta) - x\sin(\theta)$ in equation 2.27. This produces a spatial receptive field that is maximally responsive to a grating with $\Theta = \theta$. Similarly, a receptive field centered at the point (x_0, y_0) rather than at the origin can be constructed by making the substitutions $x \to x - x_0$ and $y \to y - y_0$.

preferred orientation θ

rf center x_0, y_0

Temporal Receptive Fields

Figure 2.13 reveals the temporal development of the space-time receptive field of a neuron in the cat primary visual cortex through a series of snap shots of its spatial receptive field. More than 300 ms prior to a spike, there is little correlation between the visual stimulus and the upcoming spike. Around 210 ms before the spike ($\tau = 210$ ms), a two-lobed OFF-ON receptive field, similar to the ones in figures 2.10, is evident. As τ decreases (recall that τ measures time in a reversed sense), this structure first fades

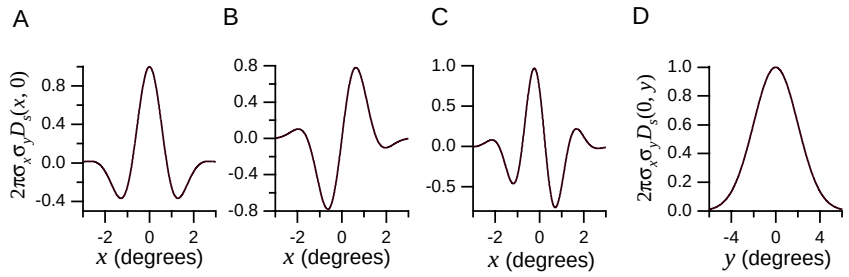

Figure 2.12: Gabor functions of the form given by equation 2.27. For convenience we plot the dimensionless function $2\pi\sigma_x\sigma_y D_s$. A) A Gabor function with $\sigma_x = 1^\circ$, $1/k = 0.5^\circ$, and $\phi = 0$ plotted as a function of x for $y = 0$. This function is symmetric about $x = 0$. B) A Gabor function with $\sigma_x = 1^\circ$, $1/k = 0.5^\circ$, and $\phi = \pi/2$ plotted as a function of x for $y = 0$. This function is antisymmetric about $x = 0$ and corresponds to using a sine instead of a cosine function in equation 2.27. C) A Gabor function with $\sigma_x = 1^\circ$, $1/k = 0.33^\circ$, and $\phi = \pi/4$ plotted as a function of x for $y = 0$. This function has no particular symmetry properties with respect to $x = 0$. D) The Gabor function of equation 2.27 with $\sigma_y = 2^\circ$ plotted as a function of y for $x = 0$. This function is simply a Gaussian.

away and then reverses, so that the receptive field 75 ms before a spike has the opposite sign from what appeared at $\tau = 210$ ms. Due to latency effects, the spatial structure of the receptive field is less significant for $\tau < 75$ ms. The stimulus preferred by this cell is thus an appropriately aligned dark-light boundary that reverses to a light-dark boundary over time.

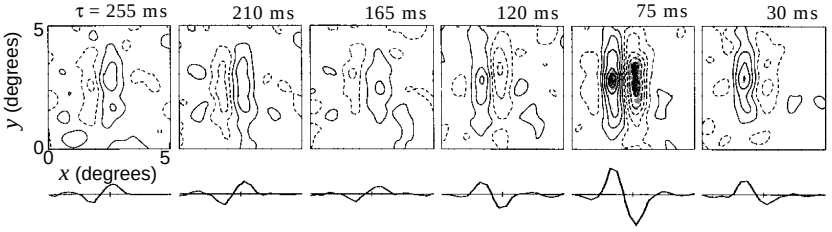

Figure 2.13: Temporal evolution of a spatial receptive field. Each panel is a plot of $D(x, y, \tau)$ for a different value of τ. As in figure 2.10, regions with solid contour curves are areas where $D(x, y, \tau) > 0$ and regions with dashed contours have $D(x, y, \tau) < 0$. The curves below the contour diagrams are one-dimension plots of the receptive field as a function of x alone. The receptive field is maximally different from zero for $\tau = 75$ ms with the spatial receptive field reversed from what it was at $\tau = 210$ ms. (Adapted from DeAngelis et al., 1995.)

Reversal effects like those seen in figure 2.13 are a common feature of space-time receptive fields. Although the magnitudes and signs of the different spatial regions in figure 2.13 vary over time, their locations and shapes remain fairly constant. This indicates that the neuron has, to a good approximation, a separable space-time receptive field. When a space-time

receptive field is separable, the reversal can be described by a function $D_t(\tau)$ that rises from zero, becomes positive, then negative, and ultimately goes to zero as τ increases. Adelson and Bergen (1985) proposed the function shown in Figure 2.14,

$$D_t(\tau) = \alpha \exp(-\alpha\tau)\left(\frac{(\alpha\tau)^5}{5!} - \frac{(\alpha\tau)^7}{7!}\right) \tag{2.29}$$

for $\tau \geq 0$, and $D_t(\tau) = 0$ for $\tau < 0$. Here, α is a constant that sets the scale for the temporal development of the function. Single phase responses are also seen for V1 neurons and these can be described by eliminating the second term in equation 2.29. Three-phase responses, which are sometimes seen, must be described by a more complicated function.

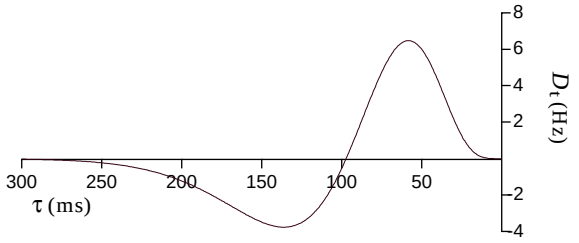

Figure 2.14: Temporal structure of a receptive field. The function $D_t(\tau)$ of equation 2.29 with $\alpha = 1/(15\text{ ms})$.

Response of a Simple Cell to a Counterphase Grating

The response of a simple cell to a counterphase grating stimulus (equation 2.18) can be estimated by computing the function $L(t)$. For the separable receptive field given by the product of the spatial factor in equation 2.27 and the temporal factor in 2.29, the linear estimate of the response can be written a product of two terms,

$$L(t) = L_s L_t(t)\,, \tag{2.30}$$

where

$$L_s = \int dxdy\, D_s(x,y) A \cos\left(Kx\cos(\Theta) + Ky\sin(\Theta) - \Phi\right)\,. \tag{2.31}$$

and

$$L_t(t) = \int_0^\infty d\tau\, D_t(\tau)\cos\left(\omega(t-\tau)\right)\,. \tag{2.32}$$

The reader is invited to compute these integrals for the case $\sigma_x = \sigma_y = \sigma$. To show the selectivity of the resulting spatial receptive fields, we plot (in

figure 2.15) L_s as functions of the parameters Θ, K, and Φ that determine the orientation, spatial frequency, and spatial phase of the stimulus. It is also instructive to write out L_s for various special parameter values. First, if the spatial phase of the stimulus and the preferred spatial phase of the receptive field are zero ($\Phi = \phi = 0$), we find that

$$L_s = A \exp\left(-\frac{\sigma^2(k^2+K^2)}{2}\right) \cosh\left(\sigma^2 kK\cos(\Theta)\right) , \tag{2.33}$$

which determines the orientation and spatial frequency tuning for an optimal spatial phase. Second, for a grating with the preferred orientation $\Theta = 0$ and a spatial frequency that is not too small, the full expression for L_s can be simplified by noting that $\exp(-\sigma^2 kK) \approx 0$ for the values of $k\sigma$ normally encountered (for example, if $K = k$ and $k\sigma = 2$, $\exp(-\sigma^2 kK) = 0.02$). Using this approximation, we find

$$L_s = \frac{A}{2} \exp\left(-\frac{\sigma^2(k-K)^2}{2}\right) \cos(\phi - \Phi) \tag{2.34}$$

which reveals a Gaussian dependence on spatial frequency and a cosine dependence on spatial phase.

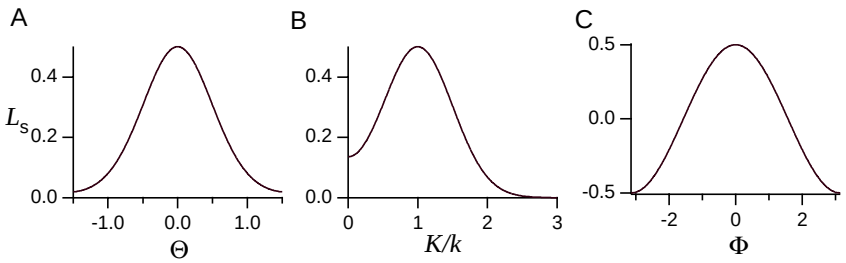

Figure 2.15: Selectivity of a Gabor filter with $\theta = \phi = 0$, $\sigma_x = \sigma_y = \sigma$ and $k\sigma = 2$ acting on a cosine grating with $A = 1$. A) L_s as a function of stimulus orientation Θ for a grating with the preferred spatial frequency and phase, $K = k$ and $\Phi = 0$. B) L_s as a function of the ratio of the stimulus spatial frequency to its preferred value, K/k, for a grating oriented in the preferred direction $\Theta = 0$ and with the preferred phase $\Phi = 0$. C) L_s as a function of stimulus spatial phase Φ for a grating with the preferred spatial frequency and orientation, $K = k$ and $\Theta = 0$.

The temporal frequency dependence of the amplitude of the linear response estimate is plotted as a function of the temporal frequency of the stimulus ($\omega/2\pi$ rather than the angular frequency ω) in figure 2.16. The peak value around 4 Hz and roll off above 10 Hz are typical for V1 neurons and for cortical neurons in other primary sensory areas as well.

Space-Time Receptive Fields

It is instructive to display the function $D(x, y, \tau)$ in a space-time plot rather than as a sequence of spatial plots (as in figure 2.13). To do this, we sup-

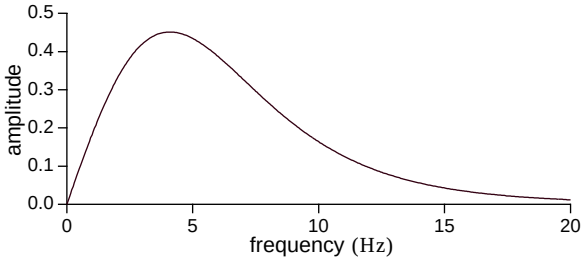

Figure 2.16: Frequency response of a model simple cell based on the temporal kernel of equation 2.29. The amplitude of the sinusoidal oscillations of $L_t(t)$ produced by a counterphase grating is plotted as a function of the temporal oscillation frequency, $\omega/2\pi$.

press the y dependence and plot x-τ projections of the space-time kernel. Space-time plots of receptive fields from two simple cells of the cat primary visual cortex are shown in figure 2.17. The receptive field in figure 2.17A is approximately separable, and it has OFF and ON subregions that

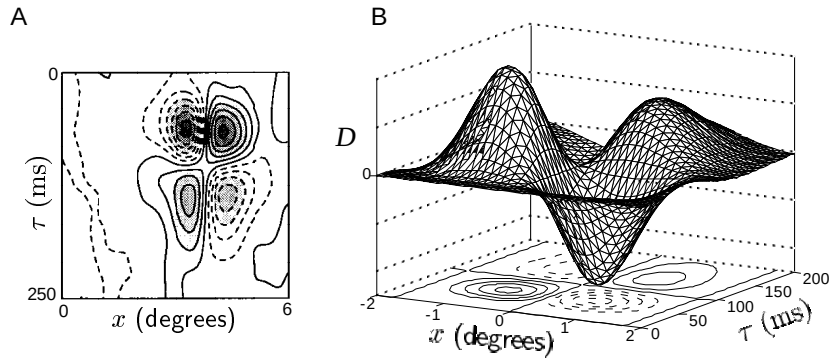

Figure 2.17: A separable space-time receptive field. A) An x-τ plot of an approximately separable space-time receptive field from cat primary visual cortex. OFF regions are shown with dashed contour lines and ON regions with solid contour lines. The receptive field has side-by-side OFF and ON regions that reverse as a function of τ. B) Mathematical descriptions of the space-time receptive field in A constructed by multiplying a Gabor function (evaluated at $y = 0$) with $\sigma_x = 1^\circ$, $1/k = 0.56^\circ$, and $\phi = \pi/2$ by the temporal kernel of equation 2.29 with $1/\alpha = 15$ ms. (A adapted from DeAngelis et al., 1995.)

reverse to ON and OFF subregions as a function of τ, similar to the reversal seen in figure 2.13. Figure 2.17B shows an x-τ contour plot of a separable space-time kernel, similar to the one in figure 2.17A, generated by multiplying a Gabor function by the temporal kernel of equation 2.29.

We can also plot the visual stimulus in a space-time diagram, suppressing the y coordinate by assuming that the image does not vary as a function of y. For example, figure 2.18A shows a grating of vertically oriented stripes

moving to the left on an *x*-*y* plot. In the *x*-*t* plot of figure 2.18B, this image appears as a series of sloped dark and light bands. These represent the projection of the image in figure 2.18A onto the *x* axis evolving as a function of time. The leftward slope of the bands corresponds to the leftward movement of the image.

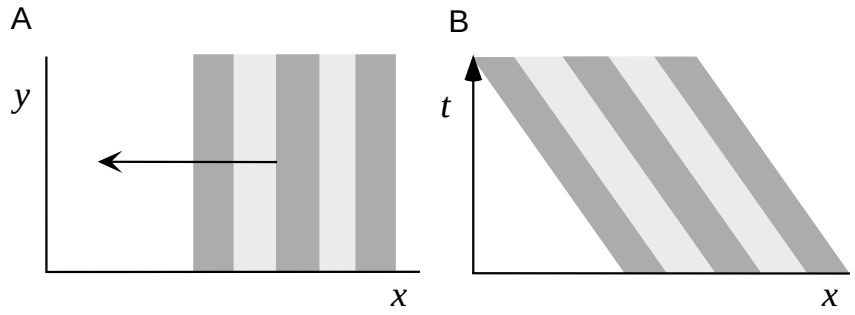

Figure 2.18: Space and space-time diagrams of a moving grating. A) A vertically oriented grating moves to the left on a two-dimensional screen. B) The space-time diagram of the image in A. The *x* location of the dark and light bands moves to the left as time progresses upward, representing the motion of the grating.

Most neurons in primary visual cortex do not respond strongly to static images, but respond vigorously to flashed and moving bars and gratings. The receptive field structure of figure 2.17 reveals why this is the case, as is shown in figures 2.19 and 2.20. The image in figures 2.19A-C is a dark bar that is flashed on for a brief period of time. To describe the linear response estimate at different times we show a cartoon of a space-time receptive field similar to the one in figure 2.17A. The receptive field is positioned at three different times in figures 2.19A, B, and C. The height of the horizontal axis of the receptive field diagram indicates the time when the estimation is being made. Figure 2.19A corresponds to an estimate of $L(t)$ at the moment when the image first appears. At this time, $L(t) = 0$. As time progresses, the receptive field diagram moves upward. Figure 2.19B generates an estimate at the moment of maximum response when the dark image overlaps the OFF area of the space-time receptive field, producing a positive contribution to $L(t)$. Figure 2.19C shows a later time when the dark image overlaps an ON region, generating a negative $L(t)$. The response for this flashed image is thus transient firing followed by suppression, as shown in Figure 2.19D.

Figures 2.19E and F show why a static dark bar is an ineffective stimulus. The static bar overlaps both the OFF region for small τ and the reversed ON region for large τ, generating opposing positive and negative contributions to $L(t)$. The flashed dark bar of figures 2.19A-C is a more effective stimulus because there is a time when it overlaps only the OFF region.

Figure 2.20 shows why a moving grating is a particularly effective stimulus. The grating moves to the left in 2.20A-C. At the time corresponding to

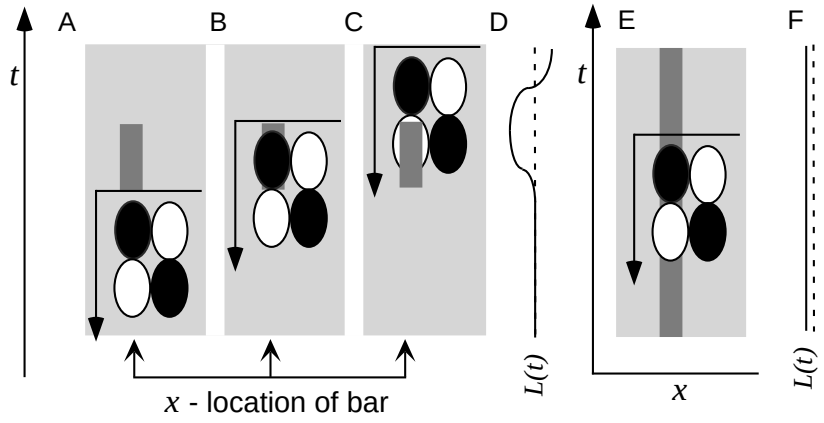

Figure 2.19: Responses to dark bars estimated from a separable space-time receptive field. Dark ovals in the receptive field diagrams are OFF regions and light circles are ON regions. The linear estimate of the response at any time is determined by positioning the receptive field diagram so that its horizontal axis matches the time of response estimation and noting how the OFF and ON regions overlap with the image. A-C) The image is a dark bar that is flashed on for a short interval of time. There is no response (A) until the dark image overlaps the OFF region (B) when $L(t) > 0$. The response is later suppressed when the dark bar overlaps the ON region (C) and $L(t) < 0$. D) A plot of $L(t)$ versus time corresponding to the responses generated in A-C. Time runs vertically in this plot, and $L(t)$ is plotted horizontally with the dashed line indicating the zero axis and positive values plotted to the left. E) The image is a static dark bar. The bar overlaps both an OFF and an ON region generating opposing positive and negative contributions to $L(t)$. F) The weak response corresponding to E, plotted as in D.

the positioning of the receptive field diagram in 2.20A, a dark band stimulus overlaps both OFF regions and light bands overlap both ON regions. Thus, all four regions contribute positive amounts to $L(t)$. As time progresses and the receptive field moves upward in the figure, the alignment will sometimes be optimal, as in 2.20A, and sometimes non-optimal, as in 2.20B. This produces an $L(t)$ that oscillates as a function of time between positive and negative values (2.20C). Figures 2.20D-F show that a neuron with this receptive field responds equally to a grating moving to the right. Like the left-moving grating in figures 2.20A-C, the right-moving grating can overlap the receptive field in an optimal manner (2.20D) producing a strong response, or in a maximally negative manner (2.20E) producing strong suppression of response, again resulting in an oscillating response (2.20F). Separable space-time receptive fields can produce responses that are maximal for certain speeds of grating motion, but they are not sensitive to the direction of motion.

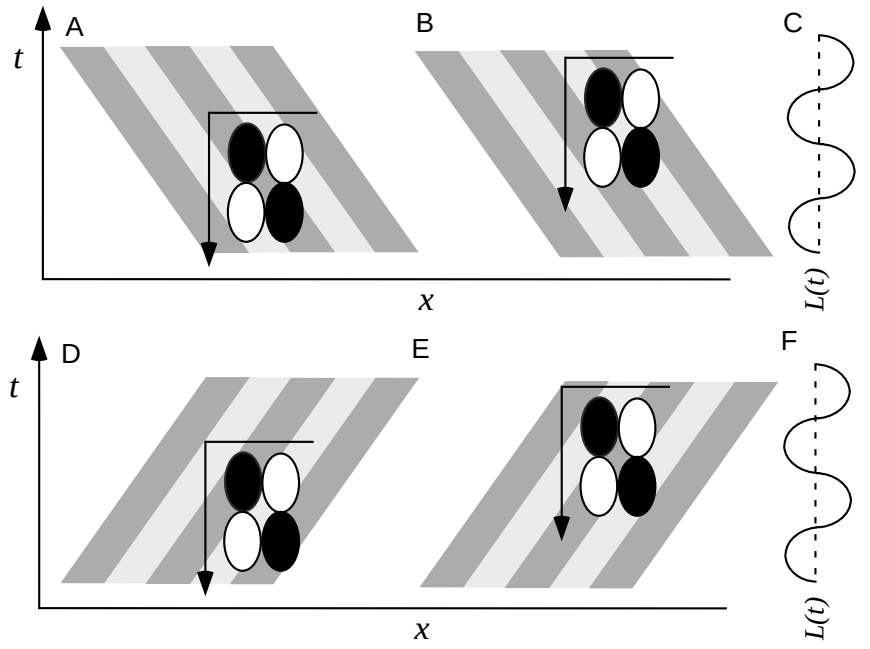

Figure 2.20: Responses to moving gratings estimated from a separable space-time receptive field. The receptive field is the same as in figure 2.19. A-C) The stimulus is a grating moving to the left. At the time corresponding to A, OFF regions overlap with dark bands and ON regions with light bands generating a strong response. At the time of the estimate in B, the alignment is reversed, and $L(t)$ is negative. C) A plot of $L(t)$ versus time corresponding to the responses generated in A-B. Time runs vertically in this plot and $L(t)$ is plotted horizontally with the dashed line indicating the zero axis and positive values plotted to the left. D-F) The stimulus is a grating moving to the right. The responses are identical to those in A-C.

Nonseparable Receptive Fields

Many neurons in primary visual cortex are selective for the direction of motion of an image. Accounting for direction selectivity requires nonseparable space-time receptive fields. An example of a nonseparable receptive field is shown in figure 2.21A. This neuron has a three-lobed OFF-ON-OFF spatial receptive field, and these subregions shift to the left as time moves forward (and τ decreases). This means that the optimal stimulus for this neuron has light and dark areas that move toward the left. One way to describe a nonseparable receptive field structure is to use a separable function constructed from a product of a Gabor function for D_s and equation 2.29 for D_t, but express these as functions of a mixture or rotation of the x and τ variables. The rotation of the space-time receptive field, as seen in figure 2.21B, is achieved by mixing the space and time coordinates using the transformation

$$D(x, y, \tau) = D_s(x', y) D_t(\tau') \tag{2.35}$$

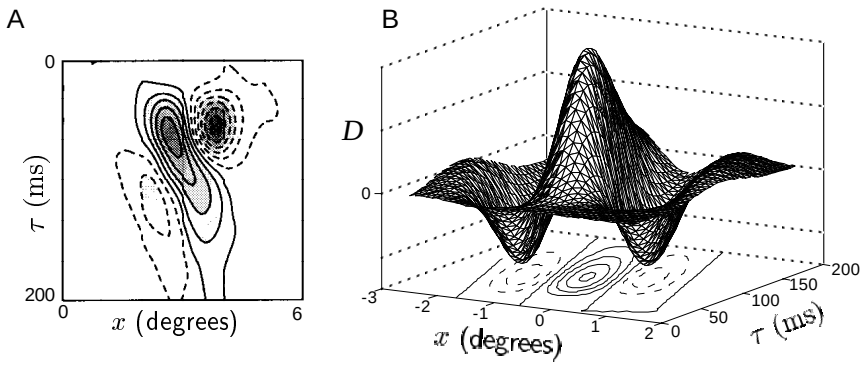

Figure 2.21: A nonseparable space-time receptive field. A) An x-τ plot of the space-time receptive field of a neuron from cat primary visual cortex. OFF regions are shown with dashed contour lines and ON regions with solid contour lines. The receptive field has a central ON region and two flanking OFF regions that shift to the left over time. B) Mathematical description of the space-time receptive field in A constructed from equations 2.35 - 2.37. The Gabor function used (evaluated at $y = 0$) had $\sigma_x = 1^\circ$, $1/k = 0.5^\circ$, and $\phi = 0$. D_t is given by the expression in equation 2.29 with $\alpha = 20$ ms except that the second term, with the seventh power function, was omitted because the receptive field does not reverse sign in this example. The x-τ rotation angle used was $\psi = \pi/9$ and the conversion factor was $c = 0.02\,^\circ/\,\text{ms}$. (A adapted from DeAngelis et al., 1995.)

with

$$x' = x\cos(\psi) - c\tau\sin(\psi) \tag{2.36}$$

and

$$\tau' = \tau\cos(\psi) + \frac{x}{c}\sin(\psi)\,. \tag{2.37}$$

The factor c converts between the units of time (ms) and space (degrees) and ψ is the space-time rotation angle. The rotation operation is not the only way to generate nonseparable space-time receptive fields. They are often constructed by adding together two or more separable space-time receptive fields with different spatial and temporal characteristics.

Figure 2.22 shows how a nonseparable space-time receptive field can produce a response that is sensitive to the direction of motion of a grating. Figures 2.22A-C show a left-moving grating and, in 2.22A, the cartoon of the receptive field is positioned at a time when a light area of the image overlaps the central ON region and dark areas overlap the flanking OFF regions. This produces a large positive $L(t)$. At other times, the alignment is non-optimal (2.22B), and over time, $L(t)$ oscillates between fairly large positive and negative values (2.22C). The nonseparable space-time receptive field does not overlap optimally with the right-moving grating of figures 2.22D-F at any time and the response is correspondingly weaker (2.22F). Thus, a neuron with a nonseparable space-time receptive field can

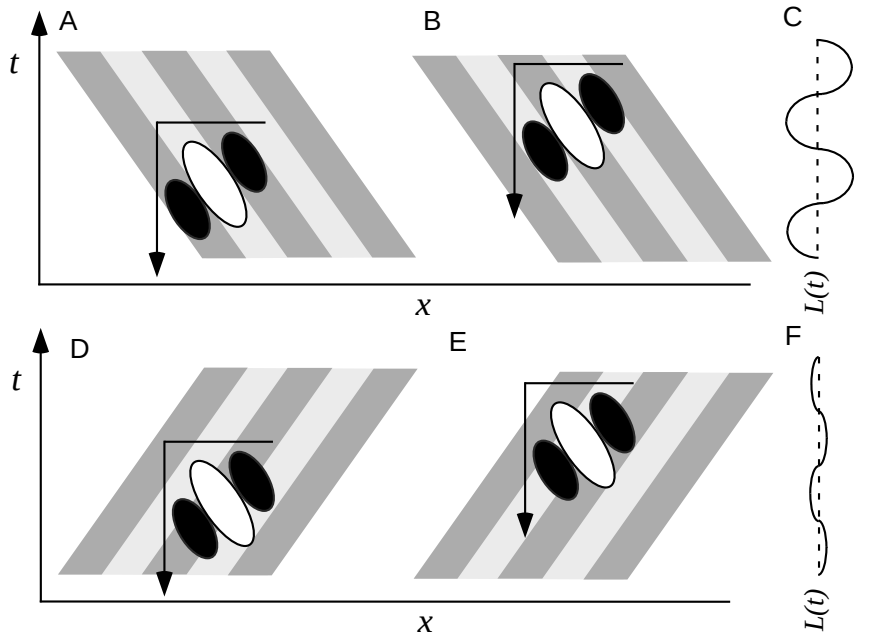

Figure 2.22: Responses to moving gratings estimated from a nonseparable space-time receptive field. Dark areas in the receptive field diagrams represent OFF regions and light areas ON regions. A-C) The stimulus is a grating moving to the left. At the time corresponding to A, OFF regions overlap with dark bands and the ON region overlaps a light band generating a strong response. At the time of the estimate in B, the alignment is reversed, and $L(t)$ is negative. C) A plot of $L(t)$ versus time corresponding to the responses generated in A-B. Time runs vertically in this plot and $L(t)$ is plotted horizontally with the dashed line indicating the zero axis. D-F) The stimulus is a grating moving to the right. Because of the tilt of the space-time receptive field, the alignment with the right-moving grating is never optimal and the response is weak (F).

direction selectivity *preferred velocity*

be selective for the direction of motion of a grating and for its velocity, responding most vigorously to an optimally spaced grating moving at a velocity given, in terms of the parameters in equation 2.36, by $c\tan(\psi)$.

Static Nonlinearities - Simple Cells

Once the linear response estimate $L(t)$ has been computed, the firing rate of a visually responsive neuron can be approximated by using equation 2.8, $r_{est}(t) = r_0 + F(L(t))$ where F is an appropriately chosen static nonlinearity. The simplest choice for F consistent with the positive nature of firing rates, is rectification, $F = G[L]_+$, with G set to fit the magnitude of the measured firing rates. However, this choice makes the firing rate a linear function of the contrast amplitude, which does not match the data on the contrast dependence of visual responses. Neural responses saturate as

contrast saturation

the contrast of the image increases and are more accurately described by $r \propto A^n/(A_{1/2}^n + A^n)$ where n is near two, and $A_{1/2}$ is a parameter equal to the contrast amplitude that produces a half-maximal response. This led Heeger (1992) to propose that an appropriate static nonlinearity to use is

$$F(L) = \frac{G[L]_+^2}{A_{1/2}^2 + G[L]_+^2} \tag{2.38}$$

because this reproduces the observed contrast dependence. A number of variants and extensions of this idea have also been considered, including, for example, that the denominator of this expression should include L factors for additional neurons with nearby receptive fields. This can account for the effects of visual stimuli outside the 'classical' receptive field. Discussion of these effects is beyond the scope of this chapter.

2.5 Static Nonlinearities - Complex Cells

Recall that a large proportion of the neurons in primary visual cortex is separated into classes of simple and complex cells. While linear methods, such as spike-triggered averages, are useful for revealing the properties of simple cells, at least to a first approximation, complex cells display features that are fundamentally incompatible with a linear description. The spatial receptive fields of complex cells cannot be divided into separate ON and OFF regions that sum linearly to generate the response. Areas where light and dark images excite the neuron overlap making it difficult to measure and interpret spike-triggered average stimuli. Nevertheless, like simple cells, complex cells are selective to the spatial frequency and orientation of a grating. However, unlike simple cells, complex cells respond to bars of light or dark no matter where they are placed within the overall receptive field. Likewise, the responses of complex cells to grating stimuli show little dependence on spatial phase. Thus, a complex cell is selective for a particular type of image independent of its exact spatial position within the receptive field. This may represent an early stage in the visual processing that ultimately leads to position-invariant object recognition.

spatial phase invariance

Complex cells also have temporal response characteristics that distinguish them from simple cells. Complex cell responses to moving gratings are approximately constant, not oscillatory as in figures 2.20 and 2.22. The firing rate of a complex cell responding to a counterphase grating oscillating with frequency ω has both a constant component and an oscillatory component with a frequency of 2ω, a phenomenon known as frequency doubling.

frequency doubling

Even though spike-triggered average stimuli and reverse correlation functions fail to capture the response properties of complex cells, complex-cell responses can be described, to a first approximation, by a relatively

straightforward extension of the reverse correlation approach. The key observation comes from equation 2.34, which shows how the linear response estimate of a simple cell depends on spatial phase for an optimally oriented grating with K not too small. Consider two such responses, labeled L_1 and L_2, with preferred spatial phases ϕ and $\phi - \pi/2$. Including both the spatial and temporal response factors, we find, for preferred spatial phase ϕ,

$$L_1 = AB(\omega, K)\cos(\phi - \Phi)\cos(\omega t - \delta) \tag{2.39}$$

where $B(\omega, K)$ is a temporal and spatial frequency-dependent amplitude factor. We do not need the explicit form of $B(\omega, K)$ here, but the reader is urged to derive it. For preferred spatial phase $\phi - \pi/2$,

$$L_2 = AB(\omega, K)\sin(\phi - \Phi)\cos(\omega t - \delta) \tag{2.40}$$

because $\cos(\phi - \pi/2 - \Phi) = \sin(\phi - \Phi)$. If we square and add these two terms, we obtain a result that does not depend on Φ,

$$L_1^2 + L_2^2 = A^2B^2(\omega, K)\cos^2(\omega t - \delta)\,, \tag{2.41}$$

because $\cos^2(\phi - \Phi) + \sin^2(\phi - \Phi) = 1$. Thus, we can describe the response of a complex cell by writing

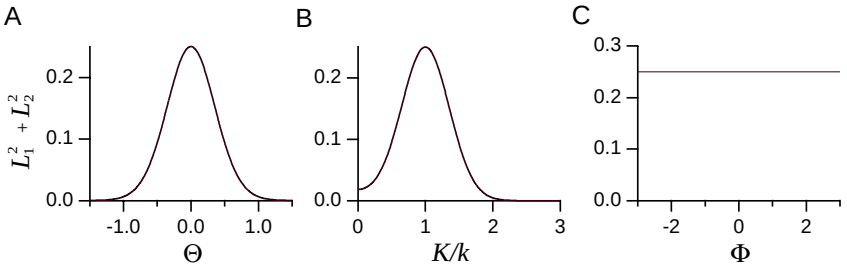

Figure 2.23: Selectivity of a complex cell model in response to a sinusoidal grating. The width and preferred spatial frequency of the Gabor functions underlying the estimated firing rate satisfy $k\sigma = 2$. A) The complex cell response estimate, $L_1^2 + L_2^2$, as a function of stimulus orientation Θ for a grating with the preferred spatial frequency $K = k$. B) $L_1^2 + L_2^2$ as a function of the ratio of the stimulus spatial frequency to its preferred value, K/k, for a grating oriented in the preferred direction $\Theta = 0$. C) $L_1^2 + L_2^2$ as a function of stimulus spatial phase Φ for a grating with the preferred spatial frequency and orientation $K = k$ and $\Theta = 0$.

$$\mathrm{r}(t) = \mathrm{r}_0 + G\left(L_1^2 + L_2^2\right)\,. \tag{2.42}$$

The selectivities of such a response estimate to grating orientation, spatial frequency, and spatial phase are shown in figure 2.23. The response of the model complex cell is tuned to orientation and spatial frequency, but the spatial phase dependence, illustrated for a simple cell in figure 2.15C, is

absent. In computing the curve for figure 2.23C, we used the exact expressions for L_1 and L_2 from the integrals in equations 2.31 and 2.32, not the approximation 2.34 used to simplify the discussion above. Although it is not visible in the figure, there is a weak dependence on Φ when the exact expressions are used.

The complex cell response given by equations 2.42 and 2.41 reproduces the frequency doubling effect seen in complex cell responses because the factor $\cos^2(\omega t - \delta)$ oscillates with frequency 2ω. This follows from the identity

$$\cos^2(\omega t - \delta) = \frac{1}{2}\cos\left(2(\omega t - \delta)\right) + \frac{1}{2}. \tag{2.43}$$

In addition, the last term on the right side of this equation generates the constant component of the complex cell response to a counterphase grating. Figure 2.24 shows a comparison of model simple and complex cell responses to a counterphase grating and illustrates this phenomenon.

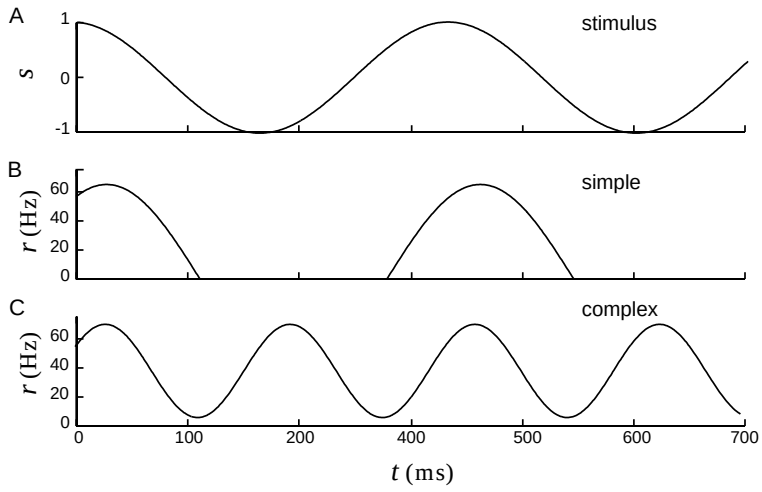

Figure 2.24: Temporal responses of model simple and complex cells to a counterphase grating. A) The stimulus $s(x, y, t)$ at a given point (x, y) plotted as a function of time. B) The rectified linear response estimate of a model simple cell to this grating with a temporal kernel given by equation 2.29 with $\alpha = 1/(15\text{ ms})$. C) The frequency doubled response of a model complex cell with the same temporal kernel but with the estimated rate given by a squaring operation rather than rectification. The background firing rate is $r_0 = 5$ Hz. Note the temporal phase shift of both B and C relative to A.

The description of a complex cell response that we have presented is called an 'energy' model because of its resemblance to the equation for the energy of a simple harmonic oscillator. The pair of linear filters used, with preferred spatial phases separated by $\pi/2$ is called a quadrature pair. Because of rectification, the terms L_1^2 and L_2^2 cannot be constructed by squaring the

energy model

outputs of single simple cells. However, they can each be constructed by summing the squares of rectified outputs from two simple cells with preferred spatial phases separated by π. Thus, we can write the complex cell response as the sum of the squares of four rectified simple cell responses,

$$r(t) = r_0 + G\left([L_1]_+^2 + [L_2]_+^2 + [L_3]_+^2 + [L_4]_+^2\right) , \tag{2.44}$$

where the different $[L]_+$ terms represent the responses of simple cells with preferred spatial phases ϕ, $\phi + \pi/2$, $\phi + \pi$, and $\phi + 3\pi/2$. While such a construction is possible, it should not be interpreted too literally because complex cells receive input from many sources including the LGN and other complex cells. Rather, this model should be viewed as purely descriptive. Mechanistic models of complex cells are described at the end of this chapter and in chapter 7.

2.6 Receptive Fields in the Retina and LGN

We end this discussion of the visual system by returning to the initial stages of the visual pathway and briefly describing the receptive field properties of neurons in the retina and LGN. Retinal ganglion cells display a wide variety of response characteristics, including nonlinear and direction-selective responses. However, a class of retinal ganglion cells (X cells in the cat or P cells in the monkey retina and LGN) can be described by a linear model built using reverse correlation methods. The receptive fields of this class of retinal ganglion cells and an analogous type of LGN relay neurons are similar, so we do not treat them separately. The spatial structure of the receptive fields of these neurons has a center-surround structure consisting either of a circular central ON region surrounded by an annular OFF region, or the opposite arrangement of a central OFF region surrounded by an ON region. Such receptive fields are called ON-center or OFF-center respectively. Figure 2.25A shows the spatial receptive fields of an ON-center cat LGN neuron.

difference of Gaussians

The spatial structure of retinal ganglion and LGN receptive fields is well-captured by a difference-of-Gaussians model in which the spatial receptive field is expressed as

$$D_s(x, y) = \pm\left(\frac{1}{2\pi\sigma_{cen}^2}\exp\left(-\frac{x^2+y^2}{2\sigma_{cen}^2}\right) - \frac{B}{2\pi\sigma_{sur}^2}\exp\left(-\frac{x^2+y^2}{2\sigma_{sur}^2}\right)\right) . \tag{2.45}$$

Here the center of the receptive field has been placed at $x = y = 0$. The first Gaussian function in equation 2.45 describes the center and the second the surround. The size of the central region is determined by the parameter σ_{cen}, while σ_{sur}, which is greater than σ_{cen}, determines the size of the surround. B controls the balance between center and surround contributions.

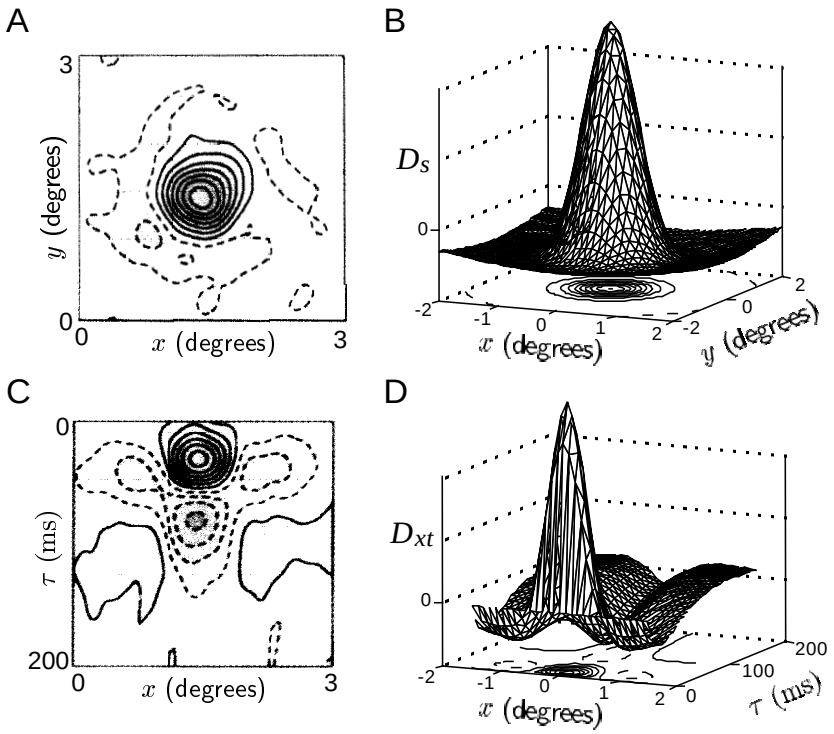

Figure 2.25: Receptive fields of LGN neurons. A) The center-surround spatial structure of the receptive field of a cat LGN X cell. This has a central ON region (solid contours) and a surrounding OFF region (dashed contours). B) A fit of the receptive field shown in A using a difference of Gaussian function (equation 2.45) with $\sigma_{\text{cen}} = 0.3^\circ$, $\sigma_{\text{sur}} = 1.5^\circ$, and $B = 5$. C) The space-time receptive field of a cat LGN X cell. Note that both the center and surround regions reverse sign as a function of τ and that the temporal evolution is slower for the surround than for the center. D) A fit of the space-time receptive field in C using 2.46 with the same parameters for the Gaussian functions as in B, and temporal factors given by equation 2.47 with $1/\alpha_{\text{cen}} = 16$ ms for the center, $1/\alpha_{\text{sur}} = 32$ ms for the surround, and $1/\beta_{\text{cen}} = 1/\beta_{\text{sur}} = 64$ ms. (A and C adapted from DeAngelis et al., 1995.)

The $\pm$ sign allows both ON-center ($+$) and OFF-center ($-$) cases to be represented. Figure 2.25B shows a spatial receptive field formed from the difference of two Gaussians that approximates the receptive field structure in figure 2.25A.

Figure 2.25C shows that the spatial structure of the receptive field reverses over time with, in this case, a central ON region reversing to an OFF region as τ increases. Similarly, the OFF surround region changes to an ON region with increasing τ, although the reversal and the onset are slower for the surround than for the central region. Because of the difference between the time course of the center and surround regions, the space-time receptive field is not separable, although the center and surround components

are individually separable. The basic features of LGN neuron space-time receptive fields are captured by the mathematical caricature

$$D(x,y,\tau) = \pm\left(\frac{D_t^{cen}(\tau)}{2\pi\sigma_{cen}^2}\exp\left(-\frac{x^2+y^2}{2\sigma_{cen}^2}\right) - \frac{BD_t^{sur}(\tau)}{2\pi\sigma_{sur}^2}\exp\left(-\frac{x^2+y^2}{2\sigma_{sur}^2}\right)\right). \tag{2.46}$$

Separate functions of time multiply the center and surround, but they can both be described by the same functions using two sets of parameters,

$$D_t^{cen,sur}(\tau) = \alpha_{cen,sur}^2\tau\exp(-\alpha_{cen,sur}\tau) - \beta_{cen,sur}^2\tau\exp(-\beta_{cen,sur}\tau). \tag{2.47}$$

The parameters α_{cen} and α_{sur} control the latency of the response in the center and surround regions respectively, and β_{cen} and β_{sur} affect the time of the reversal. This function has characteristics similar to the function 2.29, but the latency effect is less pronounced. Figure 2.25D shows the space-time receptive field of equation 2.46 with parameters chosen to match figure 2.25C.

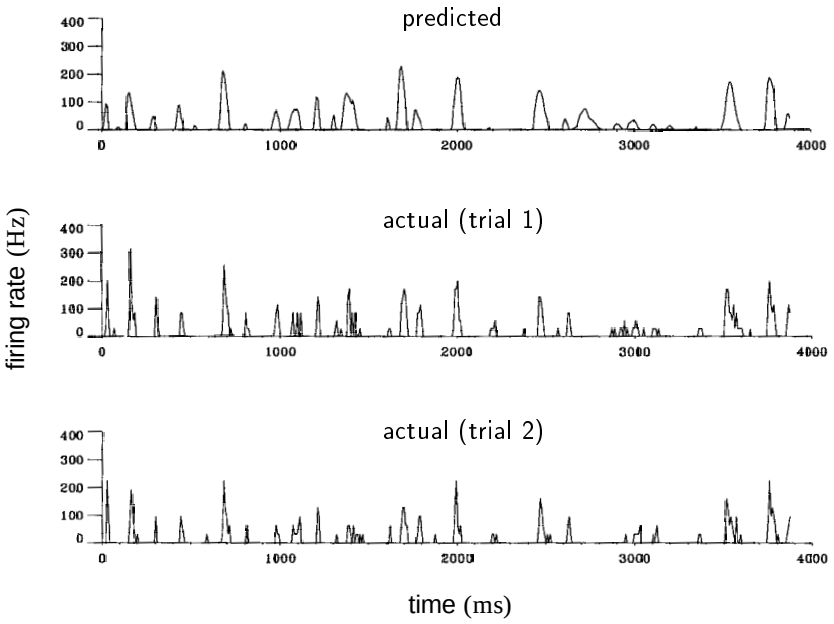

Figure 2.26: Comparison of predicted and measured firing rates for a cat LGN neuron responding to a video movie. The top panel is the rate predicted by integrating the product of the video image intensity and a linear filter obtained for this neuron from a spike-triggered average of a white-noise stimulus. The resulting linear prediction was rectified. The middle and lower panels are measured firing rates extracted from two different sets of trials. (From Dan et al., 1996.)

Figure 2.26 shows the results of a direct test of a reverse correlation model of an LGN neuron. The kernel needed to describe a particular LGN cell was first extracted using a white-noise stimulus. This, together with a rectifying static nonlinearity, was used to predict the firing rate of the neuron

in response to a video movie. The top panel in figure 2.26 shows the resulting prediction while the middle and lower panels show the actual firing rates extracted from two different groups of trials. The correlation coefficient between the predicted and actual firing rates was 0.5, which was very close to the correlation coefficient between firing rates extracted from different groups of trials. This means that the error of the prediction was no worse than the variability of the neural response itself.

2.7 Constructing V1 Receptive Fields

The models of visual receptive fields we have been discussing are purely descriptive, but they provide an important framework for studying how the circuits of the retina, LGN, and primary visual cortex generate neural responses. In an example of a more mechanistic model, Hubel and Wiesel (1962) showed how the oriented receptive fields of cortical neurons could be generated by summing the input from appropriately selected LGN neurons. Their construction, shown in figure 2.27A, consists of alternating rows of ON-center and OFF-center LGN cells providing convergent input to a cortical simple cell. The left side of figure 2.27A shows the spatial arrangement of LGN receptive fields that, when summed, form bands of ON and OFF regions resembling the receptive field of an oriented simple cell. This model accounts for the selectivity of a simple cell purely on the basis of feedforward input from the LGN. We leave the study of this model as an exercise for the reader. Other models, which we discuss in chapter 7, include the effects of recurrent intracortical connections as well.

Hubel-Wiesel simple cell model

In a previous section, we showed how the properties of complex cell responses could be accounted for using a squaring static nonlinearity. While this provides a good description of complex cells, there is little indication that complex cells actually square their inputs. Models of complex cells can be constructed without introducing a squaring nonlinearity. One such example is another model proposed by Hubel and Wiesel (1962), which is depicted in figure 2.27B. Here the phase-invariant response of a complex cell is produced by summing together the responses of several simple cells with similar orientation and spatial frequency tuning, but different preferred spatial phases. In this model, the complex cell inherits its orientation and spatial frequency preference from the simple cells that drive it, but spatial phase selectivity is reduced because the outputs of simple cells with a variety of spatial phases selectivities are summed linearly. Analysis of this model is left as an exercise. While the model generates complex cell responses, there are indications that complex cells in primary visual cortex are not exclusively driven by simple cell input. An alternative model is considered in chapter 7.

Hubel-Wiesel complex cell model

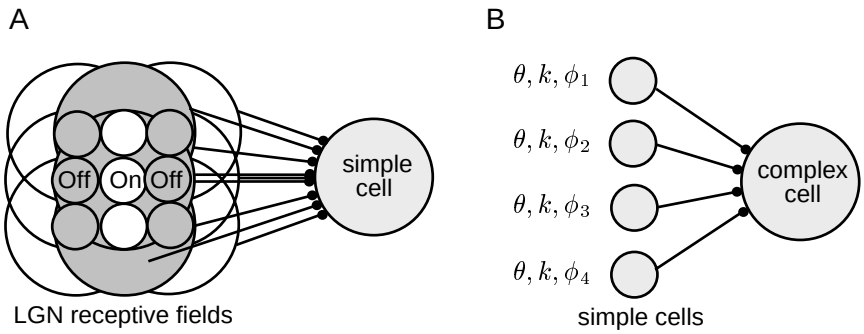

Figure 2.27: A) The Hubel-Wiesel model of orientation selectivity. The spatial arrangement of the receptive fields of nine LGN neurons are shown, with a row of three ON-center fields flanked on either side by rows of three OFF-center fields. White areas denote ON fields and grey areas OFF fields. In the model, the converging LGN inputs are summed linearly by the simple cell. This arrangement produces a receptive field oriented in the vertical direction. B) The Hubel-Wiesel model of a complex cell. Inputs from a number of simple cells with similar orientation and spatial frequency preferences (θ and k), but different spatial phase preferences (ϕ_1, ϕ_2, ϕ_3, and ϕ_4), converge on a complex cell and are summed linearly. This produces a complex cell output that is selective for orientation and spatial frequency, but not for spatial phase. The figure shows four simple cells converging on a complex cell, but additional simple cells can be included to give a more complete coverage of spatial phase.

2.8 Chapter Summary

We continued from chapter 1 our study of the ways that neurons encode information, focusing on reverse-correlation analysis, particularly as applied to neurons in the retina, visual thalamus (LGN), and primary visual cortex. We used the tools of systems identification, especially the linear filter, Wiener kernel, and static nonlinearity to build descriptive linear and nonlinear models of the transformation from dynamic stimuli to time-dependent firing rates. We discussed the complex logarithmic map governing the way that neighborhood relationships in the retina are transformed into cortex, Nyquist sampling in the retina, and Gabor functions as descriptive models of separable and nonseparable receptive fields. Models based on Gabor filters and static nonlinearities were shown to account for the basic response properties of simple and complex cells in primary visual cortex, including selectivity for orientation, spatial frequency and phase, velocity, and direction. Retinal ganglion cell and LGN responses were modeled using a difference-of-Gaussians kernel. We briefly described simple circuit models of simple and complex cells.

2.9 Appendices

A) The Optimal Kernel

Using equation 2.1 for the estimated firing rate, the expression 2.3 to be minimized is

$$E = \frac{1}{T}\int_0^T dt \left(\mathsf{r}_0 + \int_0^\infty d\tau\, D(\tau)s(t-\tau) - \mathsf{r}(t)\right)^2 . \tag{2.48}$$

functional derivative

The minimum is obtained by setting the derivative of E with respect to the function D to zero. A quantity, such as E, that depends on a function, D in this case, is called a functional, and the derivative we need is a functional derivative. Finding the extrema of functionals is the subject of a branch of mathematics called the calculus of variations. A simple way to define a functional derivative is to introduce a small time interval Δt and evaluate all functions at integer multiples of Δt. We define $\mathsf{r}_i = \mathsf{r}(i\Delta t)$, $D_k = D(k\Delta t)$, and $s_{i-k} = s((i-k)\Delta t)$. If Δt is small enough, the integrals in equation 2.48 can be approximated by sums, and we can write

$$E = \frac{\Delta t}{T}\sum_{i=0}^{T/\Delta t}\left(\mathsf{r}_0 + \Delta t\sum_{k=0}^{\infty} D_k s_{i-k} - \mathsf{r}_i\right)^2 . \tag{2.49}$$

E is minimized by setting its derivative with respect to D_j for all values of j to zero,

$$\frac{\partial E}{\partial D_j} = 0 = \frac{2\Delta t}{T}\sum_{i=0}^{T/\Delta t}\left(\mathsf{r}_0 + \Delta t\sum_{k=0}^{\infty} D_k s_{i-k} - \mathsf{r}_i\right) s_{i-j}\Delta t\,. \tag{2.50}$$

Rearranging and simplifying this expression gives the condition

$$\Delta t\sum_{k=0}^{\infty} D_k\left(\frac{\Delta t}{T}\sum_{i=0}^{T/\Delta t} s_{i-k}s_{i-j}\right) = \frac{\Delta t}{T}\sum_{i=0}^{T/\Delta t}(\mathsf{r}_i - \mathsf{r}_0)\, s_{i-j}\,. \tag{2.51}$$

If we take the limit $\Delta t \to 0$ and make the replacements $i\Delta t \to t$, $j\Delta t \to \tau$, and $k\Delta t \to \tau'$, the sums in equation 2.51 turn back into integrals, the indexed variables become functions, and we find

$$\int_0^\infty d\tau'\, D(\tau')\left(\frac{1}{T}\int_0^T dt\, s(t-\tau')s(t-\tau)\right) = \frac{1}{T}\int_0^T dt\,(\mathsf{r}(t) - \mathsf{r}_0)\, s(t-\tau)\,. \tag{2.52}$$

The term proportional to r_0 on the right side of this equation can be dropped because the time integral of s is zero. The remaining term is the firing rate-stimulus correlation function evaluated at $-\tau$, $Q_{rs}(-\tau)$. The term in large parentheses on the left side of 2.52 is the stimulus autocorrelation function. By shifting the integration variable $t \to t+\tau$, we find that it is $Q_{ss}(\tau - \tau')$, so 2.52 can be re-expressed in the form of equation 2.4.

Equation 2.6 provides the solution to equation 2.4 only for a white noise stimulus. For an arbitrary stimulus, equation 2.4 can be solved easily by the method of Fourier transforms if we ignore causality and allow the estimated rate at time t to depend on the stimulus at times later than t, so that

$$r_{est}(t) = r_0 + \int_{-\infty}^{\infty} d\tau\, D(\tau) s(t-\tau)\,. \tag{2.53}$$

The estimate written in this acausal form, satisfies a slightly modified version of equation 2.4,

$$\int_{-\infty}^{\infty} d\tau'\, Q_{ss}(\tau-\tau') D(\tau') = Q_{rs}(-\tau)\,. \tag{2.54}$$

We define the Fourier transforms (see the Mathematical Appendix)

$$\tilde{D}(\omega) = \int_{-\infty}^{\infty} dt\, D(t) \exp(i\omega t) \quad \text{and} \quad \tilde{Q}_{ss}(\omega) = \int_{-\infty}^{\infty} d\tau\, Q_{ss}(\tau) \exp(i\omega\tau) \tag{2.55}$$

as well as $\tilde{Q}_{rs}(\omega)$ defined analogously to $\tilde{Q}_{ss}(\omega)$.

Equation 2.54 is solved by taking the Fourier transform of both sides and using the convolution identity (Mathematical Appendix)

$$\int_{-\infty}^{\infty} dt\, \exp(i\omega t) \int_{-\infty}^{\infty} d\tau'\, Q_{ss}(\tau-\tau') D(\tau') = \tilde{D}(\omega)\tilde{Q}_{ss}(\omega) \tag{2.56}$$

In terms of the Fourier transforms, equation 2.54 then becomes

$$\tilde{D}(\omega)\tilde{Q}_{ss}(\omega) = \tilde{Q}_{rs}(-\omega) \tag{2.57}$$

which can be solved directly to obtain $\tilde{D}(\omega) = \tilde{Q}_{rs}(-\omega)/\tilde{Q}_{ss}(\omega)$. The inverse Fourier transform from which $D(\tau)$ is recovered is (Mathematical Appendix)

$$D(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \tilde{D}(\omega) \exp(-i\omega\tau)\,, \tag{2.58}$$

so the optimal acausal kernel when the stimulus is temporally correlated is given by

$$D(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \frac{\tilde{Q}_{rs}(-\omega)}{\tilde{Q}_{ss}(\omega)} \exp(-i\omega\tau)\,. \tag{2.59}$$

B) The Most Effective Stimulus

We seek the stimulus that produces the maximum predicted responses at time t subject to the fixed energy constraint

$$\int_0^T dt' \left(s(t')\right)^2 = \text{constant}\,. \tag{2.60}$$

We impose this constraint by the method of Lagrange multipliers (see the Mathematical Appendix), which means that we must find the unconstrained maximum value with respect to s of

$$r_{est}(t) + \lambda \int_0^T dt' \, s^2(t') = r_0 + \int_0^\infty d\tau \, D(\tau) s(t-\tau) + \lambda \int_0^T dt' \, \left(s(t')\right)^2 \quad (2.61)$$

where λ is the Lagrange multiplier. Setting the derivative of this expression with respect to the function s to zero (using the same methods used in appendix A) gives

$$D(\tau) = -2\lambda s(t-\tau) \, . \quad (2.62)$$

The value of λ (which is less than zero) is determined by requiring that condition 2.60 is satisfied, but the precise value is not important for our purposes. The essential result is the proportionality between the optimal stimulus and $D(\tau)$.

C) Bussgang's Theorem

Bussgang (1952 & 1975) proved that an estimate based on the optimal kernel for linear estimation can still be self-consistent (although not necessarily optimal) when nonlinearities are present. The self-consistency condition is that when the nonlinear estimate $r_{est} = r_0 + F(L(t))$ is substituted into equation 2.6, the relationship between the linear kernel and the firing rate-stimulus correlation function should still hold. In other words, we require that

$$D(\tau) = \frac{1}{\sigma_s^2 T} \int_0^T dt \, r_{est}(t) s(\tau - t) = \frac{1}{\sigma_s^2 T} \int_0^T dt \, F(L(t)) s(\tau - t) \, . \quad (2.63)$$

We have dropped the r_0 term because the time integral of s is zero. In general, equation 2.63 does not hold, but if the stimulus used to extract D is Gaussian white noise, equation 2.63 reduces to a simple normalization condition on the function F. This result is based on the identity, valid for a Gaussian white-noise stimulus,

$$\frac{1}{\sigma_s^2 T} \int_0^T dt \, F(L(t)) s(\tau - t) = \frac{D(\tau)}{T} \int_0^T dt \, \frac{dF(L(t))}{dL} \, . \quad (2.64)$$

For the right side of this equation to be $D(\tau)$, the remaining expression, involving the integral of the derivative of F, must be equal to one. This can be achieved by appropriate scaling of F. The critical identity 2.64 is based on integration by parts for a Gaussian weighted integral. A simplified proof is left as an exercise.

2.10 Annotated Bibliography

Marmarelis & Marmarelis (1978), **Rieke et al. (1997)** and **Gabbiani & Koch** (1998) provide general discussions of reverse correlation methods. A useful reference relevant to our presentation of their application to the visual system is **Carandini et al. (1996)**. Volterra and Wiener functional expansions are discussed in **Wiener (1958)** and **Marmarelis & Marmarelis (1978)**.

General introductions to the visual system include **Hubel & Wiesel (1962, 1977)**, **Orban (1984)**, **Hubel (1988)**, **Wandell (1995)**, and **De Valois & De Valois (1990)**. Our treatment follows **Dowling (1987)** on processing in the retina, and Schwartz (1977), Van Essen et al. (1984), and Rovamo & Virsu (1984) on aspects of the retinotopic map from the eye to the brain. Properties of this map are used to account for aspects of visual hallucinations in Ermentrout & Cowan (1979). We also follow **Movshon et al. (1978a & b)** for definitions of simple and complex cells; Daugman (1985) and Jones & Palmer (1987b) on the use of Gabor functions (Gabor, 1946) to describe visual receptive fields; and **DeAngelis et al. (1995)** on space-time receptive fields. Our description of the energy model of complex cells is based on Adelson & Bergen (1985), which is related to work by Pollen & Ronner (1982), Van Santen & Sperling (1984), and Watson & Ahumada (1985), and to earlier ideas of Reichardt (1961) and Barlow & Levick (1965). Heeger's (1992; 1993) model of contrast saturation is reviewed in **Carandini et al. (1996)** and has been applied in a approach more closely related to the representational learning models of chapter 10 by Simoncelli & Schwartz (1999). The difference-of-Gaussians model for retinal and LGN receptive fields is due to Rodieck (1965) and Enroth-Cugell and Robson (1966). A useful reference to modeling of the early visual system is Wörgötter & Koch (1991). The issue of linearity and non-linearity in early visual processing is reviewed by **Ferster (1994)**.