Chapter 1

Neural Encoding

1.1 From Stimulus to Response

Neurons are unique among the cells of the body in their ability to propagate signals rapidly over large distances. They do this by generating characteristic electrical pulses called action potentials, or more simply, spikes that can travel down nerve fibers. Neurons represent and transmit information by firing sequences of spikes in various temporal patterns. The study of neural coding involves measuring and characterizing how stimulus attributes, such as light intensity, image orientation, and movement direction; or motor actions, such as the direction of an arm movement, are represented by spikes. As an investigation of the general relationship between stimulus and response, it is a branch of the broader field of systems identification. The link between stimulus and response can be studied from two opposite points of view. Neural encoding refers to the map from stimulus to response. For example, we can catalogue how neurons respond to a wide variety of stimuli, and then construct models that attempt to predict responses to other stimuli. Neural decoding refers to the reverse map, from response to stimulus, and the challenge is to reconstruct a stimulus from the spike sequence that it evoked. Neural decoding is discussed in Chapter 3.

Characterizing the relationship between stimulus and response is difficult because neuronal responses are complex and variable. Neurons typically respond by producing irregular spike sequences, and responses can change over time even when a stimulus is held fixed. Distinguishing features that reflect the intrinsic response dynamics from those that encode the temporal characteristics of a stimulus can be difficult, especially if the time scale for significant changes in the stimulus is of the same order as the average interval between spikes. Neural responses vary from trial to trial even when the same stimulus is presented repeatedly. The source of this variability is not known, but candidates include variable levels of arousal and attention, randomness associated with various biophysical processes that affect neuronal firing, and the effects of other cognitive processes taking place during a trial. Given the complexity and variability of neural responses, many trials may be required to determine the complete relationship between stimulus and response. The number of spike sequences that can be distinguished, even with a fairly coarse degree of temporal resolution, is enormous. Exhaustively cataloging all possible response sequences and determining their probabilities of occurrence is often impossible.

Typically, many neurons respond to a given stimulus, and stimulus features are encoded by the activities of large neural populations. In studying population coding we must examine relationships of the firing patterns of the responding neurons to each other as well as the firing patterns of individual neurons.

The complexity and trial-to-trial variability of action potential sequences make it unlikely that we can describe and predict the timing of each spike deterministically. Rather we seek a model that can account for the probabilities that different spike sequences are evoked by a given stimulus. Modeling action potential sequences is a two-step process; the sequence probability is computed deterministically, and then a stochastic procedure generates the spikes from the computed probability. In the models we discuss, the spiking probability will be characterized by a time-dependent firing rate $r(t)$. We describe procedures for constructing an estimate $r_{\mathrm{est}}(t)$ of $r(t)$ deterministically. This estimate can be used to drive a stochastic spike generator. In chapter 2, we discuss how this process can be applied to visually responsive neurons in the retina, lateral geniculate nucleus, and primary visual cortex.

Recording Neuronal Responses

An action potential is a roughly 100 mV fluctuation in the electrical potential across the cell membrane of a neuron lasting for around 1 ms. In general, membrane potential fluctuations are divided into action potentials and subthreshold potentials, the component of the membrane potential below the threshold for action potential initiation. Both action potential and subthreshold membrane potentials can be measured by inserting a sharp, hollow glass electrode filled with an electrolyte into a neuron, or by attaching a broader tipped 'patch' electrode to the outer surface of the cell and breaking or perforating the membrane under the electrode. The top trace in figure 1.1 is a schematic of an intracellular recording from the soma, or cell body, of a neuron firing a sequence of action potentials. The recording shows rapid spikes riding on top of a more slowly varying subthreshold potential. The bottom trace in this figure is a schematic of an intracellular recording made some distance out on the axon of this neuron. (These traces are drawings, not real recordings, because such intracellular axon recordings, although possible in some types of cells, are difficult

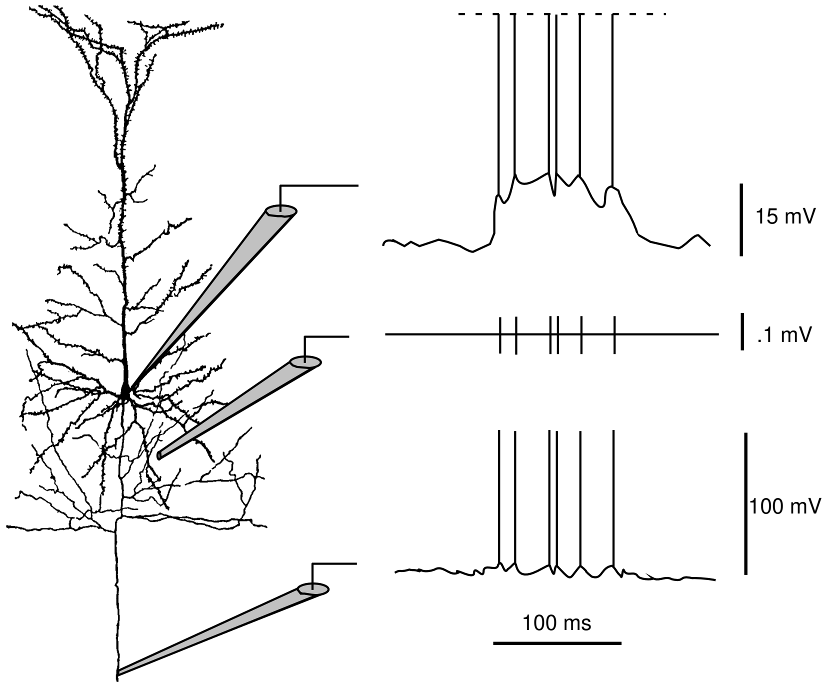

Figure 1.1: Three simulated recordings from a neuron. The top trace is from an intracellular electrode inserted into the soma of the neuron. The height of the action potentials has been clipped to show the subthreshold membrane potential more clearly. The time scale is such that the action potential trajectory cannot be resolved. The bottom trace is from an intracellular electrode placed inside the axon some distance away from the soma. The full height of the action potentials is indicated in this trace. The middle trace is a simulated extracellular recording. Action potentials appear as roughly equal positive and negative potential fluctuations with an amplitude of around 0.1 mV. This is roughly 1000 times smaller than the approximately 0.1 V amplitude of an intracellularly recorded action potential. (Neuron drawing by Cajal, 1911 from Dowling, 1992)

and rarely made.) The subthreshold membrane potential envelope, apparent in the soma recording, is completely absent on the axon, but the action potential sequence in the two recordings is identical. Subthreshold membrane potential fluctuations can only propagate for about 1 mm before becoming severely attenuated. Signals within the nervous system traveling over larger distances are carried exclusively by action potentials.

The middle trace in figure 1.1 illustrates an extracellular recording. Here an electrode is placed near a neuron but it does not penetrate the cell membrane. Such recordings only show the action potentials fired by a neuron. Extracellular recordings are typically used for *in vivo* experiments, especially those involving behaving animals. Intracellular recordings are sometimes made *in vivo*, but are more commonly used for *in vitro* preparations such as experiments on slices of neural tissue. The response studied in this chapter are action potential sequences that can be recorded extracellularly.

Firing Rates: Describing the Response

While everyone has an intuitive idea of what a firing rate is, the term has no unique definition. Figure 1.2 compares a number of ways of defining and extracting a time-dependent firing rate from a spike sequence. Figure 1.2A shows three seconds of the response of a neuron in the inferior temporal cortex recorded while a monkey watched a video. A simple way of extracting a firing rate from a spike train like this is to divide time into discrete bins of duration Δt, count the number of spikes within each bin, and divide by Δt. Figure 1.2B shows the firing rate computed using this procedure with a bin size of 100 ms.

The binning and counting procedure illustrated in figure 1.2B generates a firing rate that is a piecewise constant function of time, resembling a histogram. Because spike counts can only take integer values, the rates computed by this method will always be integer multiples of $1/\Delta t$, and thus they take discrete values. Decreasing the value of Δt increases temporal resolution by providing an estimate of the firing rate at more finely spaced intervals of time, but at the expense of decreasing the resolution for distinguishing different rates. The product of the temporal resolution Δt and the rate resolution $1/\Delta t$ is one, no matter what value of Δt is used. One way to avoid quantized firing rates is to vary the bin size so that a fixed number of spikes appears in each bin. The firing rate is then defined as that fixed number of spikes divided by the variable bin width.

Counting spikes in preassigned bins produces a firing rate that depends not only on the size of the time bins, but also on their placement. To avoid the arbitrariness in the placement of bins, we can instead take a single bin or window of duration Δt and slide it along the spike train, counting the number of spikes within the window at each location. The jagged curve in figure 1.2C shows the result of sliding a 100 ms wide window along the spike train. The firing rate defined in this way can be expressed as the sum of a window function over the times t_i for $i = 1, 2, \ldots, n$ when the n spikes in a particular sequence occurred,

$$r(t) = \sum_{i=1}^{n} f_w(t - t_i) \tag{1.1}$$

with the window function

$$f_w(t) = \begin{cases} 1/\Delta t & \text{if } -\Delta t/2 \le t < \Delta t/2 \\ 0 & \text{otherwise}\,. \end{cases} \tag{1.2}$$

Use of a sliding window avoids the arbitrariness of bin placement and produces a rate that might appear to have a better temporal resolution.

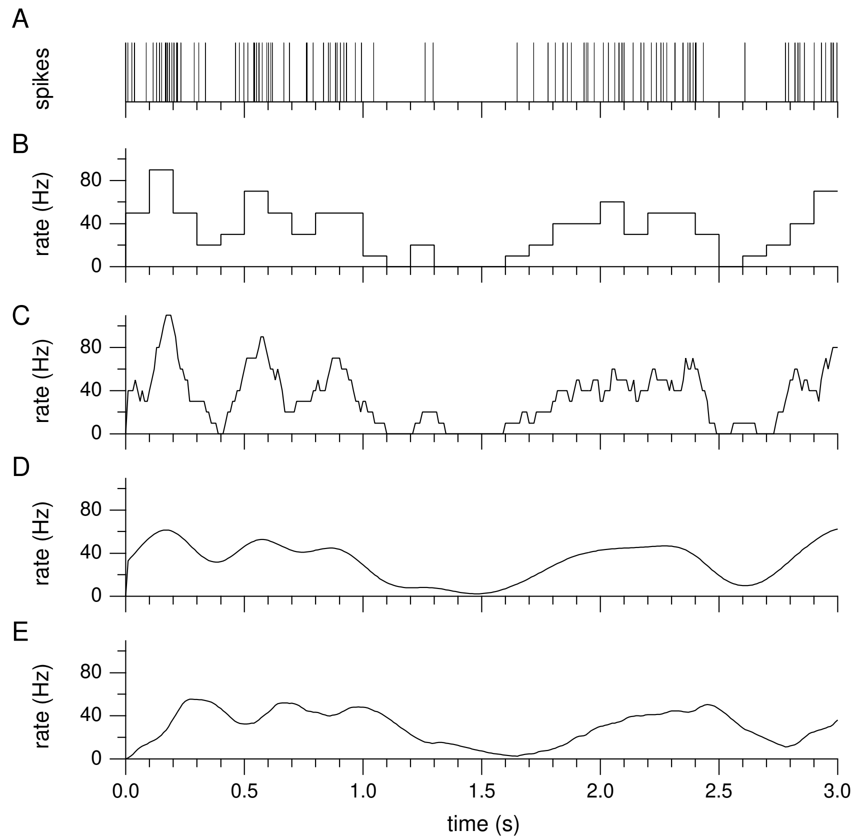

Figure 1.2: Single-trial firing rates defined by different procedures. A) A spike train from a neuron in the inferior temporal cortex of a monkey recorded while that animal watched a video on a monitor under free viewing conditions. B) Discrete-time firing rate obtained by binning time and counting spikes with $\Delta t = 100$ ms. C) Firing rate determined by sliding a rectangular window function along the spike train with $\Delta t = 100$ ms. D) Firing rate computed using a Gaussian window function with $\sigma_t = 100$ ms. E) Firing rate for an α function window with $1/\alpha = 100$ ms. (Data from Baddeley *et al*, 1997)

However, it must be remembered that the rates obtained at times separated by less than one bin width are correlated because they involve some of the same spikes.

The sum in equation 1.1 can also be written as a integral of the window function times a quantity $\rho(t)$ which we call the neural response function,

$$r(t) = \int_{-\infty}^{\infty} d\tau \, f_w(t-\tau)\rho(\tau) \, . \tag{1.3}$$

The neural response function is a sum of infinitesimally narrow, idealized spikes in the form of Dirac δ functions,

$$\rho(t) = \sum_{i=1}^{n} \delta(t - t_i) \,. \tag{1.4}$$

The equivalence of 1.1 and 1.3 follows from the basic defining equation for a δ function,

$$\int_{-\infty}^{\infty} d\tau \, \delta(t - \tau) f(\tau) = f(t) \,, \tag{1.5}$$

which holds for any function f. The neural response function is useful for re-expressing sums over spikes as integrals over time.

The integral in equation 1.3 is called a linear filter, and the window function f_w is the filter kernel. Linear filters play an important role throughout the first section of this book. We typically write them in a slightly different form from equation 1.3 obtained by making a change of integration variable $\tau \to t - \tau$ in equation 1.3,

$$r(t) = \int_{-\infty}^{\infty} d\tau \, f_w(\tau) \rho(t - \tau) \,. \tag{1.6}$$

Written this way, we can see that the kernel $f_w(\tau)$ determines how strongly the neural response function evaluated at time $t - \tau$ contributes to the firing rate defined at time t. The firing rate is computed by summing, or more properly integrating, contributions arising from all possible τ values.

The jagged appearance of the curve in figure 1.2C is caused by the discontinuous shape of the window function used. A firing rate can be computed using virtually any window function $f_w(\tau)$ that goes to zero outside a region near $\tau = 0$ provided only that its time integral is equal to one,

$$\int_{-\infty}^{\infty} d\tau \, f_w(\tau) = 1 \,. \tag{1.7}$$

For example, instead of the rectangular window function used in figure 1.2C, f_w can be a Gaussian

$$f_w(\tau) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left(-\frac{\tau^2}{2\sigma_w^2}\right) \,. \tag{1.8}$$

In this case, σ_w controls the temporal resolution of the resulting rate, playing a role analogous to Δt. A continuous window function like the Gaussian used in equation 1.1 generates a firing rate that is a smooth function of time (figure 1.2D).

Both the rectangular window and the Gaussian define the firing rate at any time using spikes fired both before and after that time. Of course, a postsynaptic neuron monitoring the spike train of a presynaptic cell only has access to spikes that have already occurred. A firing rate at time t that only depends on spikes fired before t can be defined by using a window function that vanishes when its argument is negative. A window function or kernel that satisfies this condition is called causal. One commonly used form is the α function

$$f_w(\tau) = [\alpha^2 \tau \exp(-\alpha\tau)]_+ \tag{1.9}$$

where $1/\alpha$ determines the temporal resolution of the resulting firing rate. The notation $[z]_+$ for any quantity z stands for the rectification operation,

$$[z]_+ = \begin{cases} z & \text{if } z \geq 0 \\ 0 & \text{otherwise}\,. \end{cases} \tag{1.10}$$

Figure 1.2E shows the firing rate computed from such a causal scheme. Note that the rate tends to peak later than the rate computed in 1.2D where a temporally symmetric window function was used.

A firing rate determined from equation 1.6 using any reasonable window function will change from one trial to the next because the spike times vary. This variability can be eliminated by averaging over large numbers of trials. We use angle brackets $\langle\ \rangle$ to denote averages over trials that use the same stimulus. In other words, $\langle z \rangle$ for any quantity z is the sum of the z values obtained from many trials involving the same stimulus, divided by the number of trials. The trial-averaged firing rate corresponding to 1.6 is

$$\langle r(t) \rangle = \int d\tau\, f_w(\tau) \langle \rho(t-\tau) \rangle\,. \tag{1.11}$$

Note that, because trial averaging and integration are both linear operations, we can interchange the order of averaging and integration in equation 1.11. If the average involves a sufficient number of trials, the averaged neural response function in equation 1.11 can be replaced by a probability density reflecting the relative probability of spike occurrence as a function of time.

Because spike times are continuous variables, some care must be exercised in discussing probabilities of spike occurrences. The probability for a spike to occur at any precisely specified time is zero. To get a nonzero value, we must ask for the probability that a spike occurs within a specified interval, for example the interval between times t and $t + \Delta t$. For small Δt, the probability of a spike falling in this interval is proportional to the size of the interval, Δt, and we can write it as $p[t]\Delta t$ where $p[t]$ is called the single-spike probability density. Throughout this book, we use the notation $P[\,]$ to denote probabilities and $p[\,]$ to denote probability densities. We use the notation $P[\,]$ generically for the probability of something occuring and also to denote a specific probabilility function. In the later case, the notation $P(\,)$ would be more appropriate, but switching between square brackets and parentheses is confusing, so the reader will have to use the context to distinguish between these cases.

In an expression such as 1.11, the neural response function generates a contribution to the integral whenever a spike occurs. The trial-average neural response function generates a contribution proportional to the fraction of trials on which a spike occurred at a given time. According to the law of large numbers, the fraction of trials on which a spike occurs is equal to the probability of a spike occuring on any given trial, if large numbers of trials are being consider. Thus, within integrals such as 1.11, we can replace the trial-averaged neural response function with the single-spike probability density, $\langle \rho(t) \rangle \to p[t]$, and write

$$\langle r(t) \rangle = \int_{-\infty}^{\infty} d\tau \, f_w(\tau) p[t-\tau] \,. \tag{1.12}$$

This establishes an important relationship between the average firing rate $\langle r(t) \rangle$ and the single-spike probability density $p[t]$; the average firing rate is a temporally smoothed, or filtered, version of the single-spike probability density. Window functions typically have a single narrow peak near $\tau = 0$ and fall rapidly to zero away from this peak. If $p[t-\tau]$ is approximately constant over the range τ values where $f_w(\tau)$ is appreciably different from zero, we can write $f_w(t-\tau)p[\tau] \approx f_w(t-\tau)p[t]$ to obtain

$$\langle r(t) \rangle \approx \int_{-\infty}^{\infty} d\tau \, f_w(t-\tau) p[t] = p[t] \tag{1.13}$$

by condition 1.7. We assume that the firing rates we discuss obey this identity (the window function can always be chosen to make this the case unless $p[t]$ varies arbitrarily quickly) so that, the probability of a spike occurring in an interval from t to $t + \Delta t$ for sufficiently small Δt, can be approximated as $\langle r(t) \rangle \Delta t$. We will use this result to generate spike sequences from estimates of the time-dependent firing rate.

Describing the Stimulus

Neural responses typically depend on many different properties of the stimuli that drive a neuron. In this chapter, we characterize responses of neurons as functions of just one of the stimulus attributes to which they may be sensitive. The value of this single attribute at time t is denoted by $S(t)$. In chapter 2, we consider more complete stimulus characterizations.

Neurons responding to sensory stimuli face the difficult task of encoding parameters that can vary over an enormous dynamic range. For example, photoreceptors in the retina can respond to single photons or can operate in bright light with an influx of 10^5 photons/s. To deal with such wide-ranging stimuli, sensory neurons often respond most strongly to rapid changes in stimulus properties and are relatively insensitive to steady-state levels. Steady-state responses are highly compressed functions of stimulus intensity, typically with logarithmic or weak power-law dependences. The psychophysical correlate of this compression is the Weber-Fechner law stating that the perceived level of stimulation is a logarithmic function of actual stimulus intensity.

Sensory systems make numerous adaptations to adjust to the average level of stimulation using a variety of mechanisms. To explore the relationship between stimulus and response in a potentially simpler regime, we will concentrate on describing responses to fluctuations about a mean stimulus level. We separate the average value from fluctuations by writing the stimulus as the sum of two terms,

$$S(t) = \overline{S} + s(t)\,, \tag{1.14}$$

where $\overline{S}$ is the average value of $S(t)$ over time, and $s(t)$ represents fluctuations about this mean and has average value zero. If the stimulus is presented during the interval $0 \leq t \leq T$, the relevant time averages are

$$\overline{S} = \frac{1}{T}\int_0^T dt\, S(t) \quad \text{and} \quad \frac{1}{T}\int_0^T dt\, s(t) = 0\,. \tag{1.15}$$

By convention, a stimulus written in lower case as $s(t)$ has zero average value.

Our analysis of neural encoding will involve two different types of averages: averages over repeated trials that involve the same stimulus, which we have denoted by angle brackets, and averages over different stimuli. We could introduce a second notation for averages over stimuli, but this can be avoided by using time-dependent stimuli. Instead of presenting a number of different stimuli and averaging over them, we can string together all of the stimuli we want to consider into a single time-dependent stimulus sequence and average over time. Thus, stimulus averages are replaced by time averages. We do this solely for notational convenience. If explicit stimulus averaging is necessary, the time-averages in our equations should be replaced by averages over stimuli. It is convenient if integrals involving the stimulus are time-translationally invariant so that for any function F and any number τ

$$\int_0^T dt\, F(s(t+\tau)) = \int_\tau^{T+\tau} dt\, F(s(t)) = \int_0^T dt\, F(s(t))\,. \tag{1.16}$$

To assure this we define the stimulus outside the time limits of the trial by the relation $s(T+\tau) = s(\tau)$ for any τ.

Tuning Curves

A simple way to characterize the response of a neuron is to count the number of action potentials fired during the presentation of a stimulus and define the firing rate as this spike count divided by the duration of the presentation. This approach is most appropriate if the parameter s characterizing the stimulus is held constant over the spike-counting time period. The average of this firing rate over trials, $\langle r \rangle = f(s)$, written as a function of s is called the neural response tuning curve. The functional form of a tuning curve depends on the parameter s used to describe the stimulus. The choice of parameters used as arguments of tuning curve functions is largely a matter of convention. One rationale for choosing a particular parameter s is that it results in a tuning curve of a reasonable or desirable shape. Since tuning curves correspond to firing rates, they are measured in units of spikes per second or Hz.

Figure 1.3A, taken from a classic series of experiments by Hubel and Wiesel (1968), shows extracellular recordings of a neuron in the primary visual cortex (V1) of a monkey. While these recordings were being made, a bar of light was moved at different angles across the region of the visual field where the cell responded to light. This region is called the receptive field of the neuron. Note that the number of action potentials fired depends on the angle of orientation of the bar. The same effect is shown in figure 1.3B in the form of a response tuning curve, which indicates how the average firing rate depends on the orientation of the light bar stimulus. The data

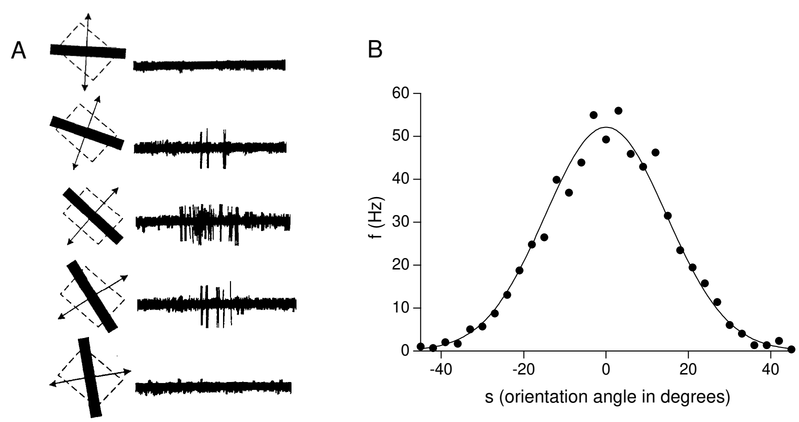

Figure 1.3: A) Recordings from a neuron in the primary visual cortex of a monkey. A bar of light was moved across the receptive field of the cell at different angles. The diagrams to the left of each trace show the receptive field as a dashed square and the light source as a black bar. The angle of the bar indicates the orientation of the light bar for the corresponding trace. B) Average firing rate of a cat V1 neuron plotted as a function of the orientation angle of the light bar stimulus. The curve is a fit using the function 1.17 with parameters $r_{\text{max}} = 52.14$ Hz, $s_{\text{max}} = 0^\circ$, and $\sigma_f = 14.73^\circ$. (A from Hubel and Wiesel, 1968; adapted from Wandell, 1995. B data points from Henry *et al* (1974).)

have been fit by a response tuning curve of the form

$$f(s) = r_{\text{max}} \exp\left(-\frac{1}{2}\left(\frac{s - s_{\text{max}}}{\sigma_f}\right)^2\right) \tag{1.17}$$

where s is the orientation angle of the light bar, s_{max} is the orientation angle evoking the maximum average response rate r_{max} (with $s - s_{\text{max}}$ taken to lie in the range between -90° and +90°), and σ_f determines the width of the tuning curve. Because $f(s)$ takes its maximum value when $s = s_{\max}$, we call $s_{\max}$ the preferred orientation angle of the neuron. It has been defined to be $0°$ in this example. Gaussian functions like 1.17 are frequently used as approximations for firing-rate tuning curves.

Response tuning curves can be used to characterize the selectivities of neurons in visual and other sensory areas to a variety of stimulus parameters. Tuning curves can also be measured for neurons in motor areas, in which case the average firing rate is expressed as a function of one or more parameters describing a motor action. Georgopoulos and collaborators (1982 & 1988) have recorded responses from neurons in different motor areas of the brain after training a monkey to perform a reaching task involving arm movements in different directions. Figure 1.4A show extracellular recordings from a neuron in the primary motor cortex (M1) made while the monkey performed arm reaching movements in different directions. The firing pattern of the cell, in particular the rate at which spikes are generated, is correlated with the direction of arm movement and thus encodes information about this aspect of the motor action.

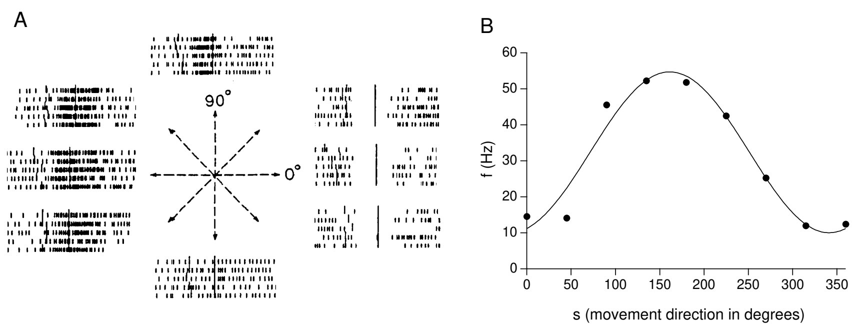

Figure 1.4: Recordings from the primary motor cortex of a monkey performing an arm reaching task. The hand of the monkey started from a central resting location and reaching movements were made in the directions indicated by the arrows. The five stacked traces for each direction show action potentials fired on five different trials. B) Average firing rate of another neuron in the primary motor cortex of a monkey plotted as a function of the direction in which the monkey moved its arm. The curve is a fit using the function 1.18 with parameters $r_{\max} = 54.69$ Hz, $r_0 = 32.34$ Hz, and $s_{\max} = 161.25°$.(A adapted from Georgopoulos *et al*, 1982 which is also the source of the data points in B)

Figure 1.4B shows the response tuning curve of a neuron in the primary motor cortex of a monkey recorded while the animal made arm movements in various directions. Here the data points have been fit by a tuning curve of the form

$$f(s) = r_0 + (r_{\max} - r_0)\cos(s - s_{\max}) \tag{1.18}$$

where s is the reaching angle of the arm, $s_{\max}$ is the reaching angle that produces the maximum response $r_{\max}$, and r_0 is an offset or background firing rate that shifts the tuning curve up from the zero axis. Cosine tuning functions are another form often used to fit firing-rate data. The minimum firing rate predicted by equation 1.18 is $2r_0 - r_{\max}$. For the neuron of figure 1.3B, this is a positive quantity, but for some M1 neurons $2r_0 - r_{\max} < 0$, and the function 1.18 is negative over some range of angles. Since firing rates cannot be negative, the cosine tuning curve must be clipped or rectified in these cases,

$$f(s) = [r_0 + (r_{\max} - r_0)\cos(s - s_{\max})]_+ \,. \tag{1.19}$$

Figure 1.5B shows how the average firing rate of a V1 neuron depends on retinal disparity and illustrates another important type of tuning curve.

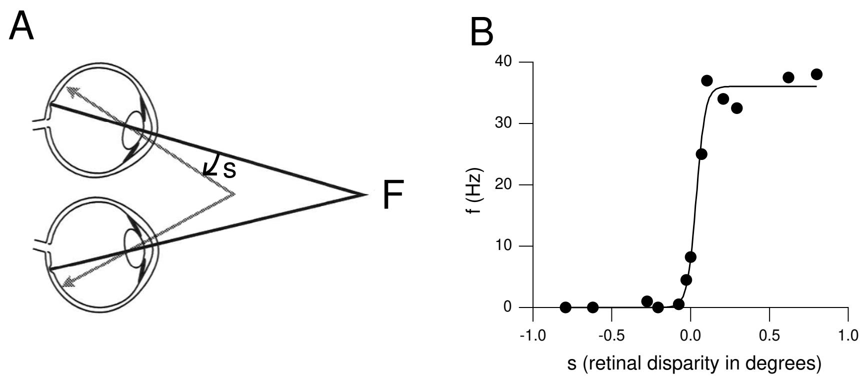

Figure 1.5: A) Definition of retinal disparity. The grey lines show the location on each retina of an object located nearer than the fixation point F. The image from the fixation point falls at the fovea in each eye, the small pit where the black lines meet the retina. The image from a nearer object falls to the left of the fovea in the left eye and to the right of the fovea in the right eye. For objects further away than the fixation point, this would be reversed. The disparity angle s is indicated in the figure. (Adapted from Wandell, 1995) B) Average firing rate of a cat V1 neuron responding to separate bars of light illuminating each eye plotted as a function of the disparity. Because this neuron fires for positive s values it is called a far-tuned cell. The data points are from Poggio and Talbot (1981). The curve is a fit using the function 1.20 with parameters $r_{\max} = 36.03$ Hz, $s_{1/2} = 0.036^\circ$, and $\Delta_s = 0.029^\circ$.

Retinal disparity is a difference in the location of an image between the two eyes (figure 1.5A). Some neurons in area V1 are sensitive to disparity, representing an early stage in the representation of viewing distance. In figure 1.5B, the data points have been fit with a tuning curve called a sigmoidal function,

$$f(s) = \frac{r_{\max}}{1 + \exp\left((s_{1/2} - s)/\Delta_s\right)} \,. \tag{1.20}$$

In this case, s is the retinal disparity, the parameter $s_{1/2}$ is the disparity that produces a firing rate half as big as the maximum value $r_{\max}$, and Δ_s controls how quickly the firing rate increases as a function of s. If Δ_s is negative, the firing rate will be a monotonically decreasing function of s rather than a monotonically increasing function.

Firing-Rate Variability

Tuning curves allow us to predict the firing rate $\langle r \rangle = f(s)$ averaged over trials, but they do not describe how r varies about its mean value $f(s)$ from trial to trial. As we have mentioned, while the map from stimulus to average response may be described deterministically, single-trial responses can probably only be modeled in a probabilistic manner. One way to build a probabilistic model is to write $r = f(s) + \eta$ where η is a random variable generated from a probability density chosen to match the variability in the data, for example a Gaussian. Negative rates that can arise from such a procedure can be eliminated by rectification, writing $r = [f(s) + \eta]_+$. If the standard deviation σ_η is independent of $f(s)$, the variability is called additive noise. The standard deviation can also depend on $f(s)$, and multiplicative noise corresponds to having σ_η proportional to $f(s)$.

Response variability extends beyond the level of firing rates to the entire temporal pattern of action potentials. Later in the chapter, we discuss another type of probabilistic model of the neuronal response that uses a stochastic spike generator to produce response variability. This approach, which we use for the remainder of this chapter, takes a deterministic estimate of the firing rate $r_{\rm est}(t)$ and produces a stochastic spiking pattern from it. The spike generator produces variable numbers and patterns of action potentials, even if the same estimated firing rate is used on each trial. For a fixed stimulus value s, we can use the tuning curve to provide an estimate of the firing rate, $r_{\rm est} = f(s)$. Even though this estimate is deterministic, the firing rate r computed from the spikes produced stochastically by the model will vary from trial to trial. We first discuss different ways of estimating firing rates and then introduce the stochastic spike generator.

1.2 What Makes a Neuron Fire?

Response tuning curves characterize the average response of a neuron to a given stimulus. It is also useful to consider the complementary procedure of averaging the stimuli that produce a given response. To average stimuli in this way, we need to specify what fixed response we will use to 'trigger' the average. The most obvious choice is the firing of an action potential. Thus, we ask, "what on average did the stimulus do before an action potential was fired?" The resulting quantity, called the spike-triggered average stimulus, provides a useful way of characterizing neuronal selectivity.

The Spike-Triggered Average

The spike-triggered average stimulus $C(\tau)$ is the average value of the stimulus a time interval τ before a spike is fired. The average is the sum over all the spikes in a trial occurring at times t_i for $i = 1, 2, \ldots, n$ and over all trials, divided by the total number of spikes in all the trials. For a large number of trials, the total number of spikes is approximately the average number of spikes per trial, $\langle n \rangle$, times the number of trials, and thus

$$C(\tau) = \frac{1}{\langle n \rangle} \left\langle \sum_{i=1}^{n} s(t_i - \tau) \right\rangle . \tag{1.21}$$

Figure 1.6 provides a schematic description of the computation of the spike-triggered average. Each time a spike appears, the stimulus in a time window preceding the spike is recorded. Although the range of τ values in equation 1.21 is unlimited, the response is typically affected only by the stimulus in a window a few hundred milliseconds wide immediately preceding a spike. Therefore, we expect $C(\tau)$ to be near zero for $\tau < 0$ (if the stimulus is uncorrelated in time) and for positive τ values larger than the correlation time between the stimulus and the response. In practice, the stimulus is only recorded over a time of order this correlation time (the shaded areas in figure 1.6). The recorded stimuli for all spikes are then summed and the procedure is repeated over multiple trials.

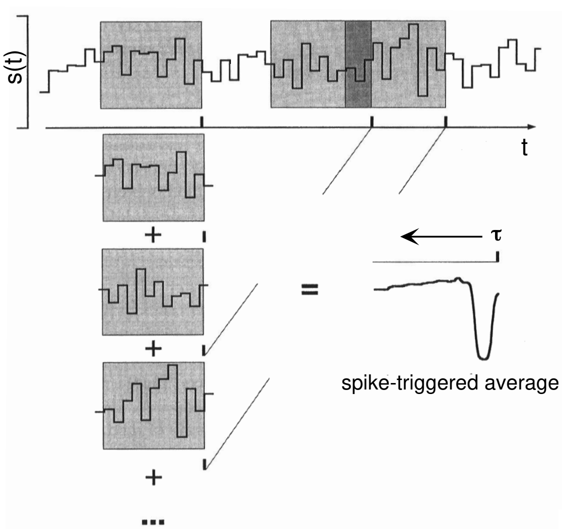

Figure 1.6: Schematic of the procedure for computing the spike-triggered average stimulus. Each grey rectangle contains the stimulus prior to one of the spikes shown along the time axis. These are averaged to produce the waveform shown at the lower right, which is the average stimulus before a spike. The stimulus in this example is a piecewise constant function of time. (Adapted from Rieke *et al* 1997)

The spike-triggered average is closely related to the correlation function of the trial-averaged neural response function and the stimulus. Correlation functions are a useful way of determining how two quantities, evaluated at different times, are related to each other. The correlation function depends on the time difference in the evaluation of the two quantities being considered. If one quantity is evaluated at time t and the other at time $t+\tau$ the correlation is obtained by averaging over all t values and is a function of τ. The correlation function of the trial-averaged neural response function and the stimulus is defined as

$$Q_{\rho s}(\tau) = \frac{1}{T}\int_0^T dt \, \langle \rho(t) \rangle \, s(t+\tau) \, . \tag{1.22}$$

If we define the average firing rate over the entire set of trials as $\bar{r} = \langle n \rangle / T$, and perform the integrals over the δ functions arising from the definition 1.4 of the neural response function, we find the relationship

$$\frac{Q_{\rho s}(-\tau)}{\bar{r}} = \frac{1}{\bar{r}T}\left\langle \sum_{i=1}^{n} s(t_i+\tau) \right\rangle = C(\tau) \tag{1.23}$$

by equation 1.21 (since $\bar{r}T = \langle n \rangle$). Because of this relation, the spike-triggered average stimulus is sometimes called the reverse correlation function. It is proportional to the correlation of the average neural response function with the stimulus at preceding times. As above, the trial-averaged neural response function can be replaced by the single-spike probability density, $\langle \rho(t) \rangle \rightarrow p[t]$ in the integral of equation 1.22 if large numbers of trials are being consider. In addition, we assume that the trial-averaged firing rate provides a good approximation of the single-spike probability density, $p[t] \approx \langle r(t) \rangle$. In this case, trial-averaged correlation function of the neural response function and the stimulus is approximately equal to the correlation function of the average firing rate and the stimulus,

$$Q_{\rho s}(\tau) \approx Q_{rs}(\tau) = \frac{1}{T}\int_0^T dt \, \langle r(t) \rangle s(t+\tau) \, , \tag{1.24}$$

so we can write

$$C(\tau) = \frac{Q_{\rho s}(-\tau)}{\bar{r}} \approx \frac{Q_{rs}(-\tau)}{\bar{r}} \tag{1.25}$$

$Q_{\rho s}$ and Q_{rs} play important roles in studies of neural encoding and decoding.

The spike-triggered average stimulus is widely used to study and characterize neural responses. Because $C(\tau)$ is the average value of the stimulus a time τ *before* a spike, larger values of τ represent times further in the past relative to the time of the triggering spike. For this reason, we will plot spike-triggered averages with the time axis going backward compared with the normal convention. Figure 1.7 shows the spike-triggered average stimulus for a neuron in the electrosensory lateral-line lobe of the weakly electric fish *Eigenmania*. Weakly electric fish generate oscillating

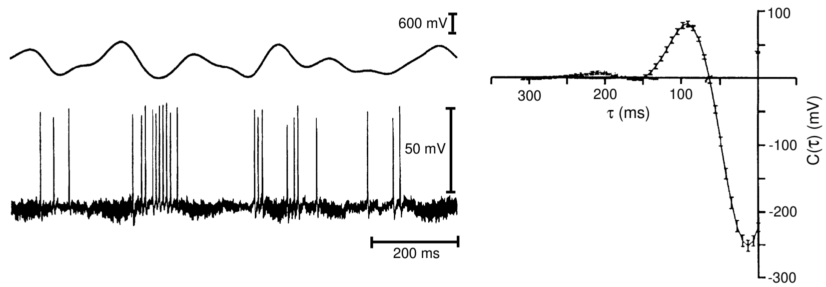

Figure 1.7: The spike-triggered average stimulus for a neuron of the electrosensory lateral-line lobe of the weakly electric fish *Eigenmania*. The upper left trace is the potential used to generate the electric field to which this neuron is sensitive. The evoked spike train is plotted below the stimulus potential. The plot at right is the spike-triggered average stimulus. (From Gabbiani *et al*, 1996)

electric fields from an internal electric organ. Distortions in this electric field produced by nearby objects are detected by sensors spread over the skin of the fish. The lateral-line lobe acts as a relay station along the processing pathway for electrosensory signals. In experiments by Gabbiani, Metzner, Wessel and Koch (1996), a fluctuating electric potential, shown as the upper left trace in figure 1.7, was used to elicit a response from an electrosensory lateral-line lobe neuron, as seen in the lower left trace. The spike-triggered average stimulus, plotted at the right, indicates that, on average, the electric potential made a positive upswing followed by a large negative deviation prior to a spike being fired by this neuron (because the spike-triggered average is plotted with a reversed time axis, the average stimulus can be read off in the usual left to right order). Since it is an average, the spike-triggered average has the same units as the stimulus, in this case mV.

Interpretation of spike-triggered averages is complicated by the fact that they depend on the particular set of stimuli that are averaged. How should this set be chosen? We will show in section 3 that, at least under an assumption of linearity, there are compelling reasons to use a stimulus that is uncorrelated from one time to the next, a white-noise stimulus. A heuristic argument supporting the use of such stimuli is that, in asking what makes a neuron fire, we want to minimize the bias introduced by our choice of stimuli. For example, we want to sample the response of a neuron to stimulus fluctuations at all frequencies with equal weight, at least up to some cutoff frequency beyond which the neuron cannot respond. White-noise stimuli have exactly this property (see Appendix A). Figure 1.7 was made using a white-noise stimulus with high frequencies filtered out.

White-Noise Stimuli

The defining characteristic of a white-noise stimulus is that its value at any one time is uncorrelated with its value at any other time. This condition can be expressed using the stimulus-stimulus correlation function, also called the stimulus autocorrelation function,

$$Q_{ss}(\tau) = \frac{1}{T}\int_0^T dt\, s(t)s(t+\tau)\,. \tag{1.26}$$

For white-noise, the stimulus-stimulus correlation function (in the limit $T \to \infty$) is zero except when $\tau = 0$,

$$Q_{ss}(\tau) = \sigma_s^2 \delta(\tau) \tag{1.27}$$

where σ_s^2 is a constant that sets the magnitude of the variability of the stimulus.

To construct white-noise stimuli, we divide time into bins of length Δt and consider times that are integer multiples of this basic unit, $t = m\Delta t$ for $m = 1, 2, \ldots, M$, where $M\Delta t = T$. The function $s(t)$ is then constructed as a discrete sequence of stimulus values. This produces a step-like stimulus waveform, like the one that appears in figure 1.6, with a constant stimulus value s_m presented during time bin m. In terms of the discrete-time values s_m, the condition that the stimulus is uncorrelated is

$$\frac{1}{M}\sum_{m=1}^{M} s_m s_{m+p} = \begin{cases} \sigma_s^2/\Delta t & \text{if} \quad p = 0 \\ 0 & \text{otherwise}\,. \end{cases} \tag{1.28}$$

An approximation of white noise can be generated by choosing each s_m independently from a probability density with mean zero and variance $\sigma_s^2/\Delta t$. Any reasonable probability density function satisfying these two conditions can be used to generate the stimulus values within each time bin. The factor of Δt in the variance indicates that the variability must be increased as the time binning gets finer. The binning of time into discrete intervals of size Δt means that the noise generated is only white (*i.e.* has a flat power spectrum) up to frequencies of order $1/\Delta t$. A special class of white-noise stimuli, Gaussian white-noise, results when the probability density used to generate the s_m values is a Gaussian function. There are a number of efficient schemes for rapidly generating approximately white-noise stimuli (see the references at the end of this chapter). In practice, no physical system can generate noise that is white to arbitrarily high frequencies, and likewise no neuron is responsive to arbitrarily high frequencies. White noise can be low-pass filtered as long as the cut-off frequency is well above the sensitivity of the neuron under investigation.

Multiple-Spike-Triggered Averages and Spike-Triggered Correlations

In addition to triggering on single spikes, average stimulus values can be computed by triggering on various combinations of spikes. Figure 1.8 shows some examples of two-spike triggers. These results come from a study by de Ruyter van Steveninck and Bialek (1988) of the H1 movement-sensitive visual neuron of the blowfly. The H1 neuron detects rotational motion during flight to generate and guide stabilizing motor corrections. It responds to angular rotations of the visual scene. In the experiments, the fly is held fixed while a visual image is rotated around it with a time-varying angular velocity $s(t)$. Figure 1.8A, showing the spike-triggered average stimulus, indicates that this neuron responds to positive angular

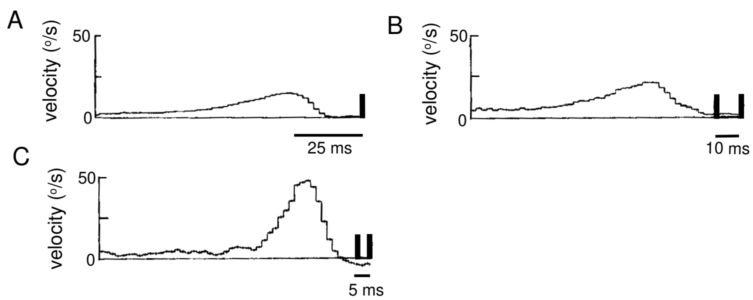

Figure 1.8: Single and multi-spike triggered average stimuli and a spike-triggered correlation function for a blowfly H1 neuron responding to a rotating visual image. A) The average stimulus velocity triggered on a single spike. B) The average stimulus velocity before two spikes with a separation of 10 ± 1 ms. C) The average stimulus before two spikes with a separation of 5 ± 1 ms. (Adapted from Rieke et al., 1997)

velocities after a latency of about 15 ms. Figure 1.8B is the average stimulus prior to the appearance of two spikes separated by 10 ms ($\pm$ 1 ms). The 10 ms two-spike average is equivalent to the sum of two single-spike-triggered average stimuli displaced from each other by 10 ms. Thus, for 10 ms separations, two spikes occurring together tell us no more as a two-spike unit than they would individually. This result changes when shorter separations are considered. Figure 1.8C shows the average stimulus triggered on two spikes separated by 5 ms ($\pm$ 1 ms). The average stimulus triggered on a pair of spikes separated by 5 ms is not the same as the sum of the average stimuli for each spike separately. Nevertheless, both encoding and decoding for the H1 neuron can be analyzed quite successfully by considering each spike independently, as discussed in Rieke *et al* (1997). This is because spikes separated by less than 5 ms do not occur frequently, and the additional information they carry considered as a pair is not very significant.

Spike-triggered averages of other stimulus-dependent quantities can provide additional insight into neural encoding, for example a spike-triggered covariance function. Obviously spike-triggered averages over higher-order stimulus combinations can be considered as well.

1.3 Modeling Firing Rates

Reverse correlation functions can be used as part of a general approach for modeling how one dynamic variable depends on another. We will use the method to construct an estimate $r_{\text{est}}(t)$ of a firing rate $r(t)$ evoked by a stimulus $s(t)$. The activity of a neuron at time t typically depends on what the stimulus did over a period of time starting a few hundred ms prior to t and ending perhaps tens of ms before t, because of integration and latency. The simplest way to construct an estimate is to assume that the firing rate at time t can be expressed as a weighted sum of the values taken by the stimulus at prior times. Since time is a continuous variable this 'sum' actually takes the form of an integral, and we write

$$r_{\text{est}}(t) = r_0 + \int_0^\infty d\tau\, D(\tau)s(t-\tau)\,. \tag{1.29}$$

The term r_0 accounts for any background firing that may occur when $s = 0$ (recall that this does not mean that the full stimulus S is zero, but rather that it takes its average value). $D(\tau)$ 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 1.83 is a linear filter just like the expression for the firing rate in equation 1.6.

We have provided a heuristic justification for the terms in equation 1.83 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

$$\begin{aligned} r_{\text{est}}(t) &= r_0 + \int d\tau\, D(\tau)s(t-\tau) + \int d\tau_1 d\tau_2\, D_2(\tau_1,\tau_2)s(t-\tau_1)s(t-\tau_2) + \\ &\quad \int d\tau_1 d\tau_2 d\tau_3\, D_3(\tau_1,\tau_2,\tau_3)s(t-\tau_1)s(t-\tau_2)s(t-\tau_3) + \ldots \end{aligned} \tag{1.30}$$

This series was rearranged by Wiener to make it easier to compute the coefficient functions (see Marmarelis and Marmarelis, 1978, for example). The first two terms of the Volterra and Wiener expansions are identical and the two expressions on the right side of equation 1.83 are precisely these two terms. For this reason, D is called the first Wiener kernel, the linear kernel, or, when higher-order terms are not being considered, simply the kernel.

To construct the best possible estimate of the firing rate given by an expression of the form 1.83, we choose the kernel D to minimize the squared difference between the estimated response to a stimulus and the actual measured response averaged over both time and trials,

$$E = \frac{1}{T}\int_0^T dt \left\langle (r_{\text{est}}(t) - r(t))^2 \right\rangle . \tag{1.31}$$

This expression is minimized in Appendix B by setting its derivative with respect to the function D to zero. The result is that D satisfies an equation involving the response-stimulus correlation function of equation 1.24 and the stimulus autocorrelation function of equation 1.26,

$$\int_0^\infty d\tau' \, Q_{ss}(\tau - \tau')D(\tau') = Q_{rs}(-\tau) . \tag{1.32}$$

The method we are describing is called reverse correlation because the response-stimulus correlation function is evaluated at $-\tau$.

Equation 1.32 can also be solved most easily if the stimulus is white noise, although it can be solved in the general case (see exercise 7). For a white-noise stimulus satisfying 1.27, the left side of equation 1.32 is

$$\sigma_s^2 \int_0^\infty d\tau' \, \delta(\tau - \tau')D(\tau') = \sigma_s^2 D(\tau) \tag{1.33}$$

and the kernel that provides the best linear estimate of the firing rate is

$$D(\tau) = \frac{Q_{rs}(-\tau)}{\sigma_s^2} . \tag{1.34}$$

In cases where the correlation functions of the firing rate and the neural response function with the stimulus are approximately equal, we use the relationship 1.25 to relate the optimal linear kernel to the spike-triggered average,

$$D(\tau) = \frac{Q_{rs}(-\tau)}{\sigma_s^2} \approx \frac{Q_{\rho s}(-\tau)}{\sigma_s^2} = \frac{\bar{r}C(\tau)}{\sigma_s^2} , \tag{1.35}$$

where $\bar{r}$ is the average firing rate of the neuron. This equation is the basis of the standard method used to determine the optimal kernel, computing the spike-triggered average stimulus in response to a white-noise stimulus.

Figure 1.9 shows a prediction of the firing rate of the H1 neuron introduced previously. The angular velocity of the rotated image is plotted in 1.9A and typical responses are shown in 1.9B. 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 1.9C. 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, presummably because of nonlinear effects not captured by the linear kernel. Some of these effects can be modeled by a static nonlinear function, as discussed below. Others may require including higher-order terms in the Volterra or Wiener expansion.

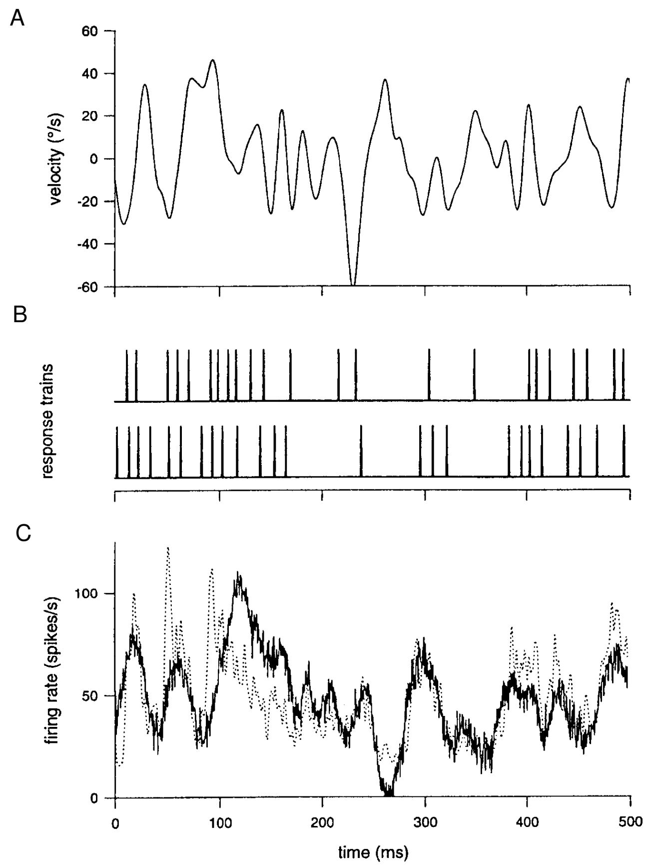

Figure 1.9: Prediction of the firing rate for an H1 neuron responding to a moving visual image. A) The angular velocity of the rotating visual image used to stimulate the neuron. B) Two of the 100 spike sequences used in this experiment to construct the spike-triggered average and to compute a time-dependent firing rate. C) The dashed line is the time-dependent firing rate computed directly from the spike trains. The solid line is an estimate of the firing rate constructed from linearly filtering the stimulus with an optimal kernel. (Adapted from Rieke *et al*, 1997)

The Optimal Stimulus

Equation 1.83 can be used to compute the stimulus predicted to evoke the maximum firing rate, subject to a constraint on stimulus power. A constraint is essential because the linear estimate 1.83 is an unbounded function of the stimulus. The constraint we use is that the integral of the square of the stimulus is held fixed,

$$\int_0^T dt\, s^2(t) = \text{constant}. \tag{1.36}$$

We call this integral the stimulus power. The stimulus that produces 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 at any given time, evaluated an interval τ prior to that time, is proportional to the optimal linear kernal $D(\tau)$, 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 1.83. At fixed stimulus power, the integral in 1.83 measures the overlap between the stimulus and the predicted optimal stimulus. In other words, it indicates how well the actual stimulus matches the predicted optimal 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 provides a linear estimate of the firing rate in response to a given stimulus. Neurons and nervous systems tend to be highly nonlinear, so a linear estimation procedure must be viewed as a crude first 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 predicted rate increases indefinitely as the stimulus strength is increased. The firing rate of a real neuron cannot be negative, and it will saturate at some point as the stimulus is increased. 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,

$$L(t) = \int_0^\infty d\tau D(\tau)s(t-\tau)\,. \tag{1.37}$$

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

$$r_{\text{est}}(t) = r_0 + F(L(t)) \tag{1.38}$$

where F is an arbitrary function. F is called a static nonlinearity to stress that it is a function of an integral of the stimulus evaluated at a single time, not a functional of the full stimulus waveform. If F is bounded from above and below, the estimated firing rate will never be negative or unrealistically large.

The function F can be extracted from data by means of a graphical procedure. First, a linear estimate of the firing rate is computed using the optimal kernel. 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.

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

$$F(L) = g[L - L_0]_+ \tag{1.39}$$

where L_0 is the threshold value that L must attain before firing will begin. Above the threshold, the firing rate is a linear function of L with g acting as the constant of proportionality. This function does not saturate, but this 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,

$$F(L) = \frac{r_{\max}}{1 + \exp\big(g(L_0 - L)\big)} \, . \tag{1.40}$$

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

$$F(L) = r_{\max}[\tanh\big(g(L - L_0)\big)]_+ \, . \tag{1.41}$$

Figure 1.10 shows the different nonlinear functions that we have discussed.

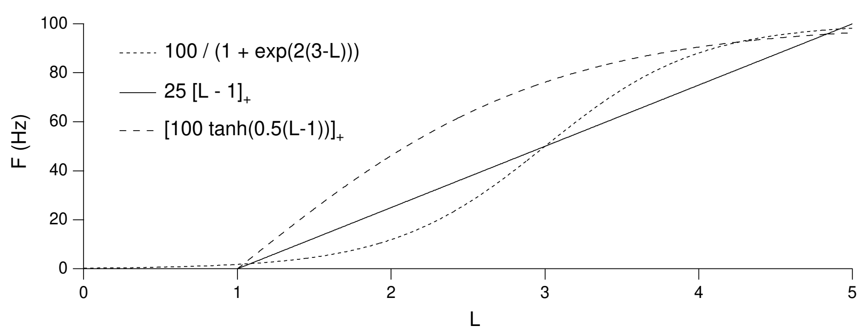

Figure 1.10: Different static nonlinearities used in modeling neural responses. The units of L are arbitrary. The functions and parameters used are given in the legend.

Although the static nonlinearity can be any function, this model of the firing rate is still restrictive because it allows for no dependence on weighted correlations of the stimulus such as the higher-order terms in the Volterra series. Furthermore, once the static nonlinearity is introduced, the linear kernel derived from equation 1.32 is no longer optimal. Nevertheless, there are interesting cases when 1.38 provides a reasonable description of neural responses. A theorem due to Bussgang (1952 & 1975) suggests that Gaussian white noise is the best stimulus to use to extract the linear kernel when static nonlinearities are present.

In some cases, the linear term of the Volterra series fails to predict the response and second-order terms provide a better approximation. For example, responses of auditory neurons that phase lock to stimulus oscillations, are best described by second-order kernels. 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 1.37 by an appropriately chosen function of s, so that

$$L(t) = \int_0^\infty d\tau D(\tau) f(s(t-\tau)) \,. \tag{1.42}$$

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 only that the integral of the kernel D is equal to one. For time-dependent stimuli, we can think of equation 1.42 as a dynamic extension of the response tuning curve. An example that uses the response tuning curve in this way is given later in the chapter.

1.4 Adding the Spikes

The models we have discussed generate a prediction of the firing rate $r(t)$ which is related to the probability of firing a single spike in a small interval around the time t. In general, this is not sufficient information to determine the probability that a particular spike sequence will occur in response to a given stimulus. Because spike times are continuous variables, the probability of different spike sequences is characterized by a probability density of spike times, $p[t_1, t_2, \ldots, t_n]$. The probability $P[t_1, t_2, \ldots, t_n]$ that a sequence of n spikes occurs with spike i falling between times t_i and $t_i + \Delta t$ for $i = 1, 2, \ldots, n$ is given in terms of this density by $p[t_1, t_2, \ldots, t_n](\Delta t)^n$. The firing rate $r(t)$ only provides enough information to compute the full spike-sequence probability density $p[t_1, t_2, \ldots, t_n]$ if individual action potentials are statistically independent, *ie* if the probability of generating any given action potential is independent of the presence or timing of any other spikes appearing in the sequence. Fortunately, as discussed below, there are a number of cases where independent spiking appears to be a reasonable approximation.

A stochastic process that generates a sequence of events, such as spikes, is called a point process. In general, the probability of an event occuring at any given time could depend on the entire history of preceding events. A point process for which this dependence extends only to the immediately preceding event, so that the intervals between successive events are independent, is called a renewal process. If there is no dependence at all on preceding events, so that the events themselves are statistically independent, we have a Poisson process. The Poisson model is simple, but it nevertheless provides a good description of neuronal firing in a fair number of cases. For these reasons we will discuss it in some detail. To make the presentation easier to follow we separate two cases, the homogeneous Poisson process, for which the firing rate is constant over time, and the inhomogeneous Poisson process, which involves a time-dependent firing rate.

Certain features of neuronal firing violate the independence assumption that forms the basis of the Poisson model, at least if a constant firing rate is used. Following the generation of an action potential, there is an interval of time known as the absolute refractory period during which the neuron cannot fire another spike. For a longer interval known as the relative refractory period, the likelihood of a spike being fired is reduced. Refractory effects can be incorporated into the Poisson spike generator by setting the firing rate to zero immediately after a spike is fired, and then letting it return to its predicted value according to some dynamic rule such as an exponential recovery (see exercise 1).

Some neurons fire action potentials in clusters or bursts of spikes that cannot be described by a Poisson process with a fixed rate. Bursting can be included in the Poisson model by allowing the firing rate to fluctuate to describe the high rate of firing during a burst. Sometimes the distribution of bursts themselves can be described by a Poisson process. In summary, a simple Poisson spike generator with a judiciously chosen time-dependent firing rate provides a reasonable description of spike generation in many cases. Note, however, that this is a *descriptive* model of neuronal spike trains; it does not explain or account for the source of neuronal variability.

The Homogeneous Poisson Process

When the firing rate is independent of time, the Poisson process generates every sequence of n spikes with equal probability, so the probability density $p[t_1, t_2, \ldots, t_n]$ depends only on n and not on the individual spike times. The dependence on n can be characterized by computing $P_T[n]$, the probability that n spikes occur within a trial of duration T, when the firing rate of the neuron is r. We do this by dividing the time T into M bins of size Δt. We assume that Δt is small enough so that we never get two spikes within any one bin (at the end of the calculation we will take the limit $\Delta t \to 0$). The n spikes can be placed in any n of these M bins and the number of ways that this can be done is given by the binomial coefficient $M!/(M-n)!n!$. This combinatorial factor must be multiplied by the probability of having n bins with single spikes in them and $M-n$ bins without any spikes in them. The probability of a spike appearing in a small bin of size Δt is $r\Delta t$ and the probability of having n such bins is $(r\Delta t)^n$. Similarly the probability of having the remaining $M-n$ bins without any spikes is $(1-r\Delta t)^{M-n}$. Putting all these factors together we find,

$$P_T[n] = \lim_{\Delta t \to 0} \frac{M!}{(M-n)!n!}(r\Delta t)^n (1-r\Delta t)^{M-n} . \tag{1.43}$$

To take the limit we note that as $\Delta t \to 0$, M grows without bound because $M\Delta t = T$. Since the number of spikes n remains finite, we can write $M - n \approx M = T/\Delta t$. Using this approximation and defining $\epsilon = -r\Delta t$ we find that

$$\lim_{\Delta t \to 0}(1-r\Delta t)^{M-n} = \lim_{\epsilon \to 0}\left((1+\epsilon)^{1/\epsilon}\right)^{-rT} = e^{-rT} \tag{1.44}$$

since the quantity in the outer parentheses is, by definition, e. For large M, $M!/(M-n)! \approx M^n = (T/\Delta t)^n$ so

$$P_T[n] = \frac{(rT)^n}{n!}\exp(-rT) . \tag{1.45}$$

This function is called the Poisson distribution.

The probabilities $P_T[n]$, for a few n values, are plotted as a function of the time period T (measured in units of $1/r$) in figure 1.11A. Note that as

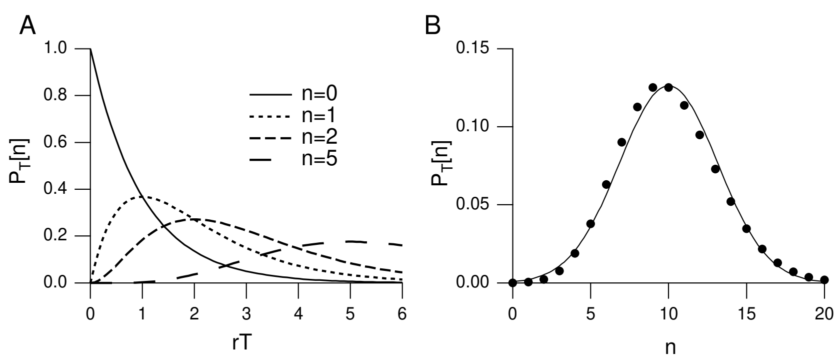

Figure 1.11: A) The probability that a homogeneous Poisson process generates n spikes in a time period of duration T plotted for $n = 0$, 1, 2, and 5. The probability is plotted as function of the rate times the duration of the interval, rT, to make the plot applicable for any rate. B) The probability of finding n spikes during a time period for which $rT = 10$ (dots) compared with a Gaussian distribution with mean and variance equal to 10 (line).

n increases, the maximum probability occurs for longer intervals and for long intervals large n values are more likely than small ones. Figure 1.11B shows the probabilities of various number of spikes occurring during trials for which the average number of spikes is 10. For large rT (a large number of expected spikes), the Poisson distribution approaches a Gaussian distribution with mean and variance equal to rT, and figure 1.11B shows that this approximation is already quite good for $rT = 10$.

We can compute the variance of spike counts produced by a Poisson process from the probabilities 1.45. For spikes counted over an interval of duration T, the variance of the spike count (derived in Appendix C) is

$$\sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2 = rT \,. \tag{1.46}$$

Thus the variance and mean of the spike count are equal. The ratio of these two quantities, $\sigma_n^2 / \langle n \rangle$, is called the Fano factor and takes the value one for a homogeneous Poisson process, independent of the time interval T.

The probability density of time intervals between adjacent spikes is called the interspike interval distribution, and it is a useful statistic for characterizing spiking patterns. Suppose that a spike occurs a time t_i for some i value. The probability of a homogeneous Poisson process generating the next spike somewhere in the interval $t_i + \tau \le t_{i+1} < t_i + \tau + \Delta\tau$, for small $\Delta\tau$, is the probability that no spike is fired for a time τ, which from 1.45 with $n = 0$ is $\exp(-r\tau)$, times the probability $r\Delta t$ of generating a spike within the following small interval $\Delta\tau$. Thus, the probability of an interspike interval falling between τ and $\tau + \Delta\tau$ is

$$P[\tau \le t_{i+1} - t_i < \tau + \Delta\tau] = r\Delta\tau \exp(-r\tau) \,. \tag{1.47}$$

The probability density of interspike intervals is, by definition, this probability with the factor $\Delta\tau$ removed. Thus, the interspike interval distribution for a homogeneous Poisson spike train is an exponential. The most likely interspike intervals are short ones, and long intervals have a probability that falls exponentially as a function of their duration.

From the interspike interval distribution of a homogeneous Poisson spike train, we can compute the mean interspike interval, averaged over trials,

$$\langle \tau \rangle = \int d\tau \, r\tau \exp(-r\tau) = \frac{1}{r} \,, \tag{1.48}$$

and the variance of the interspike intervals,

$$\sigma_\tau^2 = \int d\tau \, \tau^2 r \exp(-r\tau) - \langle \tau \rangle^2 = \frac{1}{r^2} \,. \tag{1.49}$$

The ratio of the standard deviation to the mean is called the coefficient of variation

$$C_V = \frac{\sigma_\tau}{\langle \tau \rangle} \tag{1.50}$$

and it takes the value one for a homogeneous Poisson process. This is a necessary, though not sufficient, condition to identify a Poisson spike train. Recall that the Fano factor for a Poisson process is also one. For any renewal process, $F(T) \to C_V^2$ for large values of T (Cox, 1962).

The Spike-Train Autocorrelation Function

The spike interval distribution measures the distribution of times between successive action potentials in a train. It is useful to generalize this concept and determine the distribution of times between any two spike in a train. This is called the spike-train autocorrelation function, and it is particularly useful for detecting patterns in spike trains, most notably oscillations. In practice, the spike-train autocorrelation function, like any probability density, is constructed from data by dividing time into bins and plotting a histogram. The value of the histogram for bin m (where m is a positive or negative integer) is computed by determining the number of the times that any two spikes in the train (not necessarily adjacent spikes) are separated by a time interval lying between $(m-1/2)\Delta\tau$ and $(m+1/2)\Delta\tau$ with $\Delta\tau$ the bin size. Call this number N_m. If the intervals between the n^2 spike pairs in the train were uniformly distributed over the range from zero to T into the different time bins, there would be $n^2\Delta\tau/T$ intervals in each bin. It is customary to remove this uniform term from the autocorrelation histogram by subtracting $n^2\Delta\tau/T$ from N_m for all m. By convention, the spike-train autocorrelation histogram is defined by dividing the resulting numbers by T, the duration of the spike sequence, so the value of the histogram in bin m is $H_m = N_m/T - n^2\Delta\tau/T^2$. This histogram has the units of a firing rate. The zero time intervals obtained by measuring the separation between each spike and itself are included in the autocorrelation function, but they are sometimes omitted from plotted histograms.

As we have seen, the distribution of interspike intervals for adjacent spikes in a homogeneous Poisson spike train is exponential (equation 1.47). In contrast, the intervals between *any* two spikes (not necessarily adjacent) in such a train are uniformly distributed. As a result, the subtraction procedure outlined above gives $H_m=0$ for all bins except for the $m=0$ bin that contains the contribution of the zero intervals between spikes and themselves. There are, on average, rT spikes in the train, so this bin gets an additional uncancelled contribution of r when we divide by T. The spike-train autocorrelation function $Q_{\rho\rho}(\tau)$ is the density or distribution related to the autocorrelation histogram. It is obtained by taking the $\Delta\tau \to 0$ limit of the autocorrelation histogram divided by $\Delta\tau$. The contribution $r/\Delta\tau$ right at the point $\tau = 0$ becomes $r\delta(\tau)$ in this limit so for a homogeneous Poisson spike train

$$Q_{\rho\rho}(\tau) = r\delta(\tau)\,. \tag{1.51}$$

For any spike train, the autocorrelation function is defined as the autocorrelation of the neural response function of equation 1.4 with the average value removed. Using the notation $\bar{r} = \langle n\rangle/T$, the spike-train autocorrelation function is

$$Q_{\rho\rho}(\tau) = \frac{1}{T}\int_0^T dt\ \langle(\rho(t)-\bar{r})\,(\rho(t+\tau)-\bar{r})\rangle\ , \tag{1.52}$$

and it is an even function of τ, $Q_{\rho\rho}(\tau) = Q_{\rho\rho}(-\tau)$. Because the average is subtracted from the neural response function in this expression, $Q_{\rho\rho}$ should really be called an autocovariance, not an autocorrelation, but in practice it isn't. Sometimes, the δ function term representing the correlation of spikes with themselves is removed by writing

$$Q_{\rho\rho}(\tau) = \bar{r}\delta(\tau) + Q^{+}_{\rho\rho}(\tau)\,. \tag{1.53}$$

A cross-correlation function between spike trains from two different neurons can be defined by analogy with the autocorrelation function by determining the distribution of intervals between pairs of spikes, one taken from each train. The cross-correlation function is not necessarily even. A peak at zero interval in a cross-correlation function signifies that the two neurons are firing synchronously. Asymmetric shifts in this peak away from zero result from fixed delays between the firing of the two neurons indicating non-synchronous but phase-locked firing. Periodic structure in either an autocorrelation or cross-correlation function or histogram indicates that the firing probability oscillates. Such periodic structure is seen in the histograms of figure 1.12 showing 40 Hz oscillations in neurons of cat primary visual cortex that are synchronized between the two cerebral hemispheres. Cross-correlation functions are often normalized by dividing the quantity described above by the square root of the product of the autocorrelation functions of the two spike trains being considered.

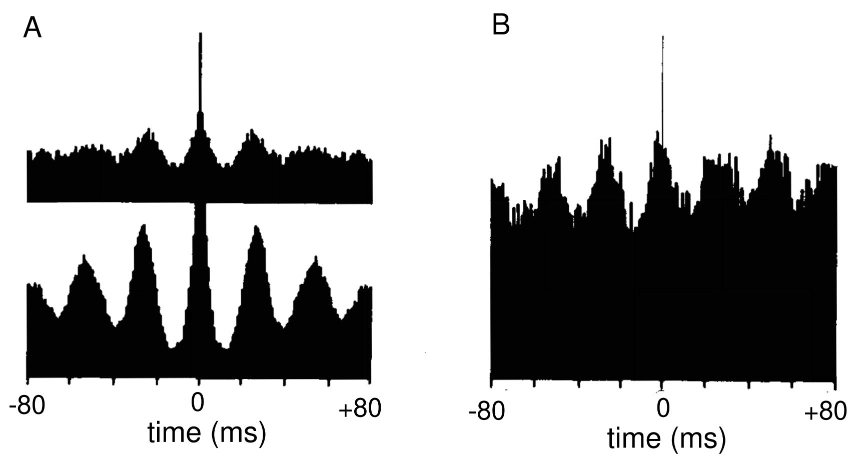

Figure 1.12: Autocorrelation and cross-correlation histograms for neurons in the primary visual cortex of a cat. A) Autocorrelation histograms for neurons recorded in the right (upper) and left (lower) hemispheres show a periodic pattern indicating oscillations at about 40 Hz. The lower diagram indicates stronger oscillations in the left hemisphere. B) The cross-correlation histogram for these two neurons shows that their oscillation are synchronized with virtually no phase shift. (From Engel *et al*, 1991)

The Inhomogeneous Poisson Process

When the firing rate depends on time, different sequences of n spikes occur with different probabilities, and $p[t_1, t_2, \ldots, t_n]$ depends on the spike times, although only through the time-dependent firing rate $r(t)$. Assuming that the spike times are ordered so that $0 \le t_1 \le t_2 \le \ldots \le t_n \le T$, the probability density for n spike times (derived in Appendix D) is

$$p[t_1, t_2, \ldots, t_n] = \exp\left(-\int_0^T dt\, r(t)\right) \prod_{i=1}^{n} r(t_i)\,. \qquad (1.54)$$

Note that for a homogeneous Poisson process with $r(t) = r$, this reduces to

$$p[t_1, t_2, \ldots, t_n] = \exp(-rT) r^n \qquad (1.55)$$

which is independent of the spike times as mentioned previously. The results in equations 1.54 and 1.55 apply if the spike times have been written in temporal order as discussed above. If the spike times are not ordered, so that, for example, we are interested in the probabilty of *any* spike occuring near the time t_1, not necessarily the first spike, these expression should be divided by a factor of $n!$ to account for the number of different possible orderings of spike times.

Comparison with Data

The Poisson process is analytically tractable, and it can be used to construct a useful spike generator, but does it match data on neural response variability? To address this question we examine Fano factors, interspike interval distributions, and coefficients of variation.

The Fano factor describes the relationship between the mean spike count over a given interval and the spike-count variance. Mean spike counts $\langle n \rangle$ and variances σ_n^2 extracted from neuronal recordings have been fit to the equation $\sigma_n^2 = A\langle n \rangle^B$ with the multiplier A and exponent B constants extracted from the data. The values of both A and B typically lie between 1.0 and 1.5. Since the Poisson model predicts $A = B = 1$, this indicates that the data show a higher degree of variability than the Poisson model would predict. However, many of these experiments involve anesthetized animals, and O'Keefe, Bair and Movshon (1997) have found that response variability is higher in anesthetized than in alert animals. Figure 1.13 shows their data for spike-count means and variances extracted from recordings of MT neurons in alert macaque monkeys using a number of different moving stimuli. The individual means and variances are scattered in figure 1.13A, but they cluster around the diagonal which is the Poisson prediction. Similarly, the results show A and B values close to one, the Poisson values (figure 1.13B).

Interspike interval distributions are extracted from data as interspike interval histograms by counting the number of intervals falling in discrete

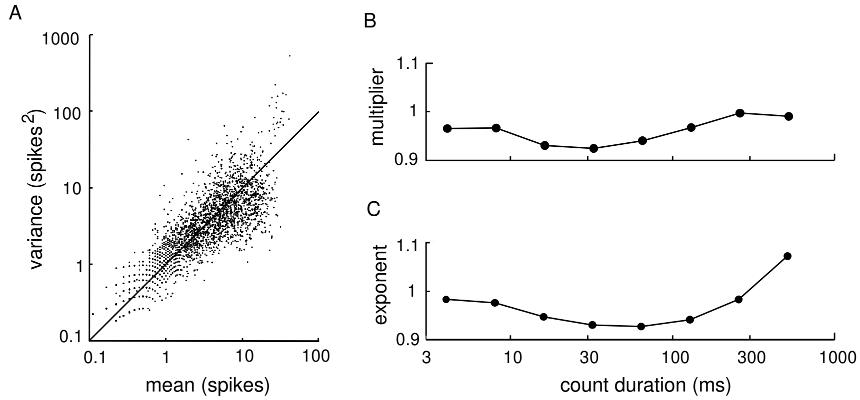

Figure 1.13: Variability of MT neurons in alert macaque monkeys responding to moving visual images. A) Variance of the spike counts for a 256 ms counting period plotted against the mean spike count. The straight line is the prediction of the Poisson model. Data are from 94 cells recorded under a variety of stimulus conditions. B) The multiplier A in the relationship between spike-count variance and mean as a function of the duration of the counting interval. C) The exponent B as a function of the duration of the counting interval. (From O'Keefe *et al*, 1997)

time bins. Bair, Koch, Newsome and Britten (1994) have extracted interspike interval histograms from responses of monkey MT neurons to moving dot images. A typical result for neurons that do fire bursts of action potentials is shown in figure 1.14A. For interspike intervals longer than

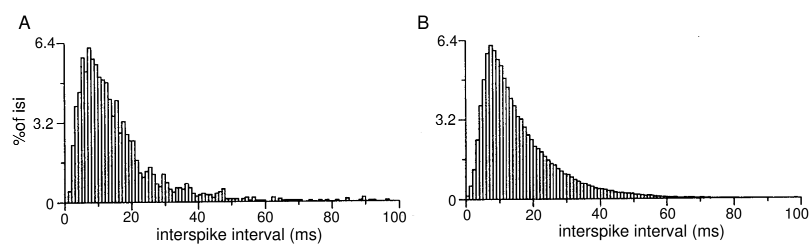

Figure 1.14: (A) Interspike interval distribution from an MT neuron responding to a moving random dot image. The probability of interspike intervals falling into the different bins, expressed as a percentage, is plotted against interspike interval. B) Interspike interval histogram generated from a Poisson model with a stochastic refractory period. (From Bair *et al*, 1994)

about 10 ms, the shape of this histogram is exponential in agreement with equation 1.47. However, for shorter intervals there is a discrepancy. While the homogeneous Poisson distribution 1.47 rises for short interspike intervals, the experimental results show a rapid decrease. This is the result of refractoriness making short interspike intervals less likely than the Poisson model would predict. Data on interspike intervals can be fit more accurately by a gamma distribution of the form

$$p[\tau] = \frac{r(r\tau)^q \exp(-r\tau)}{q!} \tag{1.56}$$

with $q > 0$ than by the exponential distribution of the Poisson model which has $q = 0$.

Figure 1.14B shows a theoretical histogram obtained by adding a refractory period to the Poisson model. After each spike, Bair *et al* (1994) inserted a refractory period of variable duration. Spiking was prohibited during the refractory period, and then was described once again by a homogeneous Poisson process. The refractory period was randomly chosen from a Gaussian distribution with a mean of 5 ms and a standard deviation of 2 ms (only random draws that generated positive refractory periods were included). The resulting interspike interval distribution of figure 1.14B agrees quite well with the data.

C_V values extracted from the spike trains of monkey V1 and MT neurons are shown in figure 1.15. The data have been divided into groups based on the mean interspike interval, and the coefficient of variation is plotted

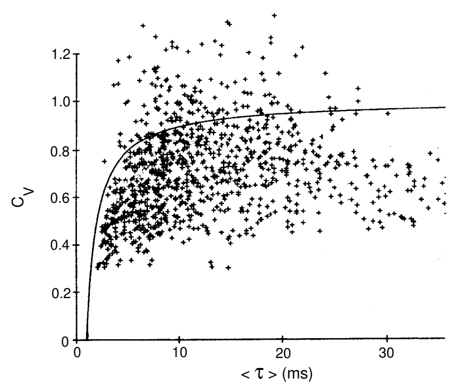

Figure 1.15: Coefficients of variation for a large number of V1 and MT neurons plotted as a function of mean interspike interval. The solid curve is the result of a Poisson model with a refractory period. (From Softky and Koch, 1992)

here as a function of this mean interval, equivalent to $1/r$. Except for short mean interspike intervals, the values are near one, although they tend to cluster slightly lower than one, the Poisson value. The small C_V values for short interspike intervals are due to the refractory period. The solid curve is the prediction of a Poisson model with refractoriness.

The Poisson model with refractoriness provides a remarkably good description of a large amount of data, especially considering its simplicity. However, the Poisson model, even when it successfully describes data, does not provide a mechanistic explanation of neuronal response variability. Spike generation, by itself, is highly reliable in real neurons. Figure

1.16 compares the response of V1 cells to constant current injection *in vivo* and *in vitro*. The *in vitro* response is a regular and reproducible spike train (left panel). The same current injection paradigm applied *in vivo* produces

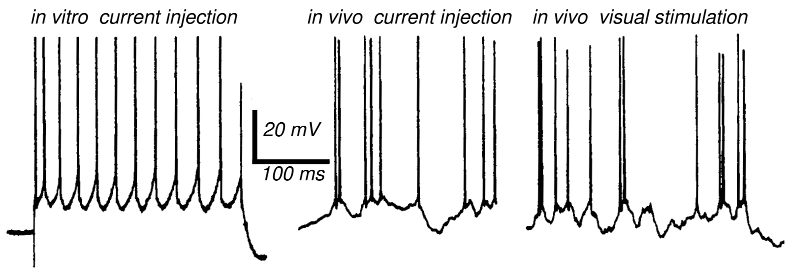

Figure 1.16: Intracellular recordings from cat V1 neurons. The left panel is the response of a neuron in an *in vitro* slice preparation to constant current injection. The center and right panels show recordings from neurons *in vivo* responding to either injected current (center), or a moving visual image (right). (From Holt *et al*, 1996)

a highly irregular pattern of firing (center panel) similar to the response to a moving bar stimulus (right panel). Although some of the basic statistical properties of firing variability may be captured by the Poisson model of spike generation, the spike generating mechanism itself in real neurons is clearly not responsible for the variability. We will explore ideas about possible mechanisms in chapter 6.

A Spiking Model

To build a descriptive model of the spike sequences evoked by different stimuli, we use the estimated firing rate $r_{\text{est}}(t)$ predicted from knowledge of the stimulus to drive a Poisson spike generator. A simple procedure for generating spikes in a computer program is based on the fact that the estimated probability of firing a spike during a short interval of duration Δt is $r_{\text{est}}(t)\Delta t$. The program progresses through time in small steps of size Δt and generates, at each time step, a random number x_{rand} chosen uniformly in the range between zero and one. If $r_{\text{est}}(t)\Delta t > x_{\text{rand}}$ at that time step, a spike is fired, otherwise it is not. For a given value of Δt, a more accurate and efficient method for generating a spike train, called the rejection method or spike thinning, is based on the exponential probability distribution of intervals between spikes. This method requires a bound $r_{\max}$ on the estimated firing rate such that $r_{\text{est}}(t) \leq r_{\max}$ at all times. We first generate a spike sequence corresponding to the constant rate $r_{\max}$ writing $t_{i+1} = t_i - \ln(x_{\text{rand}})/r_{\max}$ with $i = 1, 2, \ldots, n$ representing the different spikes. The spikes are then thinned by generating another x_{rand} for each

i and removing the spike at time t_i from the train if $r_{\text{est}}(t_i)/r_{\text{max}} < x_{\text{rand}}$. If $r_{\text{est}}(t_i)/r_{\text{max}} \geq x_{\text{rand}}$, spike i is retained. This thinning corrects for the difference between the estimated time-dependent rate and the maximum rate.

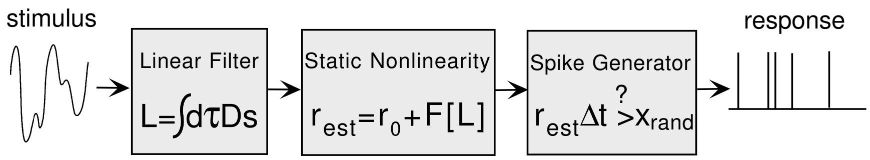

Figure 1.17: A model for predicting spiking responses to stimuli based on a linear filter. The integral of the stimulus s times the optimal kernel D is first computed. The estimated firing rate is the background rate r_0 plus a nonlinear function of the output of the linear filter calculation. Finally, the estimated time-dependent firing rate is used to drive a Poisson process that generates spikes.

Figure 1.17 shows the structure of a model consisting of 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 spike thinning method could be used instead. Also, the linear filter acts directly on the stimulus s in figure 1.17, but it could act instead on some function $f(s)$ such as the response tuning curve. Figure 1.18 shows an example of a model of an orientation selective V1 neuron constructed in this way. In this model, the estimated firing rate is determined by an exponential kernel acting on the response tuning curve of figure 1.3A,

$$r_{\text{est}} = \left[\frac{1}{\tau_r} \int_0^\infty d\tau \, \exp(-\tau/\tau_r) f(s(t-\tau)) \right]_+ \tag{1.57}$$

with $\tau_r = 30$ ms. In figure 1.18, the orientation angle increases in a sequence of steps. The firing rate responds to these changes by exponentially approaching the steady state value given by the tuning curve. The Poisson process generates an irregular firing pattern that reflects the underlying rate but varies from trial to trial.

1.5 Rate versus Temporal Coding

The nature of the neural code is a topic of intense debate within the neuroscience community. Much of the discussion has focused on whether neurons use rate coding or temporal coding, often without a clear definition of these terms. We define the notion of an independent-spike code to capture the essence of a rate code and discuss different definitions of temporal coding.

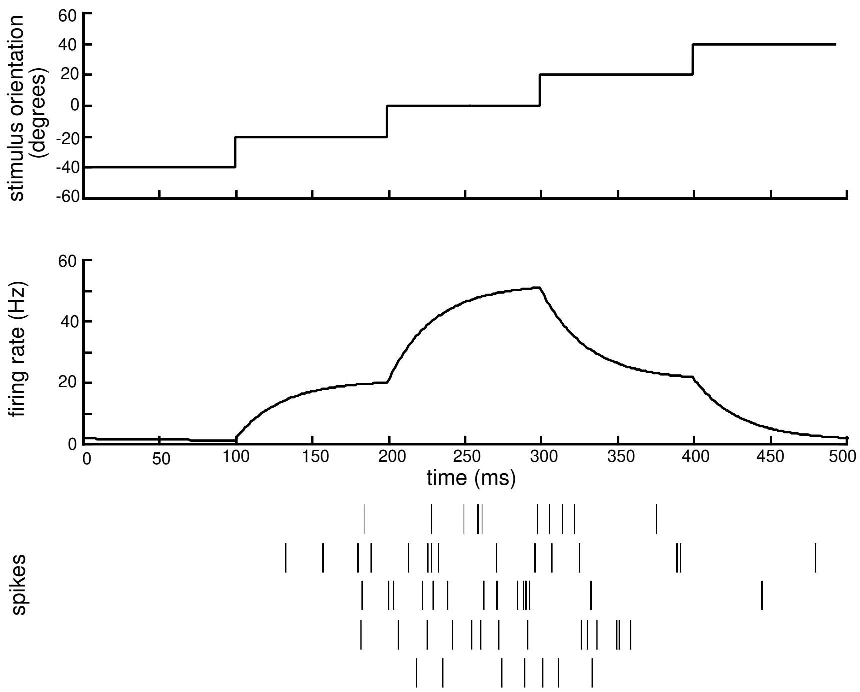

Figure 1.18: Model of an orientation selective neuron. The orientation angle (top panel) was increased from an initial value of -40° by 10° every 100 ms. The firing rate (middle panel) was used to generate spikes (bottom panel) using a Poisson spike generator. The bottom panel shows spike sequences generated on five different trials.

Independent-Spike Codes

The neural response, and its relation to the stimulus, is completely characterized by the probability distribution of spike times as a function of the stimulus. If spike generation can be described as an inhomogeneous Poisson process, this probability distribution can be computed from the time-dependent firing rate $r(t)$ using equation 1.54. In this case, $r(t)$ contains all the information about the stimulus that can be extracted from the spike train, and the neural code could reasonably be called a rate code. Unfortunately, this definition does not agree with common usage. Instead, we will call a code based solely on the time-dependent firing rate an *independent-spike code*, referring to the fact that the generation of each spike is independent of all the other spikes in the train. If spikes are not independent in this way, we call the code a *correlation code,* since correlations between spike times may carry additional information. Almost all the neural codes that have been studied have been analyzed as independent-spike codes.

When we study the response of more than one neuron, we must consider the independence of individual neurons in coding, as well as the independence of individual spikes. The analysis of population responses is easiest if the response of each neuron is considered statistically independent. The hypothesis that the activity of neurons is mutually independent is widely used in the analysis of population coding (chapter 3), even though it is less likely to be valid than the independent-spike hypothesis. The independent-neuron hypothesis does not mean that the spike trains of different neurons are not combined into an ensemble code. Rather it means that they can be combined without taking into account correlations. To test the validity of this assumption, we must ask whether firing correlations between neurons provide more information about a stimulus than their individual firing patterns alone. Two types of correlations are relevant; correlations between firing times and correlations between firing rates of different neurons.

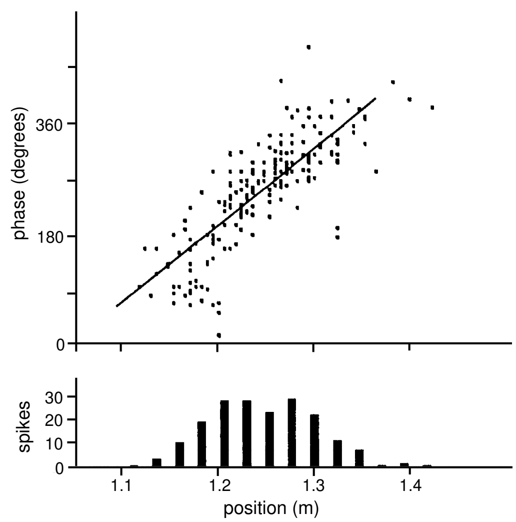

Figure 1.19: Position versus phase for a hippocampal place cell. Each dot in the upper figure shows the phase of the theta rhythm plotted against the position of the animal at the time when a spike was fired. The linear relation shows that information about position is contained in the relative phase of firing. The lower plot is a conventional place field tuning curve of spike count versus position. (Adapted from O'Keefe and Recce, 1993)

Place cell coding of spatial location in the rat hippocampus is an example where additional information appears to be carried by correlations between the firing patterns of neurons in a population. The hippocampus is a primitive cortical region of the brain (lacking the multilayered structure of neocortex) deep inside the temporal lobe. It plays an important role in memory formation and is involved in a variety of spatial tasks. The firing rates of many hippocampal neurons, recorded when a rat is moving around a familiar environment, depend on the position of the animal and are restricted to a spatially localized area called the place field of the cell. In addition, when a rat explores an environment, hippocampal neurons fire collectively in a rhythmic pattern with a frequency in the theta range, 7-12 Hz. O'Keefe and Recce (1993) discovered that relating the spiking time of an individual place cell to the phase of the population theta rhythm gave additional information about location of the rat not provided by place cells considered individually. The relationship between location and phase of place cell firing shown in figure 1.19 means, for example, that we can distinguish two locations on opposite sides of the peak of a single neuron's tuning curve that correspond to the same firing rate, by knowing when the spikes occurred relative to the theta rhythm.

In general, the independent-spike and independent-neuron hypotheses for neural coding are extremely powerful and widely used. Examples where they appear to break down are of great interest but are outnumbered by cases in which their application has been highly successful.

Temporal Codes

If the independent-spike hypothesis is valid and $r(t)$ governs the relationship between stimulus and response, the nature of the neural code depends on the behavior of $r(t)$. If $r(t)$ varies slowly with time, the code is called a rate code, and if it varies rapidly, the code is called temporal. Figure 1.20 provides an example of different firing-rate behaviors for a neuron in area MT of a monkey recorded over multiple trials with three different stimuli (consisting of moving random dots). The activity on the top panel would typically be regarded as reflecting rate coding, and the activity on the bottom as reflecting temporal coding. However, the identification of rate and temporal coding in this way is ambiguous because it is not obvious what criterion should be used to characterize the changes in $r(t)$ as slow or rapid.

One possibility is to use the spikes to distinguish slow from rapid, so that a temporal code is identified when peaks in the time-dependent firing rate occur with roughly the same frequency as the spikes themselves (see for example Rieke *et al*, 1997). In this case, each peak corresponds to the firing of only one, or at most a few action potentials. While this definition makes intuitive sense, it is problematic to extend it to the case of population coding. When many neurons are involved, any single neuron may fire only a few spikes before its time-dependent firing rate changes, but collectively the population may produce a large number of spikes over the same time period. Thus, a neuron that appears to employ a temporal code, by this definition, may be part of a population that does not.

Theunissen and Miller (1995) proposed using the stimulus, not the response, to establish what makes a temporal code. They define a temporal code as one in which information is carried by details of spike timing on a scale shorter than the fastest time characterizing variations of the stimulus.

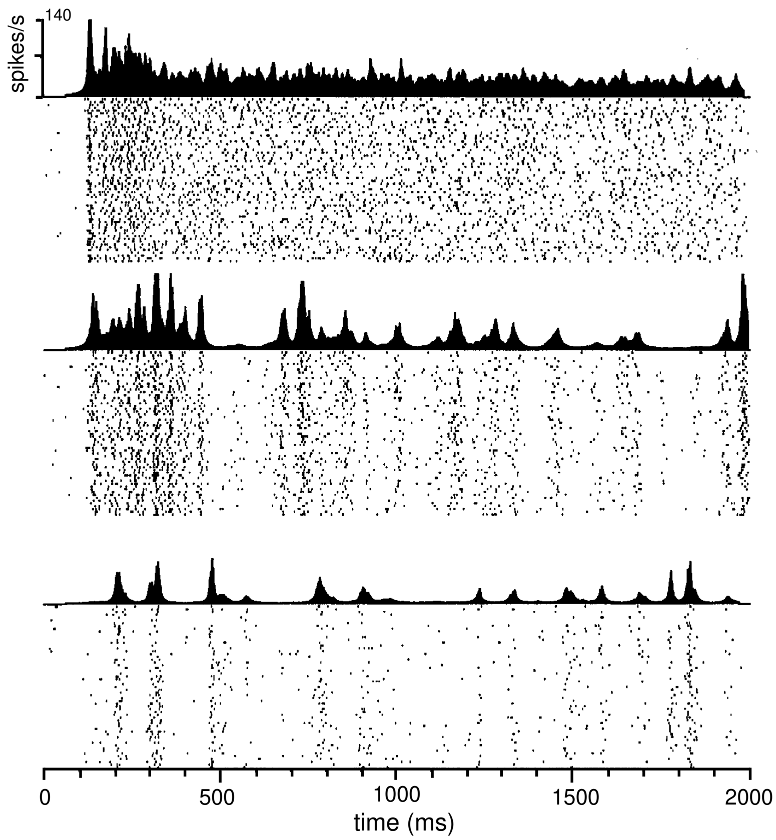

Figure 1.20: Time-dependent firing rates for different stimulus parameters. The rasters show multiple recordings of an MT neuron responding to the same moving random dot stimulus. Firing rates were constructed from the multiple trials by counting spikes and averaging over the trials shown immediately below the firing rate histograms. The three different results are from the same neuron but using different stimuli. The stimuli were always patterns of moving random dots but the coherence of the motion was varied (see chapter 3 for more information about this stimulus). (From Bair and Koch, 1996)

More precisely, this requires that information about the stimulus be carried by Fourier components of $r(t)$ at frequencies higher than those present in the stimulus. Many of the cases where a temporal code has been reported using spikes to define the nature of the code would be called rate codes if the stimulus were used instead.

1.6 Appendices

A) White-Noise Power Spectrum

The Fourier transform of the stimulus autocorrelation function,

$$\tilde{Q}_{ss}(\omega) = \int d\tau\, Q_{ss}(\tau) e^{i\omega\tau}, \tag{1.58}$$

is called its power spectrum. We can compute the power spectrum for a white-noise stimulus using the fact the $Q_{ss}(\tau) = \sigma_s^2 \delta(\tau)$ for white noise,

$$\tilde{Q}_{ss}(\omega) = \sigma_s^2 \int dt\, \delta(\tau) e^{i\omega\tau} = \sigma_s^2 \,. \tag{1.59}$$

This is the eponymous characteristic of white noise; its power spectrum is the same for all frequencies.

Using the definition of the stimulus autocorrelation function, we can also write

$$\tilde{Q}_{ss}(\omega) = \frac{1}{T} \int d\tau dt\, s(t) s(t+\tau) e^{i\omega\tau} \,. \tag{1.60}$$

Note that this expression is the Fourier transform of a convolution of the stimulus with itself (see chapter 12). As a result, the Fourier transform of the stimulus autocorrelation function can be re-expressed as

$$\tilde{Q}_{ss}(\omega) = \frac{1}{T} |\tilde{s}(\omega)|^2 \tag{1.61}$$

where $s(\omega)$ is the Fourier transform of $s(t)$. This provides another definition of the power spectrum of a stimulus, it is proportional to the absolute square of the Fourier transform of the stimulus.

B) The Optimal Linear Kernel and Optimal Stimulus

Using equation 1.83 for the estimated firing rate, the expression 1.31 to be minimized is

$$E = \frac{1}{T} \int_0^T dt \left\langle \left(r_0 + \int_0^\infty d\tau\, D(\tau) s(t-\tau) - r(t) \right)^2 \right\rangle . \tag{1.62}$$

The minimum is obtained by setting the derivative of E with respect to the function D to zero. This is called a *functional* derivative. To compute it, we introduce a small time interval Δt and evaluate all functions at integer multiples of Δt. We define $r_i = r(i\Delta t)$, $D_k = D(k\Delta t)$, and $s_{i-k} =$

$s((i-k)\Delta t)$. If Δt is small enough, the integrals in equation 1.62 can be approximated by sums, and we can write

$$E = \frac{\Delta t}{T} \sum_i \left\langle \left(r_0 + \Delta t \sum_k D_k s_{i-k} - r_i \right)^2 \right\rangle . \tag{1.63}$$

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

$$\frac{\partial E}{\partial D_j} = 0 = \frac{2\Delta t}{T} \sum_i \left(r_0 + \Delta t \sum_k D_k s_{i-k} - \langle r_i \rangle \right) s_{i-j} \Delta t . \tag{1.64}$$

Rearranging and simplifying this expression gives the condition

$$\Delta t \sum_k D_k \left(\frac{\Delta t}{T} \sum_i s_{i-k} s_{i-j} \right) = \frac{\Delta t}{T} \sum_i \left(\langle r_i \rangle - r_0 \right) s_{i-j} . \tag{1.65}$$

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 1.65 turn back into integrals, the indexed variables become functions, and we find

$$\int_0^\infty d\tau' \, D(\tau') \left(\frac{1}{T} \int_0^T dt \, s(t-\tau') s(t-\tau) \right) = \frac{1}{T} \int_0^T dt \, \left(\langle r(t) \rangle - r_0 \right) s(t-\tau) . \tag{1.66}$$

The term proportional to r_0 on the right side of equation 1.66 can be dropped because the time-average of s is zero. The remaining term is the response-stimulus correlation function 1.24 evaluated at $-\tau$, $Q_{rs}(-\tau)$. The term in large parentheses on the left side of 1.66 is the stimulus autocorrelation function, 1.26. By shifting the integration variable $t \to t + \tau$, we find that it is $Q_{ss}(\tau - \tau')$, so 1.66 can be re-expressed in the form of equation 1.32.

To find the stimulus that produces the maximum predicted responses at time t_0 subject to the fixed power constraint of equation 1.36, we use the method of Lagrange multipliers. We find the maximum value with respect to s of

$$r_{\text{est}}(t_0) + \lambda \int_0^T dt \, s^2(t) = r_0 + \int_0^\infty d\tau \, D(\tau) s(t_0 - \tau) + \lambda \int_0^T dt \, s^2(t) \tag{1.67}$$

where λ is the Lagrange multiplier to be determined. Setting the derivative of this expression with respect to the function s to zero (using the same methods used above; see exercise 10) gives

$$D(\tau) = -2\lambda s(t_0 - \tau) . \tag{1.68}$$

The value of λ (which is less than zero) is determined by requiring that condition 1.36 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) Moments of the Poisson Distribution

The average number of spikes generated by a Poisson process with rate r over a time T is

$$\langle n \rangle = \sum_{n=0}^{\infty} n P_T[n] = \sum_{n=0}^{\infty} \frac{n(rT)^n}{n!} \exp(-rT)\,, \tag{1.69}$$

and the variance in the spike count is

$$\sigma_n^2(T) = \sum_{n=1}^{\infty} n^2 P_T[n] - \langle n \rangle^2 = \sum_{n=1}^{\infty} \frac{n^2(rT)^n}{n!} \exp(-rt) - \langle n \rangle^2\,. \tag{1.70}$$

To compute these quantities we need to calculate the two sums appearing in these equations. A good way to do this is to compute the generating function

$$g(\alpha) = \sum_{n=0}^{\infty} \frac{(rT)^n \exp(\alpha n)}{n!} \exp(-rt)\,. \tag{1.71}$$

The p^{th} derivative of g with respect to α, evaluated at the point $\alpha = 0$, is

$$\left.\frac{d^p g}{d\alpha^p}\right|_{\alpha=0} = \sum_{n=0}^{\infty} \frac{n^p[rT]^n}{n!} \exp(-rt)\,, \tag{1.72}$$

so once we have computed g we only need to calculate its first and second derivatives to determine the sums we need. Rearranging the terms a bit, and recalling that $\exp(z) = \sum z^n/n!$, we find

$$g(\alpha) = \exp(-rT) \sum_{n=0}^{\infty} \frac{\big(rT\exp(\alpha)\big)^n}{n!} = \exp(-rT)\exp\left(rTe^{\alpha}\right)\,. \tag{1.73}$$

The derivatives are then

$$\frac{dg}{d\alpha} = rTe^{\alpha}\exp(-rT)\exp(rTe^{\alpha}) \tag{1.74}$$

and

$$\frac{d^2 g}{d\alpha^2} = (rTe^{\alpha})^2 \exp(-rT)\exp(rTe^{\alpha}) + rTe^{\alpha}\exp(-rT)\exp(rTe^{\alpha})\,. \tag{1.75}$$

Evaluating these at $\alpha = 0$ gives the results $\langle n \rangle = rT$ and $\sigma_n^2(T) = (rT)^2 + rT - (rT)^2 = rT$.

D) Inhomogeneous Poisson Statistics

To calculate the probability denstiy for a particular spike sequence t_i with $i = 1, 2, \ldots, n$ we need to determine the probability that no spike will be generated during the time interval from t_i to t_{i+1} between adjacent spikes. We compute this by dividing this interval into M bins of size Δt assuming that $M\Delta t = t_{i+1} - t_i$. We will ultimately take the limit $\Delta t \to 0$. The firing rate during bin m within this interval is $r(t_i + m\Delta t)$. Since the probability of firing a spike in this bin is $r(t_i + m\Delta t)\Delta t$, the probability of not firing a spike is $1 - r(t_i + m\Delta t)\Delta t$. To have no spikes during the entire interval, we must string together M such bins, and the probability of this occurring is the product of the individual probabilities,

$$P[\text{no spikes}] = \prod_{m=1}^{M} (1 - r(t_i + m\Delta t)\Delta t) \, . \tag{1.76}$$

We evaluate this expression by taking its logarithm,

$$\ln P[\text{no spikes}] = \sum_{m=1}^{M} \ln (1 - r(t_i + m\Delta t)\Delta t) \, , \tag{1.77}$$

using the fact that the logarithm of a product is the sum of the logarithms of the multiplied terms. Using the approximation $\ln(1 - r(t_i + m\Delta t)\Delta t) \approx -r(t_i + m\Delta t)\Delta t$, valid for small Δt, we can simplify this to

$$\ln P[\text{no spikes}] = -\sum_{m=1}^{M} r(t_i + m\Delta t)\Delta t \, . \tag{1.78}$$

In the limit $\Delta t \to 0$, the approximation becomes exact and this sum becomes the integral of $r(t)$ from t_i to t_{i+1},

$$\ln P[\text{no spikes}] = -\int_{t_i}^{t_{i+1}} dt \, r(t) \, . \tag{1.79}$$

Exponentiating this equation gives the result we need

$$P[\text{no spikes}] = \exp\left(-\int_{t_i}^{t_{i+1}} dt \, r(t)\right) \, . \tag{1.80}$$

The probability density $p[t_1, t_2, \ldots, t_n]$ is the product of the densities for the individual spikes and the probabilities of not generating spikes during the interspike intervals, between time 0 and the first spike, and between the time of the last spike and the end of the trial period,

$$\begin{aligned} p[t_1, t_2, \ldots, t_n] \; &= \; \exp\left(-\int_0^{t_1} dt \, r(t)\right) \exp\left(-\int_{t_n}^{T} dt \, r(t)\right) \times \\ & \quad\; r(t_n) \prod_{i=1}^{n-1} r(t_i) \exp\left(-\int_{t_i}^{t_{i+1}} dt \, r(t)\right) \, . \end{aligned} \tag{1.81}$$

The exponentials in this expression all combine because the product of exponentials is the exponential of the sum, and the different integrals in this sum all add up to form a single integral,

$$
\begin{aligned}
& \exp\left(-\int_0^{t_1} dt\, r(t)\right) \exp\left(-\int_{t_n}^{T} dt\, r(t)\right) \prod_{i=1}^{n-1} \exp\left(-\int_{t_i}^{t_{i+1}} dt\, r(t)\right) \\
= \; & \exp\left(-\left(\int_0^{t_1} dt\, r(t) + \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} dt\, r(t) + \int_{t_n}^{T} dt\, r(t)\right)\right) \\
= \; & \exp\left(-\int_0^{T} dt\, r(t)\right) . \qquad (1.82)
\end{aligned}
$$

Substituting this into 1.81 gives the result 1.54.

1.7 Further Reading and References

Further Reading

Cox, DR (1962) *Renewal Theory* London:Methuen.

Cox, DR & Isham, V (1980) *Point Processes* London:Chapman & Hall.

de Boer, E & Kuyper, P (1968) Triggered correlation. *IEEE Biomedical Engineering* **15**:169-179.

Gabbiani F & Koch C (1998) Principles of spike train analysis. In C Koch and I Segev (editors) *Methods in Neuronal Modeling: From Ions fo Networks* Cambridge MA:MIT Press.

Marmarelis, PZ & Marmarelis, VZ (1978) *Analysis of Physiological Systems: The White-Noise Approach*. New York:Plenum Press.

Rieke, FM, Warland, D, de Ruyter van Steveninck, R & Bialek, W (1997) *Spikes: Exploring the Neural Code*. Cambridge, MA:MIT Press.

Tuckwell, HC (1988) *Introduction to Theoretical Neurobiology*. Cambridge, UK:Cambridge University Press.

Werner, G & Mountcastle, VB (1965) The variability of central neural activity in a sensory system, and its implications for the central reflection of sensory events. *Journal of Neurophysiology* **26**:958-977.

Wiener, N (1958) *Nonlinear Problems in Random Theory*. New York:Wiley.

References

Baddeley, R, Abbott, LF, Booth, MJA, Sengpiel, F, Freeman, T, Wakeman, EA & Rolls, ET (1997) Responses of neurons in primary and interior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London* **B264**:1775-1783.

Bair, W & Koch, C (1996) Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Computation* **8**:1185-1202.

Bair, W, Kock, C, Newsome, WT & Britten, KH (1994) Power spectrum analysis of bursting cells in area MT in the behaving monkey. *Journal of Neuroscience* **14**:2870-2892.

Bussgang, JJ (1952) Cross-correlation functions of amplitude-distorted Gaussian signals. *MIT Research Laboratory for Electronic Technology Report* **216**:1-14.

Bussgang, JJ (1975) Cross-correlation functions of amplitude-distorted Gaussian inputs. In AH Haddad, editor, *Nonlinear Systems*. Stroudsburg, PA: Dowden, Hutchinson and Ross.

Engel, AK, Konig, P, Kreiter, AK & Singer, W (1991) Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science* **252**:1177-1179.

Gabbiani, F, Metzner, W, Wessel, R & Koch, C (1996) From stimulus encoding to feature extraction in weakly electric fish. *Nature* **384**:564-567.

Georgopoulos, AP, Kalaska, JK, Caminiti, R & Massey, JT (1982) On the relations between the directions of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience* **2**:1527-1537.

Henry, GH, Dreher, B & Bishop, PO (1974) Orientation specificity of cells in cat striate cortex. *Journal of Neurophysiology* **37**:1394-1409.

Holt, GR, Softky, GW, Koch, C & Douglas, RJ (1996) Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of Neurophysiology* **75**:1806-1814.

O'Keefe, LP, Bair, W & Movshon, JA (1997) Response variability of MT neurons in macaque monkey. *Society for Neuroscience Abstracts* **23**:1125.

O'Keefe, J & Recce, ML (1993) Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**:317-330.

Poggio, GF & Talbot, WH (1981) Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *Journal of Physiology* **315**:469-492.

de Ruyter van Steveninck, R & Bialek, W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: Coding and information transfer in short spike sequences. *Proceedings of the Royal Society of London* **B234**:379-414.

Softky, WR & Koch, C (1992) Cortical cells should spike regularly but do not. *Neural Computation* **4**:643-646.

Theunissen, FE & Miller, JP (1995) Temporal encoding in nervous systems. A rigorous definition. *Journal of Computational Neuroscience* **2**:149-162.

Wandell, B (1995) *Foundations of Vision*. Sunderland, MA:Sinauer.

1.8 Exercises

1. Generate spike sequences with a constant firing rate r_0 using a Poisson spike generator. Then, add a refractory period to the model by allowing the firing rate $r(t)$ to depend on time. Initially, $r(t) = r_0$. After every spike, set $r(t)$ to zero. Allow it to recover exponentially back to r_0 by setting $r(t + \Delta t) = r_0 + (r(t) - r_0)\exp(-\Delta t/\tau_{\mathrm{ref}})$ after every simulation time step Δt in which no spike occurs. The constant τ_{ref} controls the refractory recovery rate. Compute the Fano factor and coefficient of variation, and plot the interspike interval histogram for spike trains generated with and without a refractory period. Plot the coefficient of variation for spike trains with a refractory period as a function of τ_{ref} over the range from 1 to 20 ms.

2. Plot autocorrelation histograms of spike trains generated by a Poisson generator with A) a constant fire rate of 100 Hz, B) a constant firing rate of 100 Hz and a refractory period modeled as in exercise 1 with $\tau_{\mathrm{ref}} = 10$ ms, and C) a variable firing rate $r(t) = 100(1 + \cos(2\pi t/25\ \mathrm{ms}))$ Hz. Plot the histograms over a range from 0 to 100 ms.

3. Generate a Poisson spike train with a time-dependent firing rate $r(t) = 100(1 + \cos(2\pi t/300\ \mathrm{ms}))$ Hz. Estimate the firing rate from this spike train by letting $r_{\mathrm{est}} \rightarrow r_{\mathrm{est}} + 1/\tau_{\mathrm{est}}$ every time a spike occurs, and letting the estimate decay exponentially, $r_{\mathrm{est}} \rightarrow r_{\mathrm{est}}\exp(-\Delta t)/\tau_{\mathrm{est}})$, if no spike occurs during a time step of size Δt. Plot the average squared error of the estimate, $\int dt (r(t) - r_{\mathrm{est}}(t))^2$ as a function of τ_{est} and find the value of τ_{est} that produces the most accurate estimate for this firing pattern.

4. Using the same spike trains as in exercise 3, construct estimates of the firing rate using square, Gaussian, and other types of window functions to see which gives the most accurate estimate.

5. For a constant rate Poisson process, every sequence of N spikes occuring during a given time interval is equally likely. This seems paradoxical because we certainly do not expect to see all N spikes appearing within the first 1% of the time interval. Yet this seems as likely as any other pattern. Resolve this paradox.

6. Build a white-noise stimulus. Plot its autocorrelation function to show that it is uncorrelated and its power spectrum, which should be flat. Discuss the range of validity of these results given the value of Δt you used in constructing the stimulus.

7. Show that if the estimated rate is written in an acausal form (the estimate at time t is allowed to depend on the stimulus at times greater than t)

$$r_{\mathrm{est}}(t) = r_0 + \int_{-\infty}^{\infty} d\tau\, D(\tau) s(t - \tau)\,. \tag{1.83}$$

the optimal kernel for a linear filter estimate of the firing rate when the stimulus is temporally correlated is given by

$$D(\tau) = \frac{1}{2\pi}\int d\omega\, \frac{\tilde{Q}_{rs}(-\omega)}{\tilde{Q}_{ss}(\omega)} \exp(-i\omega\tau) \tag{1.84}$$

where $\tilde{Q}_{rs}$ and $\tilde{Q}_{ss}$ are the Fourier transforms of Q_{rs} and Q_{ss}. You will need to use Fourier transforms and the result concerning the Fourier transform of a convolution integral given in section chapter 12.

8. Construct a spiking model of an orientation selective V1 neuron. Use the tuning curves given in equations 1.17 (with the parameters given in the caption of figure 1.3A) and equation 1.57 (with $\tau_r = 30$ ms) to compute the firing rate. Use a Poisson generator to produce the spikes. Consider a variety of time-dependent orientation angles and study the ability of the spike train to encode both slowly and rapididly varying orientation angles.

9. Build a model neuron (based loosely on the electrosensory lateral-line lobe neuron shown in figure 1.7) using a Poisson generator firing at a rate predicted by equation 1.83 with $r_0 = 50$ Hz and

$$D(\tau) = \cos\left(\frac{2\pi(\tau - 20\text{ ms})}{140\text{ ms}}\right)\exp\left(-\frac{\tau}{60\text{ ms}}\right)\text{ Hz}\,.$$

Use a Gaussian white noise stimulus constructed using a time interval $\Delta t = 10$ ms with $\sigma_s^2 = 10$. Compute the firing rate and spike train for a 10 s period. From these results, compute the spike-triggered average stimulus $C(\tau)$ and the firing rate-stimulus correlation function $Q_{rs}(\tau)$ and compare them with the linear kernel given above. To what degree of accuracy do the relations

$$D(\tau) = \frac{\bar{r}C(\tau)}{\sigma_s^2} = \frac{Q_{rs}(-\tau)}{\sigma_s^2}$$

hold? Repeat this exercise with a static nonlinearity so that the firing rate is given by

$$r(t) = 10(|r_0 + \int_0^\infty d\tau\, D(\tau)s(t-\tau)|)^{1/2}\text{ Hz}$$

rather than by equation 1.83. Show that $C(\tau)$ and $Q_{rs}(-\tau)$ are still proportional to $D(\tau)$ in this case.

10. Derive equation 1.68 by minimizing expression 1.67.