1 The retina

In Section 1.1 we review the basic structure of neurons. Section 1.2 provides a description of the types of neurons found in the retina and their interactions. Finally, we list several characteristics of the signal transmitted from the retina to the rest of the brain in Section 1.3.

1.1 Basic description of a neuron

Neurons are specialized cells that process and transmit information. They typically consist of a cell body (soma), dendrites, and an axon. Communication between neurons occurs at connection points called synapses. At the majority of synapses, signals are sent from the axon of one neuron to a dendrite of another (there are many exceptions though).

Neurons maintain a voltage gradient across their membranes which depends on the intracellular-extracellular concentration difference of ions such as sodium, potassium, chloride, and calcium. Fluctuations in this gradient are induced at the synapses by other neurons through two main mechanisms:

- At chemical synapses the presynaptic neuron releases neurotransmitter molecules that activate receptor molecules. These receptors respond in either of two general ways. They may directly open ion channels in the postsynaptic cell membrane, causing ions to enter or exit the cell, which changes the local membrane potential. Alternatively, the receptor may modulate the production of chemical messengers inside the postsynaptic neuron.

- At gap junctions, which are the location of electrical synapses, there exist channels that cross the membranes of the two connecting neurons. When the membrane potential of one cell changes, ions move between the cells, modifying the membrane potential of the other cell. Unlike chemical synapses, electrical synapses are typically bidirectional.

Both kinds of synapses occur between cells in the retina.

Incremental changes in membrane voltage are called graded potentials, as opposed to all-or-nothing changes called action potentials. Action potentials occur when the membrane voltage of a neuron
reaches a certain threshold. This activates voltage-dependent ion channels. The resulting ion movement across the membrane produces a sharp rise and fall in the membrane potential, which propagates through the axon. The neuron is said to be spiking or firing. Most types of neurons in the retina do not fire—with the exception of ganglion cells, some amacrine cells and perhaps some bipolar cells— but rather rely on graded potentials to convey information. Since graded potentials are more difficult to measure than action potentials, this makes it challenging to characterize neuron interactions within the retina.

1.2 Structure of the retina

In this section we review the main nerve cells in the retina, shown in Figure 1. The exposition is based on [1–4,17,25,26].

![Simple diagram of the organization of the retina. Taken from 4.](image)

**Figure 1:** Simple diagram of the organization of the retina. Taken from [4].

1.2.1 Photoreceptor cells

Photoreceptor cells are the light sensors of the retina. They are mainly connected to neurons called horizontal and bipolar cells, which are described below. In the dark photoreceptor cells release glutamate, a neurotransmitter which either turns OFF (depolarizes) or turns ON (hyperpolarizes) a bipolar cell connected to it. This is known as the dark current. When these cells absorb light they stop releasing glutamate, which excites OFF bipolar cells and inhibits ON bipolar cells.

There are two kinds of photoreceptor cells.

- **Cones** are responsible for vision in condition of high light intensity (a.k.a. photopic vision) and also for color vision. Their response time is quicker than rods. Cones are highly concentrated at the center of the retina—especially in the fovea, an area which provides sharp central vision—and less concentrated in peripheral areas. Each cone contains one of three different photosensitive proteins (or photopsins) that have absorption peaks in the short, medium, and long wavelengths of light. These differing sensitivities are the basis for trichromatic vision.
1. Rods are responsible for vision at low levels of light (a.k.a. scotopic vision). They are concentrated in the outer edges of the retina and are absent from the fovea. They contain a single kind of photopigment, called rhodopsin, and hence cannot distinguish colors.

1.2.2 Horizontal cells

Horizontal cells are connected to photoreceptors and to bipolar cells. They have two main functions.

1. They modulate the sensitivity of photoreceptors under different lighting conditions through inhibitory feedback, so that signaling becomes less sensitive in bright light and more sensitive in dim light.

2. They produce what is known as center-surround antagonism. If an ON bipolar cell is excited by the input of one or a group of neighboring photoreceptors denoted by $P_1$, then it is usually inhibited by the activation of a set of photoreceptors surrounding $P_1$, which we denote by $P_2$. This is mediated by horizontal cells through inhibitory feedback: when any of the cells in $P_2$ are excited, the horizontal cells inhibit the output of $P_1$. It is also possible that horizontal cells produce this effect by interacting with the dendrites of bipolar cells directly, but this remains controversial. In any case, horizontal cells are responsible for the annular structure of ON and OFF bipolar cells (see below).

1.2.3 Bipolar cells

Bipolar cells receive synaptic connections from a number of photoreceptors that varies depending on the region of the retina from one at the fovea to thousands in peripheral areas. Just like rods and cones, bipolar cells produce graded potentials, rather than action potentials (however there has been recent evidence of action potentials in some bipolar cells [6]).

In mammals there are around 11 different types of bipolar cells that synapse with cones and one that synapses with rods. [47], which reports an exhaustive study carried out on mice, concludes that each cone synapses with 10 of these bipolar cells, with the exception of what are known as true blue cones which are just connected to the remaining type. This together with the fact that most types of bipolar cells synapse with distinct types of ganglion cells suggests that each kind of bipolar cell extracts complementary features from the output of the cones, forming parallel channels of information [26]. These features capture patterns in cone activation through different responses to decreased glutamate output.

- **ON-center** bipolar cells are excited by light hitting the center of their receptive field (i.e. decreased glutamate output from a rod or cone corresponding to that region) and inhibited by light hitting the surround region, so that their receptive field is annular. The corresponding feature is helpful in detecting light objects against a darker background.

- **OFF-center** have exactly the opposite response. The corresponding feature is helpful in detecting dark objects against a lighter background.

- **Sustained** bipolar cells react to slow changes in glutamate levels. The corresponding feature detects slowly varying stimuli.
• Transient bipolar cells react to rapid changes in glutamate levels. The corresponding feature detects rapidly varying stimuli.

To clarify, sustained and transient bipolar cells may in turn be ON-center or OFF-center, which yields four different features.

In addition, bipolar cells also receive connections from amacrine cells and horizontal cells. At the moment, the mechanism whereby these cells modulate the output of bipolar cells is not completely understood.

1.2.4 Amacrine cells

Amacrine cells connect to bipolar and ganglion cells and to other amacrine cells. They are responsible for lateral processing, as they provide an indirect path between bipolar and ganglion cells as well as between ganglion cells [14]. The purpose of this processing is largely unknown, although the number of different amacrine cells with very diverse morphologies that have been identified (at least 30 in mammals) suggests that it might provide a variety of complex functionalities [27]. For example, starburst amacrine cells have been shown to give rise to directionally-sensitive receptive fields in certain ganglion cells. To implement such functionalities, amacrine cells use a wide range of neurotransmitters. In general, they produce graded potentials, but some also generate action potentials.

1.2.5 Ganglion cells

One can think of ganglion cells as providing parallel channels that encode visual information through complementary processing mechanisms, which are now thought to be significantly more complex than previously suspected [17, 21]. Ganglion cells receive connections from bipolar and amacrine cells [23] as well as from other ganglion cells (usually of the same type) [22, 45].

The two most important types of ganglion cells are parasol and midget cells.

• Parasol cells receive inputs from relatively many rods and cones. They have large center-surround ON or OFF receptive fields, respond to rapid stimulus changes, can respond to low-contrast stimuli, but are not very sensitive to changes in color. They synapse with magnocellular cells in the lateral geniculate nucleus (LGN).

• Midget cells have small center-surround ON or OFF receptive fields, respond to slow stimulus changes, are sensitive to color, but to changes in contrast unless they are very significant. They synapse with parvocellular cells in the LGN.

Parasol and midget cells make up about 65% of all ganglion cells. The receptive fields of ON and OFF parasols and ON and OFF midgets form four different regular mosaics that uniformly tile the visual scene [19].

Apart from parasol and midget cells, there are at least an additional 13 types that are occur at lower densities [13]. The best known is the bistratified cell, which synapses with koniocellular
cells in the LGN and has a center receptive field without a surround area that is ON to blue stimuli and OFF to both red and green stimuli. Other types include directionally-sensitive ganglion cells and photosensitive ganglion cells, which respond to light stimulus and project to the pretectum (pupillary reflex), to the ventrolateral preoptic nucleus and to several structures involved in the control of circadian rhythms and sleep such as the suprachiasmatic nucleus. At the moment many of the response properties of low-density ganglion-cell types are unknown [17].

1.3 The response of the retina to stimuli

The response of the retina to outer stimuli is transmitted to the rest of the brain in the form of action potentials traveling through the axon of retinal ganglion cells. This section compiles some known facts about these action potentials.

1.3.1 Reaction to simple stimuli

Ganglion cells tend to be responsive to temporal and spatial changes in stimulus intensity [28]. When the ganglion cell detects an intensity step its spiking frequency changes dramatically for a brief period of time. Then it settles back to its previous spiking behavior. The cell will not react to very slow or very rapid changes; in signal processing lingo it reacts to stimuli belonging to a certain frequency band. Similarly, a small spot of light hitting the receptive field elicits stronger responses than the uniform illumination of the entire field of view (when the center and surround areas of the receptive field are simultaneously excited). If the spot is too small there is no response. Spike-triggered averaging, which we describe in Section 2.2, has revealed that the receptive fields of ON and OFF ganglion cells have the antagonistic-surround form described in Section 1.2.3 (this is not surprising, since ganglion cells are driven by bipolar cells).

1.3.2 Adaptation

Experiments show that ganglion cells are less sensitive to the same stimulus after a sustained increase in the general light intensity experienced by the retina. Indeed, the visual processing in the retina adapts dynamically to the mean light level [35]. It also adapts to sustained changes in the variations of light intensity around the mean; a mechanism which is termed contrast adaptation [10].

1.3.3 Variability of spiking pattern

When presented with the same stimulus repeatedly the ganglion cell produces somewhat different spiking patterns. A popular approach is to account for this variability by modeling the firing pattern as the realization of an inhomogeneous Poisson process with an intensity function that depends on the stimulus presented to the cell. This model is inaccurate for at least two reasons:

- Each firing event is followed by a refractory period of around 2 ms during which firing is suppressed. This period is divided into an absolute and a relative refractory period; during the latter firing activity is gradually restored. Refractoriness introduces a dependence between
firing events. An interesting approach to tackle this issue is to decouple the firing rate into a true spiking rate that does not take into account refractory periods and an additional term that models recovery from refactoriness [7] (see also [43]).

- Experiments have found that the location of firing events is significantly less variable across repeated trials than would be expected from a Poisson process [43]. This is manifested in extended periods of time for which no firing occurs across multiple repeated trials and in rapid rises of the spiking rate from zero to its maximum over the whole experiment on a time scale similar to the refractory period between spikes [8].

1.3.4 Firing patterns in neuron populations

Experiments involving a population of ganglion cells have determined that the firing of neighboring cells with similar functional types often tends to be highly synchronous [38] even in the absence of stimulus [5].

In [9], the authors analyze three types of correlation in the firing activity of salamander ganglion cells: broad (synchrony within 40-100 ms), medium (10-50 ms), and narrow (<1 ms). They conclude that broad correlations arise from shared input from common photoreceptors through intermediate neurons via chemical synapses, medium correlations from indirect connections between ganglion cells through amacrine cells via electrical synapses and short correlations from direct connections between ganglion cells via electrical synapses. See also [15] for related experiments on rabbit ganglion cells.

In [33] the spike pattern of a population of neurons is analyzed by grouping together neurons producing simultaneous spikes (within a certain time tolerance) to form different groups (each neuron is allowed to belong to several groups). Cells in the same group tend to be near, but are not necessarily nearest neighbors. The activity of each group is attributed to the activity of an interneuron, an amacrine or bipolar cell which synapses with all the members in the group. The number of active groups was higher in the dark than under a random checkerboard stimulus.

1.3.5 Encoding of visual information in spiking trains

Traditionally, information about the stimulus presented to the retina has been assumed to be encoded in the local spiking rate, or equivalently in the number of spikes over a certain interval of time. However, evidence that visual stimuli are processed very rapidly [16, 41] suggests that this cannot be the only encoding mechanism. The argument is that the interval over which an individual stimulus is processed is so small that it covers a very small number of firing events (in some cases just one). As a result the rate cannot be estimated reliably by other parts of the brain, so the information must be encoded in some other feature of the spiking train [20]. Some evidence to this effect was provided in [18] where temporal structure in spike trains is shown to provide more information about the speed of a moving target in the stimulus to the retina than the time-varying firing rates.

Most previous work analyzes the encoding by looking at individual firing patterns. However, the synchronization of spiking behavior between adjacent ganglion cells described in Section 1.3.4 makes
it very plausible that a joint encoding of stimulus features is taking place. Therefore understanding the encoding will probably involve the analysis of the firing patterns of a whole cell population [31, 33].

2 Statistical modeling

2.1 Statement of the problem

Our goal is to model how the stimuli presented to a region of the retina are related to the spike trains emitted by the retinal ganglion cells. The stimulus consists of a video \( v \) of dimensions \( \bar{a} \times \bar{b} \times \bar{t} \), where \( \bar{a} \times \bar{b} \) correspond to the pixels and \( \bar{t} \) is the number of frames. The data corresponds to \( \bar{c} \) spike trains

\[
 x_c(n) \overset{\text{def}}{=} \begin{cases} 
 1 & \text{if cell } c \text{ spikes at time } n, \\
 0 & \text{otherwise},
\end{cases}
\]

(1)

where \( 1 \leq c \leq \bar{c} \) and \( 1 \leq n \leq \bar{n} \); \( \bar{n} \) denotes the total number of samples. Note that the time discretization of the stimulus and the spike trains is not necessarily the same.

2.2 Spike-triggered averaging

Consider the problem of determining the receptive field of a ganglion cell, i.e. the spatial and temporal region of the stimulus that triggers changes in firing activity. An option would be to use a localized beam of light, modulate its size, intensity, duration and wavelength and then analyze the corresponding firing patterns. However, doing this sequentially for every neuron in a population of ganglion cells would be extremely time consuming. Instead, in this section we describe reverse correlation, an approach that allows to characterize the receptive fields of all the cells simultaneously by efficiently exploring the space of possible stimuli through the use of randomness.

The method consists of computing a quantity known as the spike-triggered average (STA) [11, 24, 32]. The STA of a cell \( c \) is the average stimulus sequence of a certain duration \( d \) preceding a spike in the spiking sequence \( x_c \),

\[
\tilde{v}_n(a, b, m) = v(a, b, n - d + m), \quad 1 \leq a \leq \bar{a}, \quad 1 \leq b \leq \bar{b}, \quad 1 \leq m \leq d,
\]

(2)

\[
\text{STA}_{x_c, v} = \frac{1}{\bar{n}} \sum_{n=1}^{\bar{n}} x_c(n) \tilde{v}_n.
\]

(3)

If the spiking pattern \( x_c \) only depends on the inner product between the stimulus and a response filter \( h \) then the STA reveals this filter, as long as the sequence of stimuli \( \tilde{v}_n \) are uncorrelated (note that this is not the case if they overlap).

**Lemma 1** ([11]). Assume that \( \tilde{v}_1, \tilde{v}_2, \ldots \) are uncorrelated and identically distributed with a discrete probability mass function \( p_\tilde{v} \) such that their variance is finite and

\[
\Pr\{\text{spike at } n|\tilde{v}_n\} = g(\langle h, \tilde{v}_n\rangle),
\]

(4)
for a certain stimulus filter \( h \in \mathbb{R}^{a \times b \times d} \) where \( g \) is an arbitrary nonlinear function. Assume in addition that \( p_{\tilde{v}} \) is symmetric around \( h \), i.e. for any \( k \in \mathbb{R}^{a \times b \times d} \) such that \( h + k \) is in the domain of \( p_{\tilde{v}} \)

\[
p_{\tilde{v}}(h + k) = p_{\tilde{v}}(h - k).
\]

Then the STA converges to a multiple of \( h \).

Proof. Let \( \tilde{v} \) denote a random variable with the same distribution as \( \tilde{v}_1, \tilde{v}_2, \ldots \) and \( 1_{\text{spike}} \) the indicator random variable that is one if a spike occurs (at a given time) and zero otherwise. If \( \Omega \) is the domain of \( \tilde{v} \), we can define a partition \( \{ \Omega_1, \Omega_2 \} \) such that for any \( u_1 \in \Omega_1 \) there is a point \( u_2 \in \Omega_2 \) such that for some values of \( \alpha \in \mathbb{R} \) and \( w \in \mathbb{R}^{a \times b \times d} \)

\[
u_1 = \alpha h + w \quad \text{and} \quad u_2 = \alpha h - w.
\]

By (6)

\[
p_{\tilde{v}}(u_1) = p_{\tilde{v}}(u_2).
\]

The sequence of random variables \( \tilde{v}_1 1_{\text{spike at } 1}, \tilde{v}_2 1_{\text{spike at } 2}, \ldots \) is uncorrelated and identically distributed with finite variance, so by the weak law of large numbers the time average of the sequence–equal to the STA– converges in probability to

\[
E(\tilde{v} 1_{\text{spike}}) = \sum_{u \in \Omega} u p_{\tilde{v}}(u) E(1_{\text{spike}}|u) = \sum_{u \in \Omega} u p_{\tilde{v}}(u) \mathbb{P}\{\text{spike}|u\} = \sum_{u \in \Omega} u p_{\tilde{v}}(u) g(\langle h, u \rangle) = \sum_{u_1 \in \Omega_1} u_1 p_{\tilde{v}}(u_1) g(\langle h, u_1 \rangle) + \sum_{u_2 \in \Omega_2} u_2 p_{\tilde{v}}(u_2) g(\langle h, u_2 \rangle)
\]

\[
= \sum_{\alpha} \sum_{w} 2 \alpha h p_{\tilde{v}}(\alpha h + w) g(\alpha ||h||_2^2) = \beta h,
\]

for a constant \( \beta \).

For ON cells the receptive field is typically revealed as an elliptical area rising above the mean with a concentric region around it that remains below the mean. For OFF cells the pattern is exactly the opposite (see Figures 1 and 2 in [12]).

Once we have obtained an estimate of the response filter \( \hat{h} \) using the STA, it is possible to estimate a nonlinear function linking the local spiking rate to the inner product between the input stimulus and \( \hat{h} \). Since this is a real-valued function of one variable, one can just plot the number of spikes in intervals of a certain length against the corresponding value of the inner product. To produce a smoother estimate, [11] computes the average number of spikes in intervals with similar values of the inner product. The resulting estimate is well approximated by an exponential function. If strong stimuli are used, the exponential function eventually saturates, assuming a sigmoidal shape.

2.2.1 STA of multiple neurons

In [33] the authors compute the STA using simultaneous spikes from groups or two or more cells. Based on an argument similar to the proof of Lemma [1] they claim that if each of the individual cells...
were firing independently then the STA of the group should be a linear combination of the STA of the individual cell. In contrast, the experiments yield joint STAs that resemble the intersection of the individual STAs.

### 2.2.2 Spike-triggered covariance

A generalization of STA is to compute an eigendecomposition of the empirical covariance matrix of the stimuli that trigger impulses, a technique known as spike-triggered covariance (STC) \[34, 39\]. The objective is to find a low-dimensional subspace characterizing the response to the stimuli, rather than just a single filter. This method has been mostly applied to neurons in other parts of the brain.

### 2.3 Generalized linear model with exponential link function

In this section we describe the statistical model proposed in \[31\] (see also \[42\]). The aim is to fit a time-varying spike rate \( \lambda_c \in \mathbb{R}^n \) that quantifies the spiking intensity at each time point. The core of the model is a set of causal linear filters that capture the dependence of the spiking behavior of the cell on the stimulus presented to the retina and also on the past spiking pattern of the cell and of other cells in the population. The spike rate depends on this linear model through an exponential link function, which allows to decompose \( \lambda_c \) in multiplicative components that quantify the contribution of each effect: \( \lambda_{\text{stim}}(\text{stimulus}), \lambda_{\text{auto}}(n) \) (past spiking pattern of the cell), \( \lambda_{\text{coup}} \) (past spiking pattern of other cells), \( \lambda_{\text{base}} \) (baseline spike rate). Note that some of these components may have a canceling effect; this is the case for the autoregressive term \( \lambda_{\text{auto}} \) which is used to model the refractory period of the spikes.

\[
\lambda_c(n) = \lambda_{\text{stim}}(n) \lambda_{\text{auto}}(n) \lambda_{\text{coup}}(n) \lambda_{\text{base}}
\]  

\[
= \exp \left\{ \sum_{c' \in \{0\} \setminus \{c\}} x_{c'} \ast h_{c,c'}(n) + \mu_{\text{base}}^c \right\},
\]

where \( h_{\text{stim}} \) is the stimulus filter, \( h_{\text{auto}} \) is an autoregressive filter that weighs the influence of the past spiking history of the cell, \( h_{\text{coup}} \) a coupling filter that incorporates the effect of the spiking pattern of cell \( c' \neq c \) and \( \mu_{\text{base}}^c \) the logarithm of the baseline spike rate of the cell.

Regarding the choice of link function, in the Supplementary Material of \[31\] the spiking rate within small bins of a certain size plotted against the inner product between the stimulus and a linear filter obtained by computing the spike-triggered average is shown to be approximately exponential (see also Section 2.2). In addition, it is reported that using other output nonlinearities such as a half-wave rectified linear function and a function of the form \( \log(1 + \exp(x)) \) results in much lower likelihoods for the observed data and yields worse cross-validation performance. A more flexible nonlinearity, parametrized by a cubic spline with 8 piecewise polynomial segments, improves crossvalidation results slightly, but does not produce significant changes in the fitted filters.

---

\[1\] In fact, they claim that if the nonlinearity tying the spiking rate and the inner product with the response filter is the same for both cells, then the joint STA should be the sum of the individual STAs (see equations (32) and (33) in [33]). However it is unclear to me that this would be the case, given that the individual filters could be very different.
A diagram of the model given by [13] is shown in Figure 2. The filters $h_{\text{stim}}$, $h_{\text{auto}}$ and $h_{\text{corr}}$ are causal, since our aim is to model the mechanisms that give rise to the spiking pattern of the cell and consequently we do not take into account future stimuli or activity. Low-dimensional representations of the filters, described in Section 2.3.1 are fit by solving the optimization problem

$$\max_{\mathcal{H}} \sum_{n=1}^{\bar{n}} x_c(n) \log \lambda_c(n) - \Delta \sum_{n=1}^{\bar{n}} \lambda_c(n),$$

where the correspondence between $\lambda_c$ and the set of parameters $\mathcal{H}$ is determined by [13]. The solution of [14] corresponds to the maximum likelihood estimate of the intensity function of an inhomogeneous Poisson process generating the data if we assume that the event of more than one spike occurring in an interval of length $\Delta$ has negligible probability [29, 37, 42]. If we renormalize $\lambda_c$ so that it represents the number of spikes per interval of length $\Delta$, then we can set $\Delta = 1$.

### 2.3.1 Low-dimensional representations of the filters in the model

#### Stimulus filter

The stimulus filter is restricted to a $5 \times 5$ pixel area in stimulus space and a duration of 30 frames, which is initially determined by computing the STA as described in Section 2.2. The parametrization can be further simplified by modeling the filter as an outer product of a spatial and a temporal component. Note that this modification makes the cost function [14] non convex [37]. The temporal component may be parametrized using a basis of predefined functions in the same way as the autoregressive and coupling filters (see Section 2.3.1 below). Once it is fitted, this component resembles a smooth filter that changes signs and therefore detects differences in the level of intensity over time.
If one plots the spatial components of the fitted filters for the ON and OFF ganglion cells separately, they both tile the spatial space of the stimulus to cover it completely. The specific form of the spatial components corresponds to elliptical antagonistic center-surround receptive fields. The area covered by the individual receptive fields changes depending on the presence of the coupling term in the model. It is larger if the term is not included. This indicates that the activity triggered by stimuli that excite regions that are far from the center of the receptive field of the cell may be explained more efficiently through the coupling terms (see Supplementary Material of [31]).

Autoregressive and coupling filters

The time filters in the model are parametrized using a basis of raised cosine bumps that allow to model fine temporal structure right after a spike and coarser temporal structure at longer delays. For the autoregressive filter a 10-dimensional basis is used, whereas for the coupling filter the fit is restricted to 4 basis vectors. The main role of the autoregressive filter is to produce a refractory period after each spike by heavily reducing the spiking rate. A sparsifying penalty on the coupling filters is applied to favor solutions for which only a small number of coupling terms are non zero. In the resulting fit the non-zero coupling terms tend to correspond mostly to neighboring cells.

2.3.2 Model evaluation

The peristimulus time histogram (PSTH) of a cell is computed by repeating the same stimulus sequence several times and then computing the histogram of spike events in time bins of a certain width. For some discussion on how to select the width of the bins see [36]. The evaluation of the model described in the previous sections is carried out by comparing its PSTH to the empirical PSTH:

1. We solve (14) to obtain the parameters that characterize the time-varying spike rate.
2. We use these parameters to determine the value of the spike rate $\lambda_c$ corresponding to a certain stimulus sequence $v$.
3. We sample from an inhomogeneous Poisson process with intensity rate $\lambda_c$ to simulate a number of firing sequences.
4. We compare the PSTH of the simulated spikes to the PSTH of real spiking sequences obtained experimentally using $v$ as the input stimulus.

2.3.3 What does the firing rate $\lambda_c$ represent?

As explained in Section 1.3.3, the firing patterns of ganglion cells cannot be modeled precisely as inhomogeneous Poisson processes due to their rapid variability and the existence of a refractory period after each spike. Both of these issues may be tackled by setting the estimated spike rate $\lambda_c$ to either very large or very low values to account for rapid changes. For example, refractory periods can be reproduced by tuning the autoregressive filters $h_{c}^{\text{auto}}$ to zero out or reduce $\lambda_c$ significantly after each spike. Similarly, spiking events that follow a stimulus almost immediately may be replicated
by setting $\lambda_c$ to equal 1 (this means that the rate is one spike per bin). As a result, $\lambda_c$ no longer represents a local firing rate but rather the probability of finding a spike at a given point.

Now the question is whether the cost function (14) yields solutions that represent the spiking probability effectively. Below we show that this is the case for a toy model, where a cell is excited by a constant stimulus that is turned on and off. The resulting estimates for the spike rate in the presence and in the absence of the stimulus are equal to the respective empirical spiking frequencies, which is a reasonable estimate. This holds in the absence of probabilistic assumptions on the process generating the spikes.

**Analysis of a toy model**

In this section we derive the spike rate obtained by solving (14) for a simple toy model. We assume that the stimulus $v$ is one dimensional and piecewise constant. Recall that we denote by $\bar{n}$ the total number of time samples. The stimulus is equal to a constant $k$ at $\bar{n}_{on}$ time steps that are arbitrary (in particular, they are not necessarily contiguous) and equal to 0 at $\bar{n}_{off}$ other time steps, so that $\bar{n} = \bar{n}_{on} + \bar{n}_{off}$. While the stimulus is on, the cell fires $s_{on}$ times, whereas when the stimulus is off the cell fires $s_{off}$ times. We fit a simplified version of (13), where the stimulus filter is parametrized as a single real number $h$,

$$\lambda_c(n) = \exp \left\{ hv(n) + \mu_c^{\text{base}} \right\}. \quad (16)$$

Following (14), we maximize the cost function

$$f(h, \mu_c^{\text{base}}) = \sum_{n=1}^\bar{n} x_c(n) \left( hv(n) + \mu_c^{\text{base}} \right) - \sum_{n=1}^\bar{n} \exp \left\{ hv(n) + \mu_c^{\text{base}} \right\} \quad (17)$$

to fit the model. The following lemma shows that this produces a very reasonable result. Note that there are no probabilistic assumptions on the cell’s firing behavior.

**Lemma 2.** The spiking rate that is obtained from maximizing (17) is equal to

$$\lambda_c(n) = \begin{cases} s_{on}/\bar{n}_{on} & \text{if the stimulus is on,} \\ s_{off}/\bar{n}_{off} & \text{if the stimulus is off.} \end{cases} \quad (18)$$

**Proof.** The cost function (17) is a concave function of $h$ and $\mu_c$. The components of the gradient are equal to

$$\frac{\partial f(h, \mu_c^{\text{base}})}{\partial \mu_c^{\text{base}}} = \sum_{n=1}^\bar{n} x_c(n) - \sum_{n=1}^\bar{n} \exp \left\{ hv(n) + \mu_c^{\text{base}} \right\} \quad (19)$$

$$= s_{on} + s_{off} - \bar{n}_{on} \exp \left\{ kh + \mu_c^{\text{base}} \right\} - \bar{n}_{off} \exp \left\{ \mu_c^{\text{base}} \right\}, \quad (20)$$

$$\frac{\partial f(h, \lambda_c^{\text{base}})}{\partial h} = \sum_{n=1}^\bar{n} x_c(n) v(n) - \sum_{n=1}^\bar{n} v(n) \exp \left\{ hv(n) + \mu_c^{\text{base}} \right\} \quad (21)$$

$$= ks_{on} - k\bar{n}_{on} \exp \left\{ kh + \mu_c^{\text{base}} \right\} \quad (22)$$

12
The result is obtained by setting these components to zero. (22) implies that at the maximizer, the rate when the stimulus is on is equal to

\[ \exp \left\{ kh + \mu_c^{\text{base}} \right\} = \frac{s_{\text{on}}}{n_{\text{on}}}. \]  (23)

Plugging this into (20) gives the optimal rate when the stimulus is off

\[ \exp \left\{ \mu_c^{\text{base}} \right\} = \frac{s_{\text{off}}}{n_{\text{off}}}. \]  (24)

For the sake of completeness, the estimated parameters are equal to

\[ \mu_c^{\text{base}} = \log \frac{s_{\text{off}}}{n_{\text{off}}}, \]  (25)

\[ h = \frac{1}{k} \log \frac{s_{\text{on}} n_{\text{off}}}{s_{\text{off}} n_{\text{on}}}. \]  (26)

3 Proposal

Our aim is to develop a statistical model capable of:

- Reproducing spiking patterns elicited by natural stimuli.
- Characterizing the effect on the firing rate of a ganglion cell of stimulus activity occurring away from its first-order receptive field, represented by the STA.
- Determining to what extent information is encoded in precise spiking times, local spiking rates or synchronization patterns.

The general structure of the model would be:

Stimuli → Input units → Hidden layers → Output unit → Comparison to actual spike trains

We propose exploring alternative options for the different parts of the model.

3.1 What data do we have?

- A sequence of images: \( v \in \mathbb{R}^{a \times b \times t} \).
- Spiking times from a population of \( \bar{c} \) ganglion cells corresponding to \( \bar{j} \) experiments were the sequence \( v \) is shown to the piece of retina. The notation we use for the sequence of spiking times is \( t_c^{(j)}(m) \), which denotes the \( m \)th spiking time in the spike train corresponding to cell \( c \) and experiment \( j \).
If we fix a bin length $\Delta$, we can define a spike train $x_c^{(j)} \in R^n$, where the number of bins $n = \frac{t}{\Delta}$, for each cell and each experiment.

$$x_c^{(j)}(\Delta, n) \overset{\text{def}}{=} \begin{cases} 1 & \text{if there exists an } m \text{ such that } (n - 1) \Delta \leq t_c^{(j)}(m) < n\Delta, \\ 0 & \text{otherwise}. \end{cases}$$  \hspace{1cm} (27)

The corresponding empirical spiking rate is equal to

$$r_c(\Delta, n) \overset{\text{def}}{=} \frac{x_c^{(j)}(\Delta, n)}{j}. \hspace{1cm} (28)$$

### 3.2 What do we want to estimate?

The aim of characterizing a spike train is rather vague. There are several options:

Local spike rate

#### 3.2.1 Local firing rate

### 3.3 Cost function

Beyond the likelihood-based cost function described in Section [2.3](#), a possibility is to try to incorporate metrics that account for more structure in the spike rates, such as the ones described in [44](#), by using ideas from structured prediction.

### 3.4 Input unit

Moving beyond a low-dimensional parametrization based on the STA of the individual cells, we propose using higher-dimensional representations of the input filters, which might reveal more structure in the fitted receptive fields. This increases the risk of overfitting, which may be tackled by regularization and perhaps by fitting shared parameters across the different input units.

### 3.5 Hidden layers

Connected hidden units within the model would allow to model cross-dependence from input units and spiking synchronicity, perhaps mimicking the effect of interneurons such as amacrine cells to some extent. A possible way of obtaining an initial estimate for the connection architecture could be an analysis of synchronous spikes as in [33](#).

### 3.6 Spiking times or spiking rates?

Due to the variability of spiking trains described in Section [1.3.3](#), it is not clear what metric to use as an objective evaluation criterion. A reasonable approach could be to use the rate over segments of different lengths, which for very small segments approaches an evaluation of spiking-time precision.
It may also be possible to estimate a stimulus-dependent metric in a data-driven way by determining whether individual spike times or local rates are more reproducible across repeated experiments.

### 3.7 Other ideas

#### 3.7.1 Analysis of spike-triggering stimuli

In the spirit of the techniques described in Sections 2.2.1 and 2.2.2 it would be interesting to perform a more sophisticated analysis of stimuli that elicit spikes, not only in the case of random stimuli, but also under natural stimuli. This could yield insight into why the generalized linear model does not perform as well for natural stimuli and perhaps also suggest alternative designs for the input units of the model.

#### 3.7.2 Estimating stimuli from spike trains

An interesting way of trying to determine whether information is encoded in precise spiking times, local spiking rates or spiking patterns across cells is to try to estimate stimuli from the spiking trains of a neuron population, as in [30,40,46].

### References

3. [http://thebrain.mcgill.ca/flash/d/d_02/d_02_cl/d_02_cl_vis/d_02_cl_vis.html](http://thebrain.mcgill.ca/flash/d/d_02/d_02_cl/d_02_cl_vis/d_02_cl_vis.html)


