Neuronal Coding in the Retina: Effects of Eye Movements and Network Interactions

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

zur Erlangung des akademischen Grades
Doctor Rerum Naturalium
(Dr. rer. nat.)

vorgelegt von

Christian Bernhard Mendl
aus Ingolstadt

München, den 7. Dezember 2011

angefertigt am
Max-Planck-Institut für Neurobiologie

1. Gutachter:	Prof. Dr. Tim Gollisch
2. Gutachter:	Prof. Dr. Erwin Frey

Tag der mündlichen Prüfung: 1. Februar 2012

Contents

Acknowledgements

First and foremost, I'd like to thank Tim Gollisch for the opportunity to join his lab and for constantly supervising my work. I gained many invaluable insights into the broad neuroscience field during the last three years. I'm especially indebted to Tim for his patient help while drafting my application for the Boehringer Ingelheim PhD fellowship. Moreover, I'm grateful for all the conferences and workshops I could attend during my PhD.

I owe a very nice and supportive atmosphere and many inspiring discussions to my fellow lab members Daniel Bölinger, Vidhyasankar Krishnamoorthy, Mona Garvert, Anna Stöckl, Daisuke Takeshita, Jing Shao, Long Sha, Kristin Völk and Esteban Real. Vidhya generously shared his detailed knowledge about the retina and biology in general; I talked a lot to Daniel about various papers, and conducted the last experiments together with Mona.

I want to thank Erwin Frey for the formal supervision of my PhD and for joining the thesis advisory committee. Special thanks to all members of this committee, consisting of Tim Gollisch, Günther Zeck, Erwin Frey and Andreas Herz for supportive criticism. Additionally, Andreas Herz hosted Daniel and me in Munich after the lab had moved to Göttingen.

I am indebted to the Boehringer Ingelheim Fonds for very generous financial support and several retreats, including the ”Sommerseminar Hirschegg” in 2010 and the Lautrach seminar ”Wissenschaftlich – verständlich” in 2011.

Finally, I'd like to thank the whole MPI of Neurobiology as well as Andreas Herz' group for the inspiring atmosphere and great time, and of course my brothers and parents.

Chapter 1

Introduction

Seeing is one of the most fundamental ways how we perceive our surroundings, and although the physical theory of light is well established, science starts only now to understand the detailed neuronal principles of vision. It is well known that the *retina* at the back of the eye is the central conversion unit from the incoming light stimulus to neuronal activity. Namely, this intricate neuronal network detects the visual stimulus via photoreceptors, processes the visual information and finally relays this information to higher brain areas via electric *action potentials* (or *spikes*), the “language” of the brain. Thus, the retina becomes not only an essential research system for understanding vision, but opens the perspective of gaining fundamental insights into neuronal coding strategies of nervous systems in general.

An important ingredient of the oculomotor strategy to acquire visual information are fast eye movements, denoted *saccades*. For example, when we look at a face, our gaze fixes a certain feature like the mouth, and then jumps to another feature like the nose. In other words, periods of visual “fixation” are interleaved with rapid eye movements. Even though saccades can be attributed to the foveal organization of the (human) retina, these eye movements are also present in animals without fovea (Land 1999). Due to the importance of saccades, a genuine understanding of vision has to include their role for visual information processing.

This thesis aims at elucidating the precise connection between saccades and neuronal coding in the retina. We try to understand whether saccades can elicit spiking responses, how these responses encode information and whether the visual stimulus before, during or after a saccade predominantly determines the neuronal activity. In order to answer these questions, we perform *in vitro* electrophysiological recordings of the action potentials generated by *retinal ganglion cells*, the output neurons of the retina. Their axons form the optic nerve, which transmits the encoded visual information to higher brain areas.

All studies in this thesis are performed on amphibian axolotl and frog retinas since they enable long and stable recordings. In particular, amphibian retinas are much less sensitive to temperature as compared to mammalian systems.

The experimental setup should record the population activity of retinal ganglion cells on the level of single spikes. For that reason, we employ a *multi-electrode array* (MEA), that is, a miniature 8×8 square grid of point-shaped electrodes on a glass plate. We flat-mount the isolated retina (without the optic nerve) onto the plate, which allows for simultaneous recording from several ganglion cells. A computer monitor and a lens project a time-varying visual stimulus directly onto the retina. This enables a systematic study of the input-output relationship based on the information processing in the retina. We imitate saccadic eye movements during experiments via corresponding shifts of the visual image.

We will see how the spike *latency* emerges as salient feature of the neuronal response following a saccade, that is, the *time difference* between the saccade onset and the first elicited spike. Interestingly, spike latencies have attracted attention in recent studies as an alternative neuronal coding strategy (Gollisch and Meister 2008), which allows for faster information processing than the commonly advertised spike rate. Along the same lines, the neuronal response immediately following a saccade might contain the most information about the visual stimulus (Segev et al. 2007).

We will study in detail how the latency depends on the presented (time-dependent) stimulus. Specifically, we will compare the experimental data to a model based on the *receptive field* of a retinal ganglion cell, that is, the area of the visual field a particular ganglion cell is sensitive to. Via theoretical calculations, we derive a simple formula for the activation of the ganglion cell by the visual stimulus. This formula solely depends on the center location of the receptive field. We then investigate how the presented image relative to this location influences the spiking response. In case the visual images before and after a saccade are the same, we study the influence of the time-dependent trajectory of the visual image during a saccade. Thus, our approach inherently takes the important role of saccades for vision into account, which has often been neglected in previous studies.

So far we have considered individual ganglion cells. However, a central topic concerning neuronal *networks* is the question how neurons interact to processes information. Only the concerted activity of billions of neurons in the human brain effect its amazing capabilities. In the retina, information processing strongly depends on the specific connections between and among the various cell classes.

The second part of this thesis aims at a quantification of pairwise and higher-order interaction strengths between retinal ganglion cells. This requires *simultaneous* recordings from several cells, which is precisely the main feature of the MEA setup.

For the quantification, we first follow the approach by Schneidman et al. (2006) and test whether Ising models from statistical physics (that is, *pairwise* interactions) can reproduce the spiking pattern probabilities of a ganglion cell population. Ising models have been extensively studied in physics; in their original form, they describe a regular arrangement of electronic spins which influence their neighbors. Since we are interested in the internal wiring of the retina, we record the *spontaneous* neuronal activity, in the absence of any visual stimulus, and thus omit potential stimulus-induced correlations. This is different from the study by Schneidman et al. (2006), who present "natural movies" to the retina. Next, we explicitly include third-order interactions into the model, and quantify the improvement for reproducing the experimental spiking statistics. Thus, we can estimate the importance of third-order interactions in the retina.

Outline of the thesis In chapter 2, we illustrate the laminar organization of the various cell types in the retina, and provide an overview of research results concerning (saccadic) eye movements and information processing in the retina.

We explain the experimental methods of this thesis in chapter 3, including a short introduction of the MEA (multi-electrode array) setup. In particular, we describe the difficulties associated with extracting individual action potentials from the recorded voltage traces.

Chapter 4 constitutes the first main results part of the present thesis, where we investigate neuronal coding schemes in the presence of saccadic eye movements. The last section of the chapter describes the effects of pharmacology on information processing in the retina.

In chapter 5, we review several concepts from statistical physics and information theory applied to neuronal networks, in order to provide the theoretical background for the following chapter.

Finally, in chapter 6, we apply these concepts to the spontaneous spiking activity of a neuronal population. We first test whether the Ising model (i.e., pairwise interactions) suffices to describe the spiking statistics, and then compare it to an extended model with explicit third-order interactions.

Chapter 2

Overview: Research on Retina

This chapter reviews several aspects of the present neurobiological research on the retina. We start with a broad introduction to the anatomy of the eye and the retina at the back of the eyecup, then focus specifically on the laminar organization of the various cell types in the retina and finally shift the attention to saccadic eye movements, which are particularly important for the present thesis.

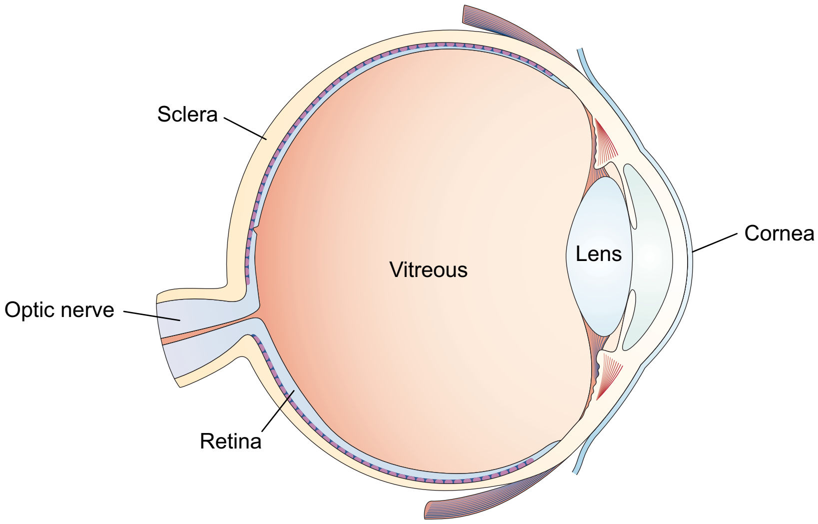


Figure 2.1: Schematic illustration of the human eye (image adapted from Martinez-Conde et al. (2004), with permission). Light enters the eye from the right and is projected via the lens onto the retina, which is located at the opposite side of the eyecup. After several processing steps inside the retina, the retinal ganglion cells transmit the visual information via their axons to higher brain areas. Note that the optic nerve consists precisely of these axons.

Figure 2.1 shows an overview of the anatomical organization of the eye, with the retina attached to the back part of the eyecup. Note that the retina is the first neuronal processing unit of visual information and thus part of the brain. The optic nerve consists of the axons of the *retinal ganglion cells* and transfers the visual information to higher brain areas.

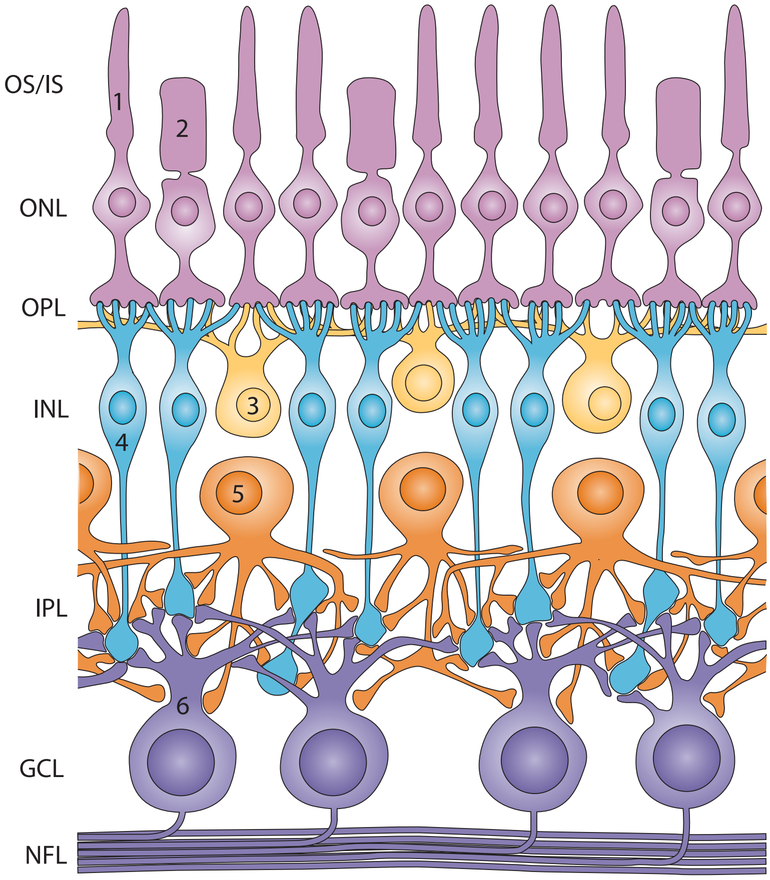


Figure 2.2: Laminar structure of the retina (image adapted from Wässle (2004), with permission). The retina consists of several cell classes, namely: rods (1), cones (2), horizontal cells (3), bipolar cells (4), amacrine cells (5) and retinal ganglion cells (6). The various layers read ONL (outer nuclear layer, containing the photoreceptors), OPL (outer plexiform layer), INL (inner nuclear layer containing the bipolar and horizontal cells), IPL (inner plexiform layer), GCL (ganglion cell layer), and NFL (optic nerve fiber layer).

Research on neuronal tissues and the retina in particular dates back to the pioneering work by Cajal (1894). Already at the end of the 19th century, Cajal characterized several cell types inside the retina and described its laminar organization (see figure 2.2). Specifically, when light reaches the retina, it is detected by two classes of photoreceptors, rods and cones. Broadly speaking, cones are responsible for color vision at bright (photopic) daylight conditions, whereas rods can detect single photons at low (scotopic) light levels (Wässle 2004). The photoreceptors transfer the visual information via synapses and electric gap junctions to horizontal and bipolar cells. Retinal ganglion cells integrate the signals from bipolar cells either directly or via intermediate amacrine cells, which establish a lateral signal pathway in the retina between bipolar cells and ganglion cells. The axons of the ganglion cells comprise the optic nerve, which relays the whole visual information to specialized brain regions.

We will provide more detailed information about the synaptic connectivity between the various cell classes in the following subsection.

2.1 Information Processing in the Retina

This section reviews research results concerning the information processing in the retina and the corresponding neuronal pathways involved in the processing.

The retina cannot be regarded as the biological counterpart to a digital camera, which faithfully records an image pixel by pixel and then passes this information along. Instead, the neurons in the retina already process the incoming visual information, with specialized ganglion cell subtypes transmitting different features on parallel pathways (Masland 2001; Wässle 2004; Gollisch and Meister 2010). In their seminal (and by now classical) paper, Lettvin et al. (1959) reported this essential finding already more than half a century ago. Namely, some ganglion cells of frogs specifically encode the local contrast of the visual image, and others the movement of edges. Similarly for primates, studies have found several distinct ganglion cell types characterized by morphology, light response properties and projections to higher brain areas (see review by Field and Chichilnisky (2007)).

Signaling pathways in the retina are distinct neuronal circuits which transmit visual information from photoreceptors to ganglion cells (see figure 2.3). Rather counter-intuitively, photoreceptors continuously emit neurotransmitters during darkness, and are *hyperpolarized* by light, i.e., the membrane potential changes further to the negative. As a result, neurotransmitter emission decreases with increasing light intensity (ON), and correspondingly, neurotransmitter emission increases with decreasing light intensity (OFF). The retina has two distinct ON and OFF pathways originating at the bipolar cell level. That is, there are subclasses of bipolar cells, denoted ON bipolar cells, which specifically receive signals from either rods or cones and report increasing light intensity, and correspondingly OFF bipolar cells reporting decreasing light intensity (Masland 2001). Thus, OFF bipolars relay the incoming neurotransmitter signal, whereas ON bipo-

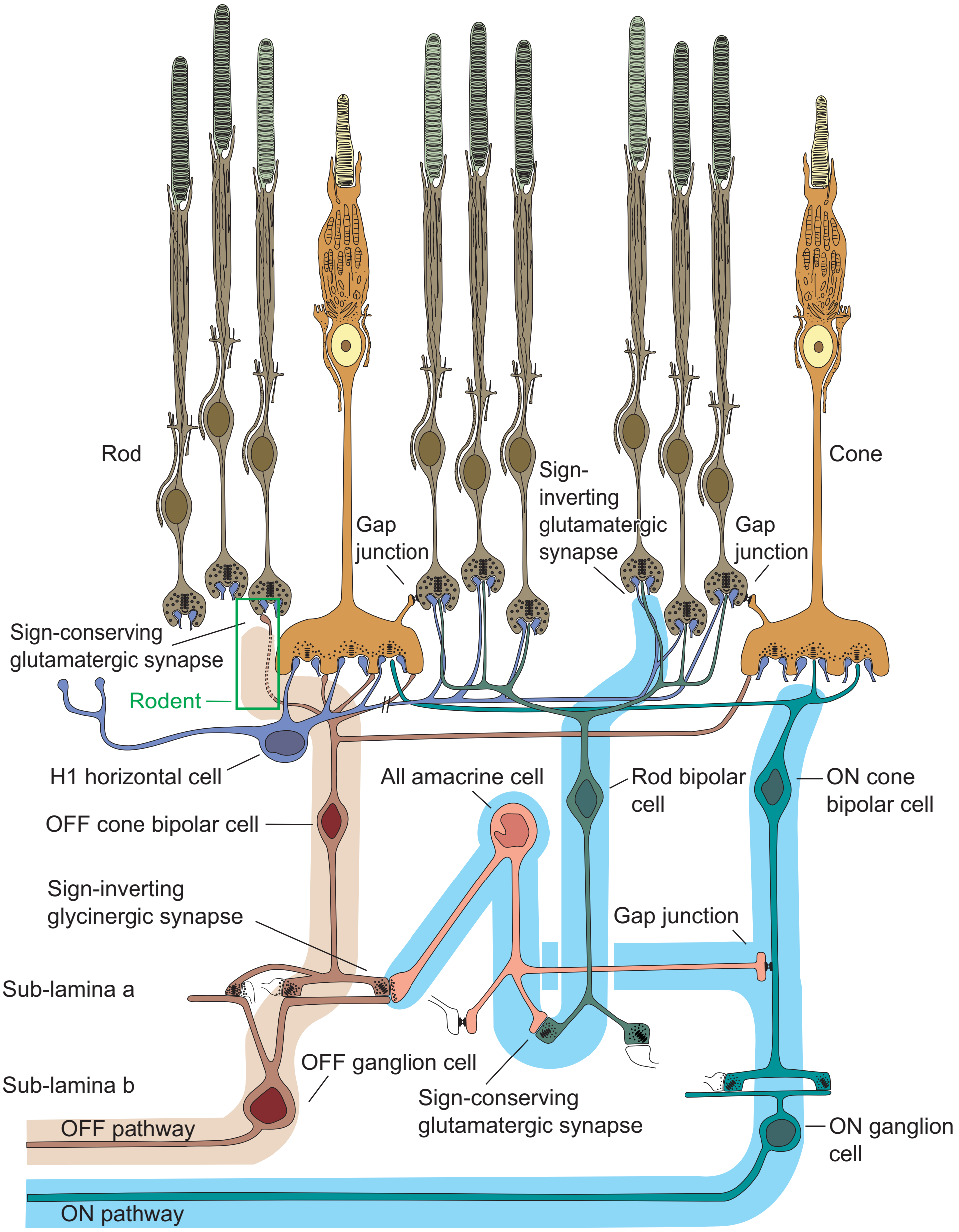


Figure 2.3: Rod and cone pathways in the mammalian retina (image source: Sharpe and Stockman (1999), with permission). Several distinct circuits within the retina transmit the signals from the photoreceptors (rods and cones) to the retinal ganglion cells.

lar cells are driven by increasing illumination thanks to a sign-inverting, metabotropic glutamate receptor. The distinction between ON and OFF pathways is reflected in ON and OFF sublayers within the inner plexiform layer (Masland 2001); see also figure 2.2. Thus, the dendrites of pure ON and OFF ganglion cells stratify in the respective sublayer. Nevertheless, there are also ON-OFF ganglion cells which receive input from both layers. Interestingly, cone pathways in the retina are evolutionary older than rod pathways (Masland 2001). Accordingly, several studies point to the conclusion that rod pathways make use of existing cone pathways (Sharpe and Stockman 1999).

Motion detection and processing is an essential aspect of vision, for example when a frog tries to catch a fly. It is relevant for the present thesis since saccadic eye movements result in a corresponding shift of the visual image on the retina. Already almost half a century ago, Barlow and Hill (1963); Barlow et al. (1964); Barlow and Levick (1965) specifically investigated motion processing in the rabbit retina and showed that certain ganglion cells report the motion *direction*. Recently, Kay et al. (2011) have characterized intrinsically direction selective ganglion cells in mice. Studies have even elucidated the dendritic computations effecting this direction selectivity (Taylor et al. 2000). In a similar vein, Kim et al. (2008) identified a subclass of retinal ganglion cells which selectively respond to *upward* motion, by combining genetic marking with physiology. In detail, the authors use transgenic mice to mark a subtype of OFF ganglion cells with asymmetric, systematically aligned dendritic arbors, and relate the morphologic structure of the dendritic tree to function (detect the direction of the visual stimulus).

In the context of saccades, Ölveczky et al. (2003) investigate how the retina can separate object from background motion, i.e., detect moving objects behind a stationary background, which nevertheless “moves” on the retina due to saccades. Indeed, the authors find and characterize a subset of retinal ganglion cells – denoted “object motion sensitive” (OMS) cells – which accomplish this task. They explain the selectivity for differential motion (object versus background) by a circuit model involving nonlinear rectification (e.g., via bipolar cells (Demb et al. 2001)) and inhibitory, polyaxonal, wide-field amacrine cells. The large receptive field of these amacrine cells is required to suppress responses elicited by *global* background motion. The model has been verified by intracellular recordings from interneurons (Baccus et al. 2008). Interestingly, OMS ganglion cells respond strongest at the onset of differential motion, and then gradually adapt to the motion (Ölveczky et al. 2007).

How precisely groups of neurons interact to process information is a largely open question in neurobiological research. With respect to motion processing, Berry et al. (1999) show how the *population* activity of retinal ganglion cells can anticipate the correct position of a moving object, correcting for the time delays in the signal transduction in the retina.

Münch et al. (2009) identify a ganglion cell type in the mouse retina which specifically responds to approaching visual objects, which is important for escape behaviors, for example. An interesting aspect of this study is the concrete example of a neural pathway with reversed signal directions under different physiological conditions.

Neuronal coding schemes are of central interest in neurobiological research, and a particular topic of the present thesis (see section 4.2). For example, the question arises whether retinal ganglion cells transmit the visual information via the "spike rate" (number of action potential per time), or whether single spikes already contain information about a stimulus.

An important aspect is the precision and reproducibility of individual spikes for several repeats of the same stimulus. Meister and Berry (1999) review this topic, including models of neuronal coding. Concerning temporal precision, Berry et al. (1997) report that the timing of single spikes can convey much more information about their particular stimulus than the spike count.

Warland et al. (1997) construct an artificial neural network to decode visual information from the spike trains of biological ganglion cells. In this study, the stimulus consists of a spatially uniform, time-varying illumination. The authors conclude that the optimal interpretation of a ganglion cell's action potential depends strongly on the simultaneous activity of other nearby cells, thus shedding light on the relevance of population codes.

Berry and Meister (1998) investigate how the refractory period influences the precise timing of individual spikes. In detail, the authors construct a model which predicts the first spike based on the stimulus alone, but includes the timing of preceding spikes to take the refractory period into account. The study is relevant for an analysis of inter-spike-intervals (ISIs) in bursts of spikes (see also section 4.2).

Humans can detect image features which are smaller than a single photoreceptor (Poggio et al. 1992). This remarkable ability is referred to as *hyper-acuity*, and the question arises how the visual system accomplishes it. One idea to answer this question stems from Ahissar and Arieli (2001), who relate spatial to temporal coding. Specifically, the authors consider the adjacent receptive fields of two close retinal ganglion cells, such that the retina translates two moving black-white borders with a small spatial offset to a temporal delay between the responses of these ganglion cells.

Gollisch and Meister (2008) show that certain ganglion cells in salamander retina transmit a briefly presented visual image via the relative timing of their first spikes (i.e., spike latencies). This will become a central aspect of the present thesis in chapter 4.

Modeling refers to establishing mathematical models of the spiking behavior of neurons, and is an important tool to elucidate and understand the effective input-output relationship of neurons. Figure 2.4 shows the basic ingredients of the widely employed LN (linear-nonlinear) model: a spatio-temporal filter (green and blue) followed by a nonlinear rectification (red) – hence the name "linear-nonlinear". With further refinements, these models describe the spiking behavior of single retinal ganglion cells quite well. The spatial filter (blue ellipse in figure 2.4) corresponds to the *receptive field* of the ganglion cell, i.e., the region of the visual field the ganglion cell is sensitive to. Keat et al. (2001) add a feedback pathway to the LN model in order to accurately predict single spikes based on the preceding visual stimulus.

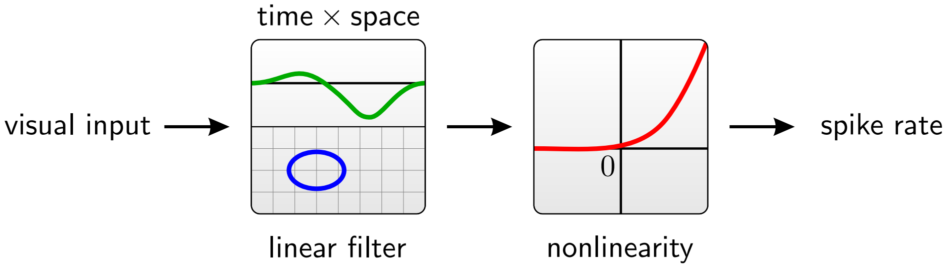


Figure 2.4: Schematic illustration a simple LN (linear-nonlinear) model describing an individual retinal ganglion cell: the visual input first passes a linear filter (in time and space dimension), and is then then rectified by a nonlinear function, which sets its negative input to zero.

To describe neuronal interactions, Pillow et al. (2008) fit a generalized linear model (GLM) with explicit coupling filters to experimental recordings from monkeys, with the idea to capture neuronal correlations. Indeed, the authors report that correlations notably contribute to the transmitted information.

Neuronal networks determine how groups of neurons interact to process information. Many key questions are still unsolved, in particular since parallel recordings from a large number of neurons are technically challenging. One viable solution might be *multi-electrode arrays*, which we employ in the present thesis (see section 3.2 and chapter 6).

Synchronous firing patterns are a central aspect of neuronal interactions. Concerning the underlying mechanisms in the retina, firing synchrony within $40 - 100\,\mathrm{ms}$ might be attributable to shared photoreceptors connected to the ganglion cells via chemical synapses, synchrony within $10 - 50\,\mathrm{ms}$ attributable to amacrine cells which forward signals to the ganglion cells via electrical junctions, and very "narrow" synchrony (< 1 ms) due to electrical junctions between ganglion cells (Brivanlou et al. 1998). Interestingly, the receptive field corresponding to synchronous firing patterns of groups of neurons is smaller than that of individual cells, according to Schnitzer and Meister (2003). Synchrony might thus contribute to a sharper resolution of the visual image.

Applying concepts from physics to neuronal networks dates back to the famous Hopfield networks (Hopfield 1982) in the context of memory. More recently, information theory has emerged as a valuable tool to quantitatively assess neuronal network processes. We will provide a detailed review of several approaches in chapter 5, and apply them to our own experimental data in chapter 6. Very briefly, Schneidman et al. (2006); Tkacik et al. (2006) and Shlens et al. (2006) investigate whether Ising models are sufficient to capture the spiking statistics of neuronal populations (see section 5.3). Schneidman et al. (2003a) and Puchalla et al. (2005) introduce the concept of synergy/redundancy of (pairs of) neurons (section 5.5). Finally, Nemenman et al. (2002) estimate the *entropy* of spike pattern probabilities while correcting for subsampling bias effects (section 5.6).

2.2 Saccades and Fixational Eye movements

Information processing by retinal ganglion cells in the presence of saccadic eye movements is a main topic of this thesis (see chapter 4). The following subsection provides an overview of saccades and fixational eye movements.

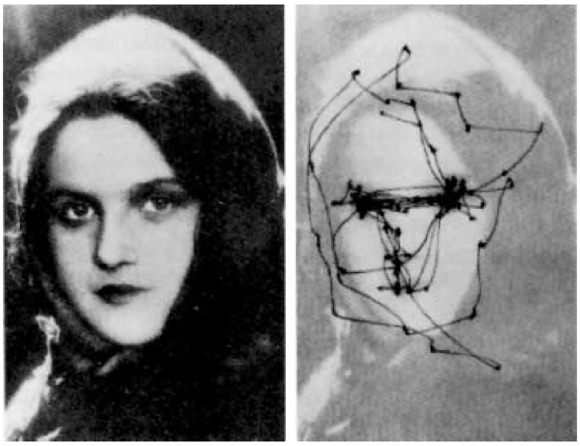


Figure 2.5: (Saccadic) eye movements during free examination of the photograph for one minute, as recorded by an eye suction device (Yarbus (1967); reprinted image from Martinez-Conde et al. (2004), with permission). Note that the viewer mainly fixates the salient features of the image, in particular, the eyes of the face.

Human saccadic eye movements have already been investigated more than 40 years ago by Yarbus (1967), using an eye suction device attached to the human eye. This effective but rather invasive method has been superseded by modern optical eye tracking devices, but the biological findings remain valid. Figure 2.5 illustrates the main aspect of saccadic eye movements. Namely, the human vision jumps from one "interesting" feature to the next. As Yarbus (1967) points out, "interesting" is based on most useful information for the observer, and cannot simply be derived from the level of detail.

Given the notion of rapid eye movements, we expect that the image on the retina gets blurred during saccades (analogous to a shaky photograph), and thus should be blocked from upstream brain areas to maintain acute vision. Termed *saccadic suppression*, Burr et al. (1994) indeed report that brain areas for motion processing are selectively suppressed. This is in line with the fact that saccadic eye movements remain largely unnoticed to humans; that is, we experience stable surroundings instead of a shaking of the visual scene, which could conceivably result from these rapid eye movements.

Fixational eye movements in humans are microscopic and unnoticed eye movements while focusing the gaze on a visual object of interest, e.g., on the salient features of the face in figure 2.5 (Martinez-Conde et al. 2004, 2006; Martinez-Conde and Macknik 2007; Martinez-Conde et al. 2009). *Fixation* refers to this focusing of the gaze; contrary to what the name suggests, human eyes do not remain completely static during these fixation periods. Fixational eye movements are a constant feature of human vision, although we are not aware of it. Namely, they are believed to counteract neuronal adaptation and visual perception fading due to the small shifts of the image on the retina (Martinez-Conde et al. 2006). In more detail, fixational eye movements consist of three main components (at least in humans): tremor, drifts and microsaccades, as illustrated in figure 2.6. Tremor has a high-frequency ($\approx 90\,\mathrm{Hz}$) and tiny amplitude (about the diameter of a single cone in the fovea), and seems to be independent in the two eyes (Martinez-Conde et al. 2004). As the name suggests, microsaccades are small involuntary saccades with $\approx 25\,\mathrm{ms}$ duration (Ditchburn 1980).

The question whether microsaccades actually contribute to visual acuity has long been debated. Ko et al. (2010) try to resolve this question and elucidate the role of microsaccades in detail. The authors report that microsaccades accurately move the eye to regions of interest and are dynamically modulated when aligning a virtual horizontal bar. Thus, the study indicates that the visual system employs microsaccades to increase the level of spatial detail.

Given the importance of fixational eye movements, the question arises

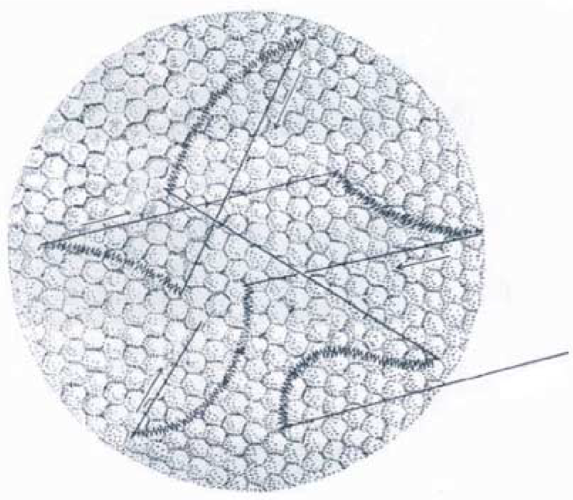


Figure 2.6: Fixational eye movements result in shifts of the image on the retina, as illustrated by the trajectory (image adapted from Martinez-Conde et al. (2004), with permission). The honeycomb structure shows single photoreceptors. Different components of fixational eye movements can clearly be distinguished: high-frequency tremor with amplitude less than a photoreceptor, drift (curved lines) and microsaccades (straight lines).

how they are actually used by neuronal circuits in the retina and upstream brain regions to arrive at sharper vision. To tackle this question, Pitkow et al. (2007) and Burak et al. (2010) simulate fixational eye movements by a random walk trajectory and construct a Markov decoder model to discriminate between two stimuli, given the respective retinal spike trains. To estimate human acuity experimentally, the authors ask subjects to distinguish between two tiny oriented bars, which correspond to only a few photoreceptors on the fovea. The authors report that their model matches the experimental acuity and show how the model could be realized by a biological neuronal network.

Eye movements of animals are quite abundant according to the review by Land (1999). This is particular interesting since most vertebrates lack foveas, which necessitate saccadic eye movements in primates to assemble high-resolution images of a scene. From an evolutionary perspective, the author offers the explanation that animals compensate for body movements, i.e., try to keep the gaze still during locomotion, in order to avoid motion blur if photoreceptors respond slowly. Note that, besides actually moving the eyes themselves, animals might alternatively move the head or the whole body.

Werner and Himstedt (1985) specifically analyze fire salamander head movements when fixating prey. This study is important for the present thesis since we work with salamanders, too (see chapter 3). Two main findings of their study are that angular velocity and duration of head movements depend on the turning angle, and that there exists a *feedback loop* during the saccade (for example when the prey moves away), which can slightly alter the amplitude of the head movement.

Greschner et al. (2002) use video-oculography to measure fixational eye and head movements of turtles, and subsequently imitate the corresponding image shifts when recording from the isolated turtle retina in vitro. According to their physiological measurements, the fixational eye movements have a frequency of $\approx 5\,\mathrm{Hz}$ and an amplitude of $\approx 5\,\mu\mathrm{m}$ on the retina. The authors show that synchronous spiking of retinal ganglion cells contributes to the information about stimulus features.

Roska and Werblin (2003) investigate inhibition induced by global image shifts (imitating saccadic eye movements) via patch-clamp recording on rabbits. The authors show that transient inhibition occurs within a specific layer of the IPL (inner plexiform layer), affecting certain subtypes of retinal ganglion cells.

Segev et al. (2007) employ archer fish as model system to investigate how eye movements contribute to retinal coding of visual information. These fish shoot down insects from below the water surface by squirting water at them. The first goal of the study is the *in vivo* measurement of eye movements using a search coil setup. The authors train several archer fishes to discriminate virtual flies of different sizes, and find that archer fish eye movements consist of saccades and fixations, where the fixation periods have a drift and tremor component (quite similar to human eye movements). The measured saccadic movement speed ($3-6\,\mu\mathrm{m/ms}$) can serve as guideline for our visual stimulation parameters. In the second part of their study, the authors record from retinal ganglion cells *in vitro* using a multi-electrode setup (similar to the present thesis, see chapter 3). They imitate eye movements via corresponding shifts of the presented visual image, explicitly simulating superimposed drift and tremor at $5\,\mathrm{Hz}$ with tremor amplitude $\approx 0.2^\circ \simeq 12\,\mu\mathrm{m}$ on the retina. The authors show that the presented object size can be decoded from the spike trains of several ganglion cells. Interestingly, most visual information about the object size is conveyed at the beginning of a fixation according to the study. We will compare this finding with our experimental data in chapter 4.

Chapter 3

Experimental Methods

This chapter explains the *experimental* methods employed in our studies. The details of the theoretical models and concepts used for analyzing the data will be given later.

Figure 3.1: Melanoid axolotl (image source: mudfooted.com)

We study amphibian retina of adult salamanders (axolotl, show in figure 3.1) and frogs (Xenopus laevis), with all data shown in chapter 4 stemming from axolotl. Analysis is performed in vitro, that is, the retina is extracted from the eye cup and then flat-mounted onto a microelectrode array (MEA). This array consists of a glass plate with 60 electrodes on it, arranged in a regular 8 × 8 square grid (with the 4 corner electrodes missing). Neighboring electrodes are 100 µm apart, approximately corresponding to one third of the dendritic tree diameter of a retinal ganglion cell. The electrodes pick up the extracellular potentials generated by the action potentials of individual retinal ganglion cells and transmit the amplified signals to a recording computer.

3.1 Retina Preparation

Animals were dark-adapted and anesthetized in an ice bath for at least 30 minutes before the actual preparation, and then decapitated prior to enucleating the eyes. Anesthetization without drugs keeps the retina in its natural condition as far as possible. Since salamanders and frogs are cold-blooded, the preparation can be performed at room temperature, different from working with mammalian retina of mouse, for example.

The retina is a soft, fragile tissue at the back part of the eye cup, with an approximate diameter of 4 mm in salamanders. Due to the tiny dimensions, we employ a microscope for preparation. A razor blade hemisects the eye into front part (containing the lens) and back part with the retina curled in it. Next, we use a smooth, bent glass-stick to gently separate the retina from the pigment epithelium (PE) attached to the photoreceptors. The PE is a black, thin cell layer between retina and eye cup mainly responsible for nutrition of the photoreceptors as well as light absorption. Finally, once extracted from the eye, we cut the retina into two halves and keep one half for later usage. The first half is mounted onto a semipermeable membrane and then transferred to the recording (MEA) setup. The membrane slightly presses the retina onto the MEA glass plate for fixation. To keep the retina alive as long as possible, we superfuse it with bicarbonate Ringer's solution containing 110 mM sodium chloride (NaCl), 2.5 mM potassium chloride (KCl), 1 mM calcium chloride ($CaCl_2$), 1.6 mM magnesium chloride ($MgCl_2$), 22 mM sodium bicarbonate ($NaHCO_3$), and additionally 10 mM D-glucose for nutrition. The solution is equilibrated with carbogen (95% O_2, 5% CO_2) to supply oxygen and fix the pH level at 7.3 – 7.4.

3.2 Visual Stimulation and Recording

Figure 3.2 schematically illustrates the recording setup. A computer monitor generates the visual stimulus, and a multi-electrode array (MEA) extracellularly records the simultaneous action potentials of several ganglion cells in the retina. In the lab there were two almost identical versions of the setup available, such that both retinas from one animal could be measured simultaneously. This way we maximize the amount of data acquired per time. Moreover, since the isolated retinas slowly degrade over time, instantaneous recordings are preferable as compared to keeping the second retina alive during the recording timespan of the first.

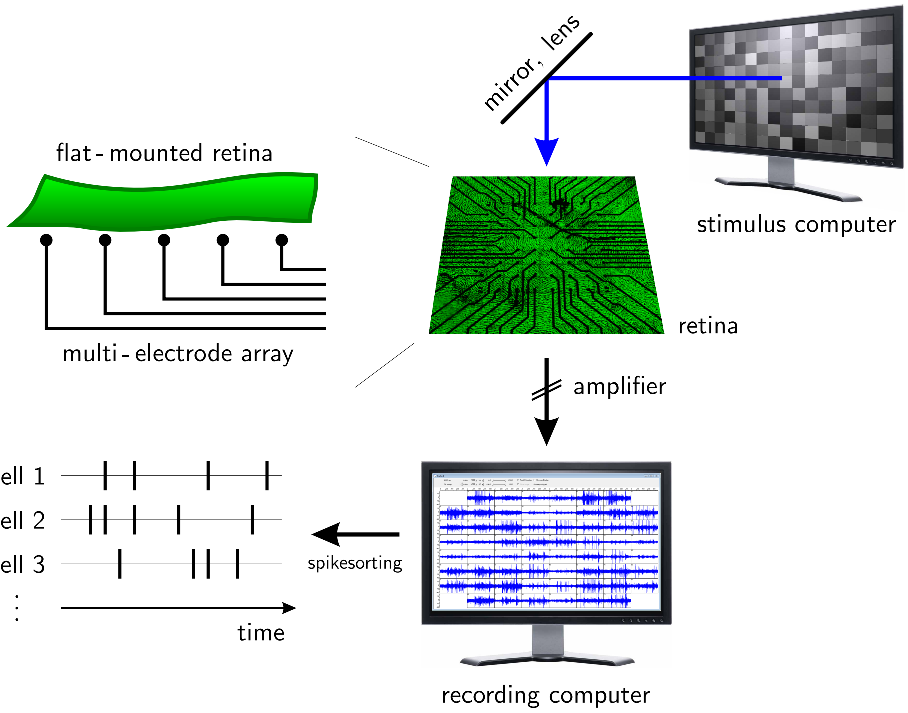


Figure 3.2: Schematic illustration of the experimental recording setup. The time-varying visual stimulus is scaled down by a lens and projected onto the retina. A multi-electrode array (MEA) extracellularly records the simultaneous output signals (action potentials) of several retinal ganglion cells. These voltage traces are then amplified and stored on the recording computer for further processing.

	setup 1	setup 2
intensity per pixel [10^{-12} W]	0.331	0.684
pixel size on retina [μm × μm]	5.65 × 5.75	6.0 × 6.25
intensity per area [10^{-3} W/m^2]	10.19	18.23

Table 3.1: Light intensities generated by the stimulus monitors and reaching the retina, for the two experimental setups

We employ standard CRT monitors for visual stimulation, with 100 Hz refresh rate and one pixel on the monitor approximately corresponding to 6 μm on the retina. Table 3.1 shows the light intensity values for both setups, measured with a calibrated photodiode amplifier.

The amplifier is built by MultiChannelSystems GmbH (Reutlingen, Germany). Concretely, we use the MEA60-Inv-System together with a 64-channel PCI-bus data acquisition card inside the recording computer. The voltage traces are sampled at 25 kHz on each channel (i.e., electrode), with

16 bit resolution. A band-pass filter passes frequencies between 300 Hz and 5 kHz. It consists of a high-pass filter with cutoff frequency 300 Hz implemented in software, and a low-pass filter with cutoff frequency 5 kHz implemented in hardware.

3.3 Spikesorting

In our context, *spikesorting* refers to the extraction of spike timings from the recorded voltage traces, and the assignment of these spikes to biological cells. As illustration, figure 3.3 shows such a voltage trace containing a burst of spikes. The accepted view holds that the spikes originating from an individual cell have the same stereotypic shape. Thus the shape itself conveys no information. Consequently, only the *time point* of each spike elicited by a cell has to be stored, which considerably reduces the amount of data.

A common approach consists of detecting spikes simply by threshold crossings. That is, whenever the voltage passes a certain value (red line in figure 3.3), the following 2 or 3 ms are interpreted as a spike. While this is quite straightforward to implement, a critical issue concerning extra-cellular multi-electrode setups (contrary to the patch clamp technique, for example) is the mapping between recording channels and biological cells. Namely,

i) a channel can receive input from more than one cell, or – conversely –

ii) several channels may pick up spike signals originating from a single cell.

To address these problems, a spikesorting algorithm typically identifies cells based on their spike shape, which acts as fingerprint for each cell. (This is no contradiction to the above statement that spike shapes convey no information within a neuronal network.) For our purposes, the “spike shape” consists of the vector of successively sampled voltages. A mathematical clustering algorithm sorts these vectors into various clusters (i.e., groups), such that each cluster corresponds to a cell. This solves (i). To treat (ii), we look for channels with (almost) coinciding spike timings, indicating a common source. We subsume each such group of channels into a virtual single channel. The spike shape now consists of the original voltage traces appended one onto another.

For employ a custom software package implementing the above features (Pouzat et al. 2002). It is built on top of `IgorPro 6.0` (WaveMetrics, Inc., Oregon, USA).

We use this package for offline analysis of the recorded data. The spike quality of each cell is rated on a scale from $1, 1.3, 1.7, 2, 2.3 \dots 5$, with 1 denoting a very reliable cell (see e.g. figure 3.3) and 5 mostly noise. The quality depends on the signal to noise ratio of the voltage trace, and whether the cell is easily distinguishable from other cells.

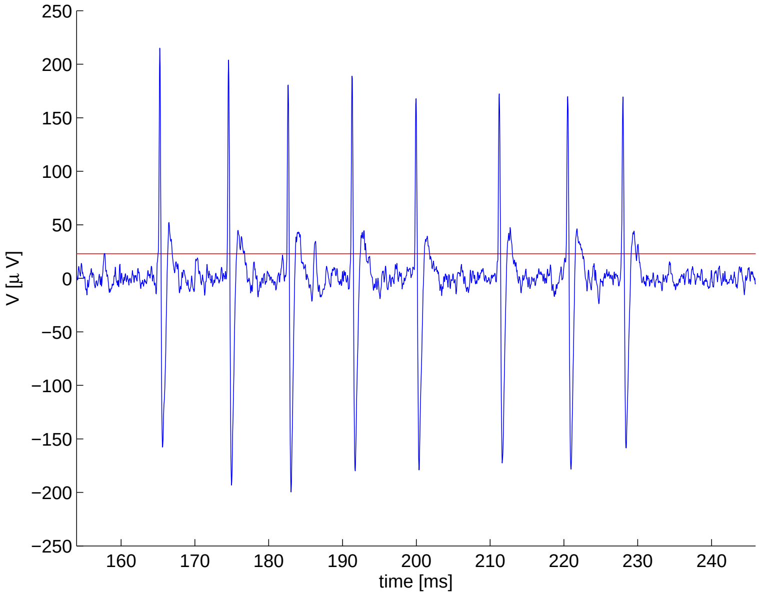
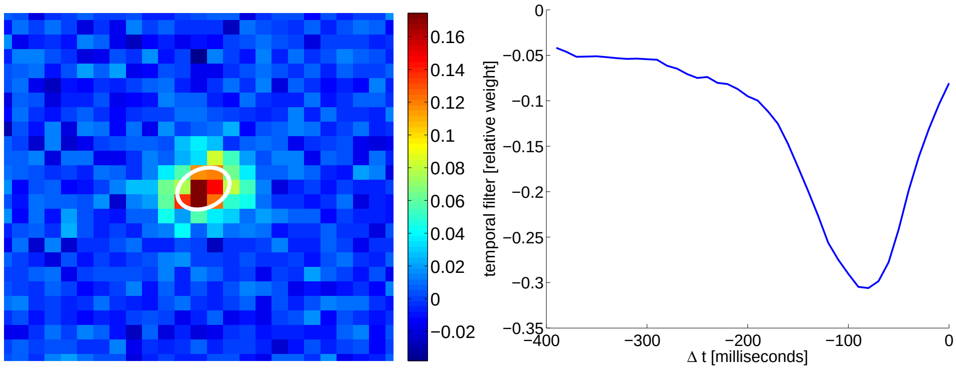


Figure 3.3: Typical voltage trace of a single channel recorded by the MEA setup, showing a burst of spikes from an individual retinal ganglion cell. Since the cell membrane acts as capacitor, the extracellular signal is the derivate of the intracellular voltage (as recorded by the patch clamp technique, for example). For this reason, and due to the 300 Hz high-pass filter, the proportion of positive and negative extracellular voltage amplitudes is similar. The red line defines the spike detection threshold, i.e., whenever the signal crosses this line, a spike is assigned to one of several “clusters”. Note that the signal overshoots the threshold a second time after each spike; these spike artifacts are filtered out by the spikesorting algorithm.

3.4 Receptive Field Calculation

In general terms, the *receptive field* of a sensory neuron is the region in stimulus space which influences that neuron’s response. Specifically for vision, the receptive field of a retinal ganglion cell is an area of the visual field. Visual cues from this area eventually reach the ganglion cell via photoreceptors and interneurons. Since the optical pathway of the eye maps the visual field one to one onto the retina, receptive fields can be specified by the corresponding area on the retina. Note that ganglion cells typically pool input from many photoreceptors; thus, the receptive field size is much larger than the area covered by a single photoreceptor.

We employ a white noise analysis (Chichilnisky 2001) to estimate the receptive field of individual retinal ganglion cells. The general idea can



(a) spatial STA, showing the receptive field **(b)** temporal filter

Figure 3.4: Spatial (a) and temporal (b) part of a ganglion cell's STA (spike-triggered average) when presented with a flickering checkerboard, and Gaussian fit of the receptive field. (White ellipse shows standard deviation.) The size of each pixel (or tile) is $60 \times 60\,\mu\mathrm{m}^2$ on the retina. Color codes for intensity and should not be confused with the fact that the actual stimulus is grayscale. The spatial and temporal contributions to the STA have been extracted mathematically via a SVD (singular value decomposition) of the STA. Time zero of the temporal filter is when the corresponding spike occurs; the curve shows that the cell needs approximately 100 ms to integrate its activation.

be summarized as follows: the cell's preferred stimulus (both in space and time) is the (average) stimulus immediately preceding each spike, denoted STA (spike-triggered average). Specifically, we present a flickering checkerboard to the retina, consisting of a regular chessboard-like tiling of the whole screen. (For the purpose of this section, the tiles can be thought of as pixels on the retina.) Each tile is approximately $60 \times 60\,\mu\mathrm{m}^2$ in size (on the retina) and has its own (grayscale) color, which is determined by generating a random number. Every $\Delta t = 40\,\mathrm{ms}$, we update the colors of all tiles by drawing a new set of independent numbers — hence the flickering. We use a deterministic computer program to generate these (pseudo-)random numbers; thus, we can easily re-generate the exact same sequence later for offline analysis. This is necessary since we calculate the STA after offline spikesorting. Under the assumption that the "memory" of a ganglion cell does not exceed 800 ms, it suffices to include $800\,\mathrm{ms}/\Delta t$ time steps preceding each spike. For a responsive ganglion cell, the STA is a short "movie clip" in which the spatial receptive field lights up. A singular-value decomposition approximates the STA (to first order) by a product of a temporal filter times a stationary spatial part. Figure 3.4 shows an example, including a 2D Gaussian fit (white ellipse) of the receptive field. This Gaussian distribution serves as reference for later analysis.

We observe that the receptive field diameter (obtained from the Gaussian fit standard deviation) ranges from 100 µm to 400 µm.

3.5 Latency Estimation

We estimate an individual cell's latency after presenting the same stimulus for several trials as illustrated in figure 3.5. To deal with situations where a cell skips spikes within a trial, we have to refrain from taking the mean value of all first spikes within each respective trial. Instead, we calculate a latency probability distribution (blue curve in figure 3.5) and take the maximum of this curve. We additionally obtain an estimate of the variance in the data from the FWHM (full width half maximum) of this curve.

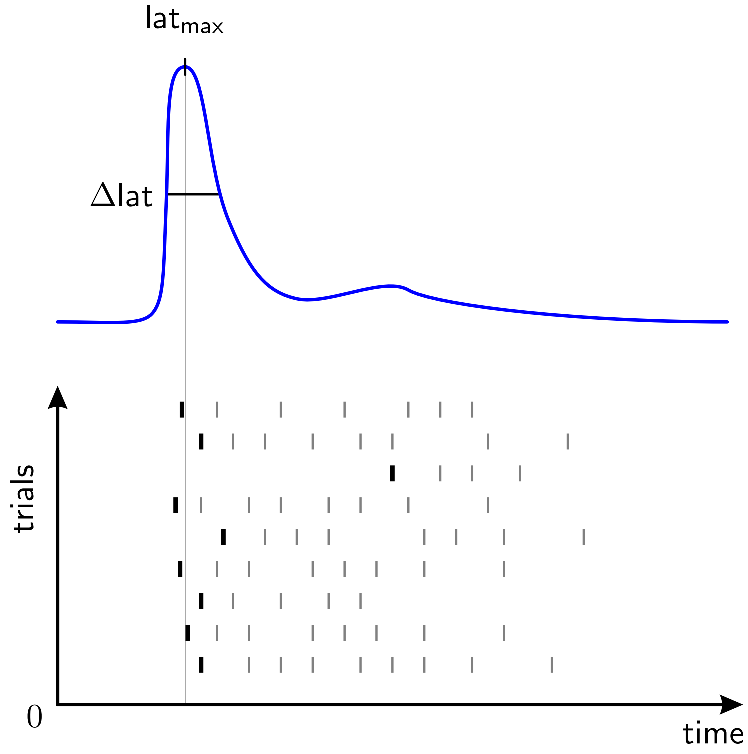


Figure 3.5: Calculation of the most probable latency lat_{max} and corresponding error Δlat from the response of a single cell (schematic illustration). The same stimulus is repeated several times ("trials") for statistical analysis. The blue curve shows the distribution of the first spike (bold) in each trial. The variance Δlat is obtained from the FWHM of this curve. A critical advantage of this approach is the robustness against outliers.

3.6 Pharmacology

The following section lists the various drugs employed in the pharmacology experiments.

To selectively block the ON-pathway, we add 200 µM of 2-amino-4-phosphono-butyrate (APB) to the Ringer's medium. APB is a metabotropic glutamate receptor agonist which blocks light-induced activation of ON bipolar cells (see also Nakajima et al. (1993); Conn and Pin (1997); Yang (2004) and figure 2.3). Although APB acts on multiple metabotropic receptors in the retina, the selective loss of ON responses is thought to derive from its action on the ON-bipolar dendrites.

For the purpose of blocking inhibition, we add the following cocktail to the Ringer's solution (table 3.2). The various concentrations are taken from Ichinose and Lukasiewicz (2005) and Rieke (2001):

concentration	name	function
5 µmol/ℓ	strychnine	glycine receptor blocker
10 µmol/ℓ	bicuculline	competitive antagonist of ionotropic $\mathrm{GABA_A}$ receptors
150 µmol/ℓ	picrotoxin	non-competitive antagonist of $\mathrm{GABA_A}$ receptors

Table 3.2: Amounts of pharmacological substances used for blocking inhibition in the retina

All these receptors are primarily located on the bipolar dendrites or axon terminals, as well as on retinal ganglion cells.

Chapter 4

Neuronal Coding in the Presence of Saccades

This chapter constitutes the first main results part of the thesis. It is concerned with a detailed analysis how individual retinal ganglion cells encode information in the context of (micro-)saccades, which we have reviewed in section 2.2. We observe that the much smaller fixational eye movements (e.g., Greschner et al. (2002)) with amplitudes about the diameter of a single photoreceptor are not sufficient to reliably elicit action potentials of the ganglion cells, at least for our visual stimulation setup. Thus, our study focuses on stimulation parameters corresponding to saccadic eye movements.

Our setting is similar to the archer fish study by Segev et al. (2007), which imitates saccadic eye movements with angular velocity 3 µm/ms on the archer fish retina. For comparison, our study typically uses 1 µm/ms (see section 4.1). In the archer fish study, the fixation time interval is 2 s, which is of the same order of magnitude as in our study.

Since we perform the measurements on the retina isolated from the eye, we have to imitate eye movements by rapid shifts of the presented image (see section 4.1 for details). These shifts are interleaved with stationary periods (lasting about 1 s), which can be interpreted as fixations. Typically, the rapid shifts elicit several action potentials (or spikes) after the shifts, whereas the ganglion cells stop firing when the presented image remains stationary. We investigate how these spike patterns depend on the presented image *before* or *after* the shift, or on the time-dependent activation during the shift. By varying the presented image, the activation strength changes. We observe that this results in different spike *latencies*, i.e., the time interval between beginning of the fixation and the first elicited spike.

In our study, the presented image is a rather simple black-white grating, resembling a zebra crossing. This allows us to calculate the activation (spatial generator signal) based on the cell's receptive field *center*, as will become clear in section 4.3. We compare the ganglion cell's spiking responses to different vertical offsets of the grating. Fitting a sine curve to the resulting latency data (see figure 4.13 below) enables a direct comparison with the spatial generator signal. In first approximation, we are able to estimate the *phase* of this sine curve from the relative receptive field position on the grating (see figure 4.14).

In order to refine our model, we have designed a stimulus modality which employs two parameter sets in alternating order between trials. For example, we reverse the movement direction in every other trial (section 4.5) to estimate the contribution of the image shift direction to the cell's response. The alternating presentation is essential to exclude any potential adaptation effects (if we were using the same parameter set for several trials). For both directions, the visual images before and after the movement shifts are the same, so the observed response differences must be due to the time-dependent activation during the shifts.

The case of different starting or target images (before or after an image shift, respectively) is treated in section 4.6. Here, we focus on the relative contribution of these images to the cell's response. Naturally, one expects that the retina transmits a "snapshot" of the new (target) image after each saccade to higher brain areas. Nevertheless, since ganglion cells are mostly sensitive to *relative* activation changes, the starting image should also have some weight.

4.1 Designing a Stimulus to Imitate Saccades

As mentioned above, our basic idea consists of imitating saccadic eye and head movements by a rapid displacement (or shift) of the presented image. We have found that the movement time interval and amplitude cannot be too small, otherwise they do not elicit spiking responses at all. The study by Werner and Himstedt (1985) supports this observation. The authors investigate the head orientation of fire salamander during prey capture and report relatively slow movements, lasting up to 3 s.

Our detailed movement profile is shown in figure 4.1. We employ the following parameters (unless noted otherwise): shift time interval $\Delta t_{\text{move}} = 100\,\text{ms}$, total stimulus period $p = 2\,\text{s}$, and movement amplitude $a = 100\,\mu\text{m}$. This leaves 900 ms for the stationary time interval. Note that the image moves downward and then upward to the starting position, after which the next trial starts and the stimulus is repeated.

As next step, we have to select an appropriate image for projection onto the retina. For our study, we use a simple black-white grating (similar to a zebra crossing) with stripe width 100 μm. Thus, the spatial period of this pattern equals 200 μm. An important advantage is the constant global illumination of the retina when the image moves, so we can assume that any spiking elicited by the movement is due to local changes. Furthermore, we

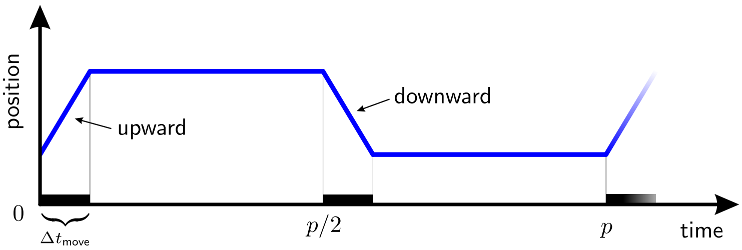


Figure 4.1: Imitation of (micro)saccades by rapid image shifts. Since the retina has to remain stationary on the recording setup, we shift the presented image to mimic eye movements. Black bars indicate movement intervals, and the blue curve the trajectory of the *retina* relative to the image. In this study, the period $p = 2\,\text{s}$, $\Delta t_{\text{move}} = 100\,\text{ms}$ and movement amplitude $a = 100\,\mu\text{m}$ (unless noted otherwise).

will show later (section 4.3) that the activation of a ganglion cell by this grating pattern mostly depends on the *center* of the cell's receptive field relative to the grating, whereas the precise shape of the receptive field is negligible for our analysis.

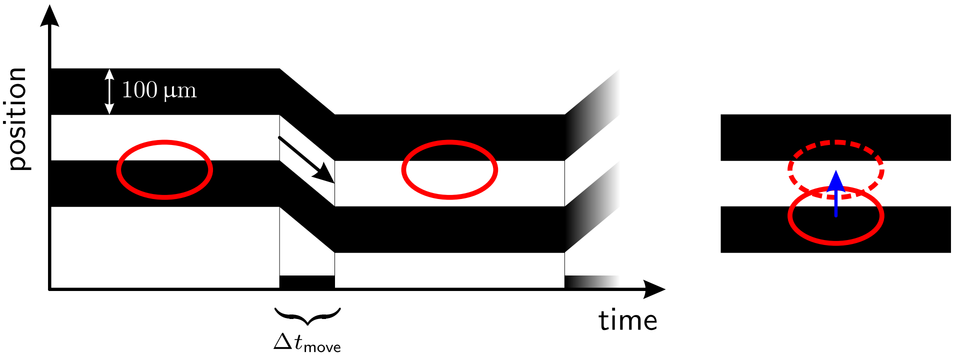


Figure 4.2: Rapid image shifts to imitate saccadic eye movements (see figure 4.1) with a periodic black-white grating as spatial stimulus. The continuous red ellipses schematically indicate the receptive field of a single ganglion cell, which has to remain stationary on the recording setup but moves relative to the grating. The equivalent illustration on the right shows the upward movement shift from the perspective of the grating. In particular, the visual stimulus inside the dashed ellipse is the same as inside the continuous one in the middle. The right side captures the actual biologically situation where the eye moves.

Figure 4.2 shows the rapid shifts together with the black-white grating. The red ellipses illustrate the receptive field of a single ganglion cell, which traverses different parts of the grating due to the shifts. The current activation of the cell can be thought of as the black-white pattern inside the ellipse. In what follows, we investigate the cell's spiking response elicited by this time-dependent activation.

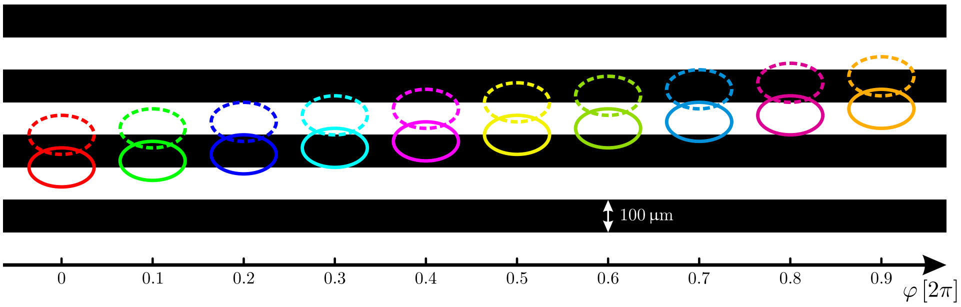


Figure 4.3: Vertical position on the grating defines the information which can be gained about the stimulus. We offset the grating 10 times such that the receptive field of an individual ganglion cell (colored ellipses) traverses different parts of the black-white grating. The grating rapidly shifts between the continuous and dashed version of each same-colored ellipse pair as shown in figure 4.2. After several trials, the next global offset to the neighboring ellipse pair occurs. Comparing the cell's responses for different ellipse positions (i.e., phase shifts φ) could elucidate how the cell encodes information.

Concretely, we want to understand how the cell's response depends on the vertical position on the grating, in order to investigate how the cell could transmit spatial information. For that purpose, we repeat the stimulus in figure 4.2 for several times and then offset the whole grating by 20 µm, such that the ganglion cell's receptive field starts from a slightly different position. After 10 repeats of these offsets, the cell reaches its original position due to the periodicity of the grating. Figure 4.3 illustrates this scheme by color-coding the receptive field position of a single ganglion cell for the various starting positions. The labeling as "phase" φ will become clearer in section 4.3, where we show that the ganglion cell's activation by the grating (i.e., the local brightness within the receptive field) closely matches a sine curve, which reaches 1 in the middle of each black stripe and correspondingly -1 in the middle of the white stripes.

As outlook, we can ask how to tackle the decoding problem, i.e., infer the ganglion cell's receptive field position from the measured spiking responses.

4.2 Informative Spike Response Features

In the previous section, we have provided the details of the visual stimulation. Now, the apparent question arises if and how retinal ganglion cells respond to this stimulus. Moreover, in the interest of decoding, we have to find out which features of the spiking response contain information about the vertical grating position.

Before investigating the experimental data, we briefly review several conceivable coding schemes (Dayan and Abbott 2005), namely

- the spike count, i.e., the number of action potentials elicited after stimulus onset, within the time window of the current trial (the neurons should not fire spontaneously, otherwise the number of spikes cannot clearly attributed to the stimulus alone).
- the time-dependent firing rate (spikes per time) at time t after stimulus onset, measured by counting the (trial-averaged) number of spikes within a short time interval t, $t + \Delta t$ and dividing by Δt. Defined this way, the firing rate is equivalent to the PSTH (peristimulus time histogram), which is just the time histogram of the spikes during stimulus presentation, as the name implies. A smaller "time bin" Δt means higher temporal resolution, but incurs lower resolution of firing rates.
- inter-spike correlations; for example, the inter-spike interval (ISI), the time difference between two succeeding spikes of a neuron, could carry information. This information would be lost if regarding spikes as independent of each other. Note that the refractory period sets a minimum limit on the time interval between succeeding spikes.
- temporal codes, i.e., the precise timing of individual spikes (up to millisecond precision); a prominent example is the latency, that is, the time interval between stimulus onset and the first spike elicited by the stimulus.
- neuronal "population codes", e.g., the synchronous firing of two neurons, or rhythmic oscillations of a neuronal population. A prominent example are place cells in the hippocampus: the spike timings of an individual place cell relative to the phase of the population theta rhythm ($7 - 12\,\mathrm{Hz}$) provide information about the current location of the animal. In particular, the conveyed information cannot be extracted from individual neurons alone but requires the population.

Historically it has been believed that the spike rate carries most or all information conveyed by a neuron. This view is supported by recordings from stretch receptor neurons almost a century ago. In the study, the neurons increase their firing rate as the load on the muscle increases (Adrian 1926; Adrian and Zotterman 1926). Nevertheless, neuroscientists also have proposed alternative coding strategies, as enumerated above. It seems likely that the particular strategy employed depends on the application and type of neuron. For our study, we will particularly investigate the role of the spike *latency* for stimulus encoding.

Measured responses to the shifted grating stimulus Figure 4.4 shows (an excerpt of) the spiking activity of two individual ganglion cells responding to the phase-shifted oscillating grating stimulus described in the previous

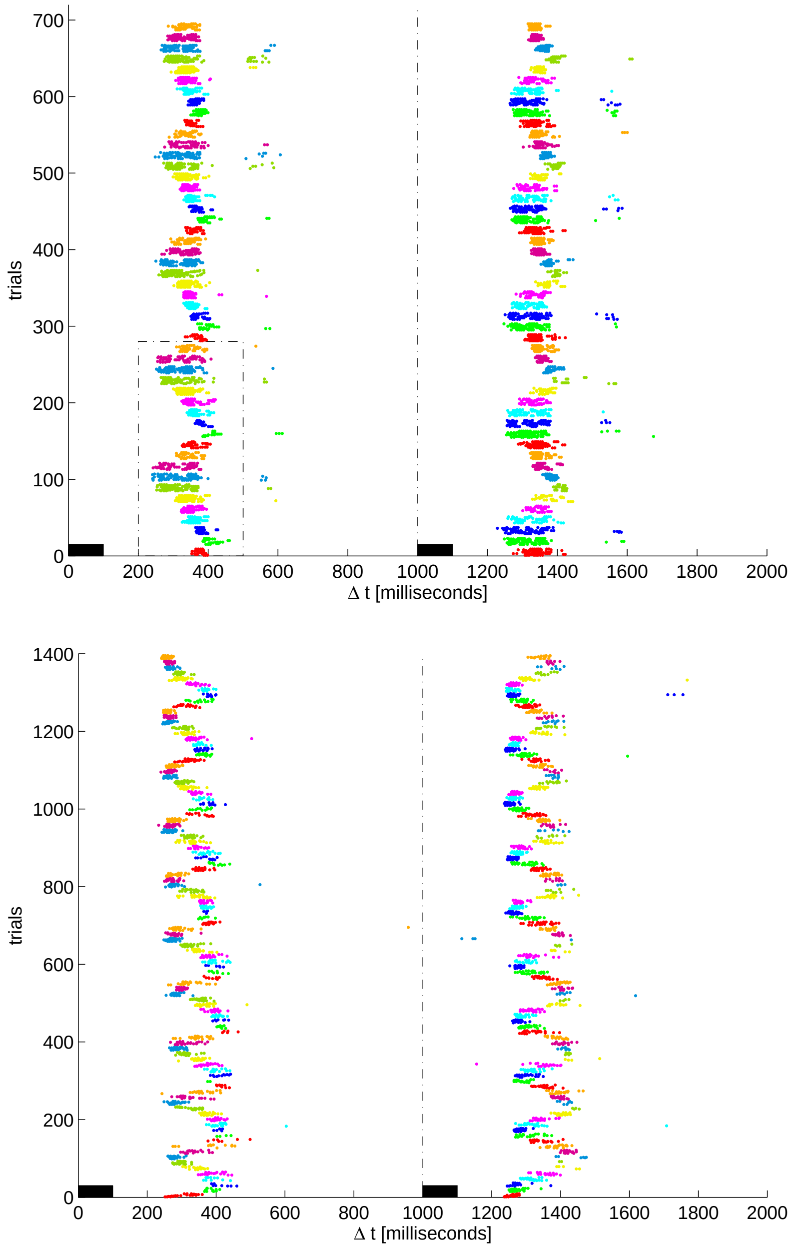


Figure 4.4: Spike raster plot of two individual retinal ganglion cells presented with the same phase-shifted oscillating grating stimulus. Each dot is a single spike, and color encodes the phase offset φ as shown in figure 4.3. The time axis is the same as illustrated in figure 4.2 with $p = 2\,\mathrm{s}$ and $\Delta t_{\mathrm{move}} = 100\,\mathrm{ms}$. Note that the cells' responses set in more than 200 ms after the image shifts (black bars at the bottom). While the general structure of the *latency* response is similar, the second cell exhibits a more transient response pattern than the first. The two cells are taken from different retinas. The dashed rectangle in the upper plot marks the zoom-in region shown in figure 4.5 below.

section (figure 4.2). The most remarkable feature of the spike patterns in figure 4.4 are the different response latencies for different phases (illustrated by the 10 different colors). That is, the first spike elicited within the first and second half of each trial (corresponding to the upward and downward movement intervals in figure 4.1) has a sinusoidal shape, when plotted against the phase offset shown in figure 4.3. Since it is plausible that latency is a fast, reliable information source, we will primarily focus on this coding strategy in the further analysis. For statistical verification, we repeat several stimulus cycles, as can be seen from the spike raster plots. By "cycle" we understand the presentation of the 10 receptive field offsets shown in figure 4.3, that is, each sequence of 10 colors in figure 4.4.

Additionally, note that the first cell in figure 4.4 abruptly stops firing about 400 ms after stimulus onset, almost independently of the phase. Due to the differences in spiking onsets, we expect the spike *count* to be noticeably different for the various phases. This observation only applies to the first cell, whereas the second cell responds transiently to the stimulus. Nevertheless, a quantitative analysis will reveal that the second cell also shows a systematic spike count variation.

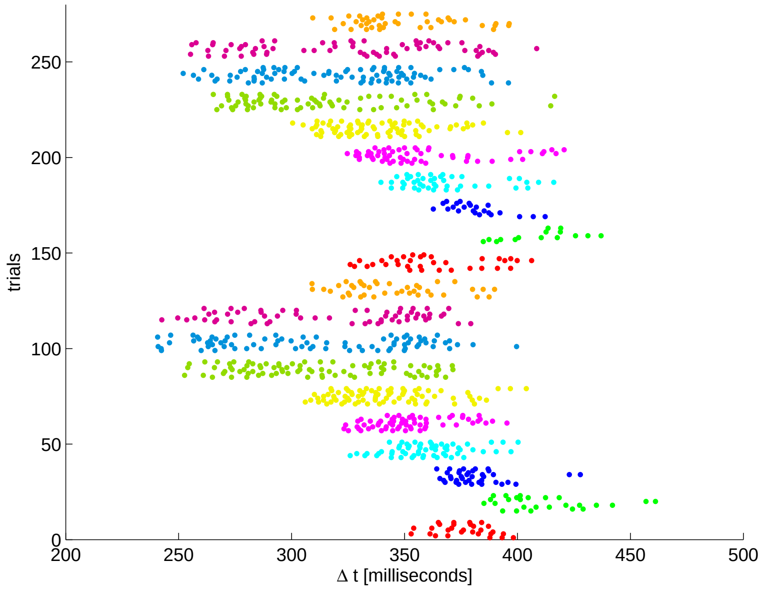
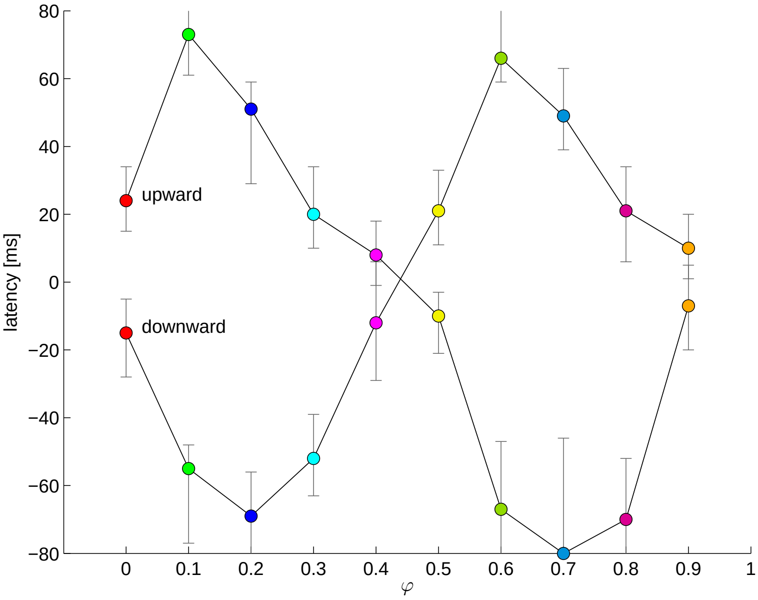
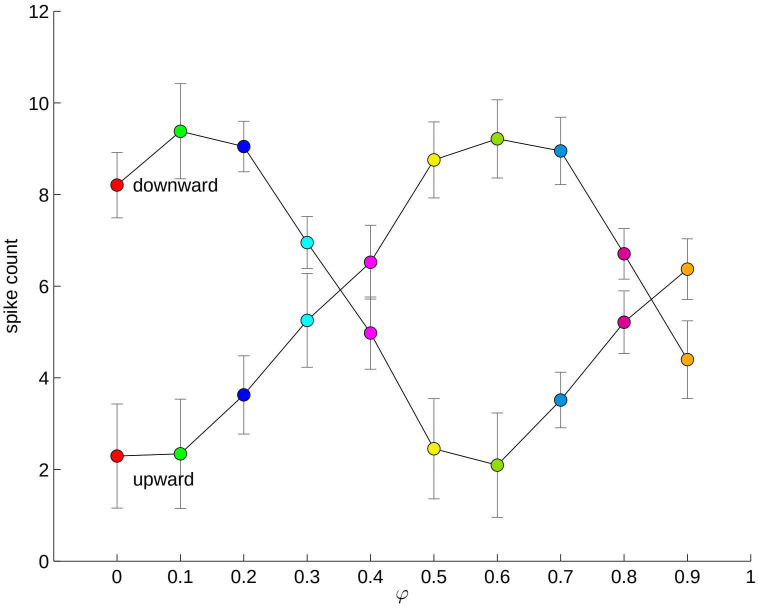


Figure 4.5: Zoom-in of the dashed rectangle in figure 4.4 to resolve the precise spiking patterns. Each dot is a single spike, and color encodes the receptive field position offsets or "phases" φ as shown in figure 4.3. There is quite a large variability within same-colored trials ($\sim$ 25 ms), which is nevertheless about 6 times smaller than the latency difference between the bright green and navy-blue color ($\sim$ 150 ms). Intriguingly, this number is even larger than the image shift interval of 100 ms. This could be interpreted as a neuronal "amplification" in the time domain.

Figure 4.6a shows a quantitative version of the spike latencies and spike

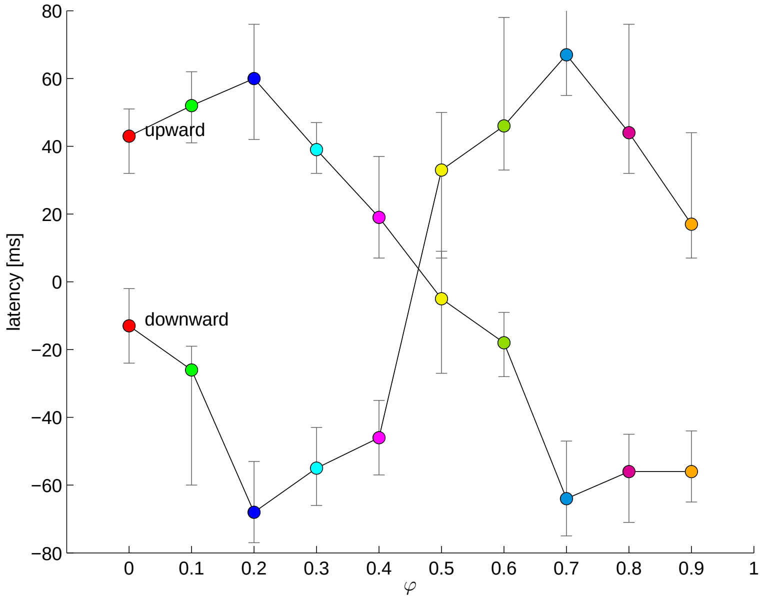


(a) Latency

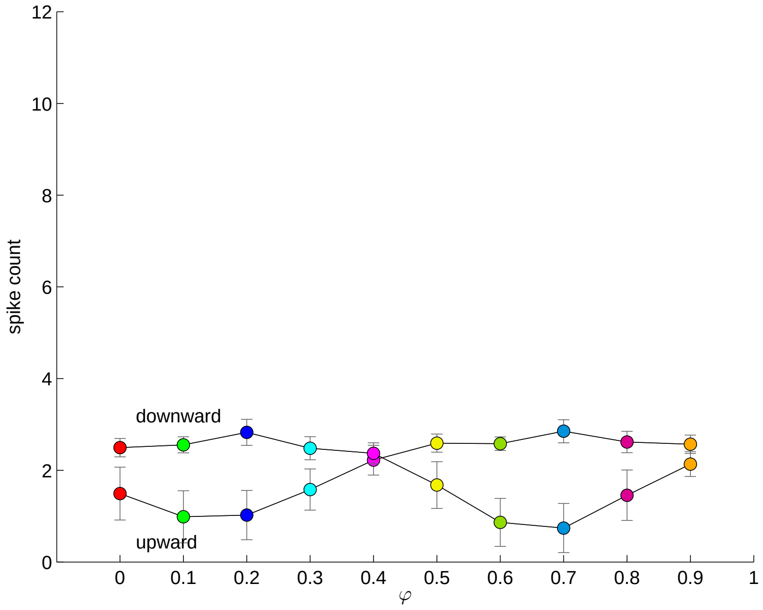


(b) Spike count

Figure 4.6: Latency and spike count of the first cell in figure 4.4, forming a sinusoidal shape. The two curves in each subplot correspond to the first and second half of the stimulus in figure 4.3. Latency values and error bars are determined as illustrated in figure 3.5.



(a) Latency



(b) Spike count

Figure 4.7: Latency and spike count of the second cell in figure 4.4, analogous to figure 4.6. The spike count tuning is less pronounced as compared to the first cell, but nevertheless clearly visible.

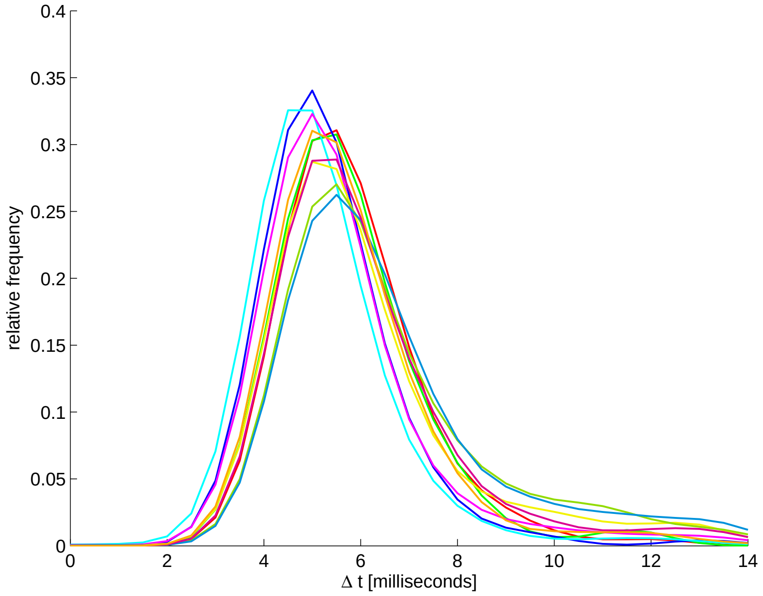
counts of the first cell. To correct for noise and slow time drifts of the cell's responses during the recording (which takes about 1 h for a stimulus modality), we subtract the latencies within a presentation cycle from the respective mean latency values. Specifically, we first select the fastest latency response within the stimulus trials of each phase offset for the current presentation cycle. Then, we compute the mean value of these 10 latencies (one for each phase offset) and finally subtract the mean from these latencies. We proceed analogously for the second-fastest latencies etc. Next, we calculate histograms of the latency data (separately for each phase offset) and employ the scheme in figure 3.5 to obtain the latency values and error bars shown in figure 4.6a. Hence, the data points are *relative* latencies.

Note the symmetric reversal of the curves in figure 4.6a. For example, the cell responds slowest to the upward movement during the spatial phase $\varphi = 0.1$ (denoted as green data point), but (almost) fastest for the downward movement during the same phase. We observe that the movement from a white grating stripe to a black stripe leads to a short latency (presumably since the ganglion cell is OFF-type), and conversely, the movement from black to white to a long latency.

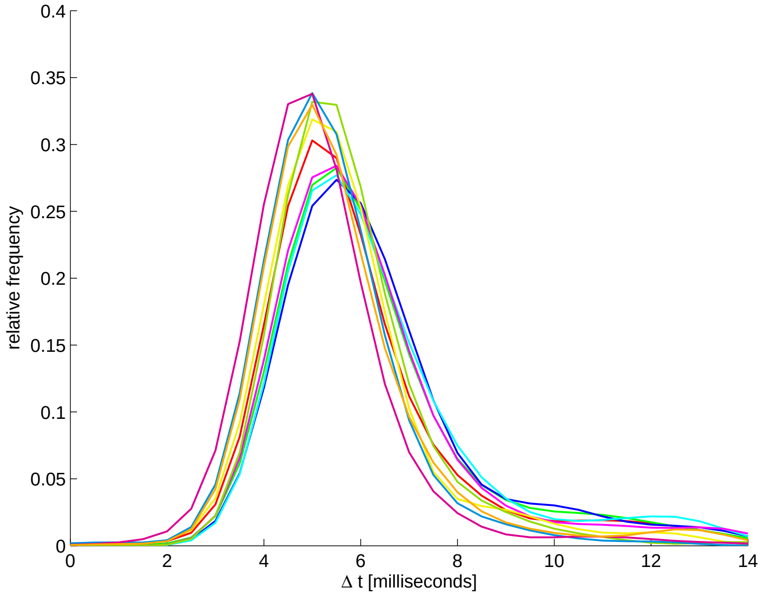
Figure 4.6b shows the average spike count of the first cell. As expected, the spike count is highest when the cell responds earliest. The latency and spike count curves are quite similar, so in this particular case, we conclude that they serve equally well for extracting information about the phase φ (see also section 4.4 below). However, note that the latency information is – by definition – immediately available when the first spike appears, whereas a spike count observer has to wait until the cell subsides spiking.

Figure 4.7 contains the latency and spike count of the second cell from figure 4.4. The latency curves look remarkably similar to the first cell and are again symmetrically reversed. (Since the two cells stem from different recordings, we can exclude any intrinsic mutual influence.) However, the variation in spike counts is much smaller for the second cell, which fires between 1 and 3 spikes in each trial. Still, there's a clearly visible tuning of spike counts with respect to the phase φ; this tuning was not obvious from the raster plot in figure 4.4. Since spike counts are integers, we speculate that – heuristically – more information about the phase φ is contained in the latency than the spike count for this particular cell. Namely, higher brain areas receive either 1, 2 or 3 for the spike count on a single trial basis, whereas the latency can be more fine-grained.

An alternative coding scheme mentioned in the above enumeration is the inter-spike interval (ISI), that is, the time interval between two succeeding spikes of a single cell. In principle, this coding scheme allows for rapid information transmission since only two spikes are required – in contrast to the spike count, for example, where typically many more spikes have to be submitted. In a related study, Berry and Meister (1998) show that the timing of individual spikes is more precise than naively expected from the



(a) upward shift



(b) downward shift

Figure 4.8: Inter-spike interval (ISI) histogram plots of the first cell in figure 4.4, after application of a Gaussian filter for smoothing (bin size is 0.5 ms). As above, color encodes the receptive field position offsets or “phases” φ as shown in figure 4.3. Note that the curves are slightly shifted with respect to each other, which could provide a small contribution to information transmission.

instantaneous stimulus-based firing probability (without taking the timing of preceding spikes into account). Figure 4.8 show ISI time histograms of the first cell from figure 4.4, for both the upward and downward movement shifts. Observe that all spike intervals are greater than 2 ms, as expected from the refractory period. Although the curves corresponding to different phases look similar, there is a small offset between the peaks of the curves, i.e., the most likely ISI. For example, the cyan curve in 4.8a ($\varphi = 2\pi \cdot 0.3$) reaches its maximum at 4.5 ms, whereas the dark blue-green ($\varphi = 2\pi \cdot 0.7$) curve's maximum appears 1 ms later. Note that the latter also has the shortest latency (see figure 4.6a). Thus, the ISI could contain a small amount of information about the stimulus. Nevertheless, the similarity of the curves suggests that this is less information than contained in the spike latency or spike count, so the further analysis will focus on the latter two.

Quantification of cell classes Obtaining latency tuning curves as demonstrated above is not possible for all cells, for example, if a cell fires spontaneous spikes which cannot be attributed to the presented stimulus, or if the rapid shifts do not elicit spikes for all phases.

We have performed a quantitative screening of 126 cells (which we have selected based on spikesorting and receptive field quality), revealing that

> 18% exhibit this kind of latency response, i.e., these cells fire spikes for each stimulus phase with a clearly distinguishable response onset, and without interfering spontaneous activity;
>
> 62% exhibited an irregular or too noisy response pattern, i.e., the variance of the first spike during identical stimulus repeats was higher than the difference for the various phases;
>
> 14% did not respond at all, or showed uncorrelated spontaneous activity;
>
> 6% responded to a subset of the stimulus phases only (thus, there is no latency data available for all phases).

Concerning the last scenario, note that the spike count could serve as a robust alternative for information transmission.

Summary Concluding, we have found hints of two potential coding schemes from the above list (latency and spike count). Additionally, the inter-spike interval (ISI) may also exhibit a subtle contribution, even though the ISI is largely independent of the currently employed phase φ. The main difference between our study and the one by Gollisch and Meister (2008) stems from our explicit imitation of saccadic eye movements, whereas their study flashes static images onto the retina.

4.3 Activation of Receptive Fields: Mathematical Analysis

In the previous two sections we have explained how we offset the presented image (figure 4.3) to study the effect on the neuronal responses, and found that the latency serves a major role in encoding the stimulus. To cast these observations into mathematical formulas, we first have to quantify how a ganglion cell is activated by the black-white grating which we project onto the retina. "Activation" (or "spatial generator signal") refers to the excitation of retinal ganglion cells, i.e., the incoming light stimulus translated via photoreceptors, horizontal cells, bipolar cells and amacrine cells into synaptic excitation (or inhibition). For our purposes, we substitute the intricate retinal network by a much simpler receptive field model (see figure 3.4), where we calculate the activation of a retinal ganglion cell from the light intensity within the cell's receptive field. We describe the receptive field by a Gaussian function in two dimensions; hence the activation is the spatial integral of the stimulus light intensity weighted by this function. The receptive field's size and (elliptical) shape corresponds to the standard deviation of the Gaussian curve.

In the present section, we will derive the following slightly counter-intuitive but useful result: the activation is (approximately) independent of the detailed size and shape of the receptive field (under the assumption that the size is of the same magnitude as the grating width). Instead, only the *center position* enters the equations, yielding a sinusoidal tuning curve of the activation. This is important because it greatly simplifies the further analysis, allowing us to compare the phase of the activation tuning curve to the phase of the latency curves from the previous section.

In mathematical terms, the receptive field profile is a 2D Gaussian normal distribution denoted $\mathcal{N}(\mu, S)$, with center position $\mu \in \mathbb{R}^2$ and positive definite variance matrix S. The (stationary) stimulus light intensity $f(x)$ depends on the spatial variable $x \in \mathbb{R}^2$. As mentioned above, the spatial activation (or "generator signal") equals

$$\begin{aligned} G(\mu, S) &= \iint_{\mathbb{R}^2} \frac{1}{2\pi\,|S|^{1/2}} \mathrm{e}^{-\frac{1}{2}(x-\mu)\cdot S^{-1}(x-\mu)} f(x)\,\mathrm{d}^2x \\ &= \iint_{\mathbb{R}^2} \frac{1}{2\pi} \mathrm{e}^{-\frac{1}{2}|y|^2} f\left(S^{1/2}y + \mu\right)\,\mathrm{d}^2y. \end{aligned} \tag{4.1}$$

First, we investigate the special case where f describes a half-plane $v \cdot x - d \leq 0$, with parameters $v \in \mathbb{R}^2 \setminus \{0\}$ and $d \in \mathbb{R}$. This translates to

$$v \cdot \left(S^{1/2}y + \mu\right) - d \leq 0 \quad \Leftrightarrow \quad \underbrace{\left(S^{1/2}v\right)}_{v'} \cdot y - \underbrace{(d - v \cdot \mu)}_{d'} \leq 0.$$

We weight the bright half of the plane by 1 and the dark half by -1 to establish the mean intensity 0. Plugged into (4.1) yields (after integrating over one dimension)

$$G_{\mathrm{HP}}(\mu, S) = 2\int_{-\infty}^{\frac{d'}{|v'|}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}y^2}\,\mathrm{d}y - 1 = \mathrm{erf}\left(\frac{d - v\cdot\mu}{\sqrt{2}\left|S^{1/2}v\right|}\right). \tag{4.2}$$

From this formula, we may define the effective receptive field size as $\sigma := \left|S^{1/2}v\right|$ and the effective receptive field center position as $z := v \cdot \mu$.

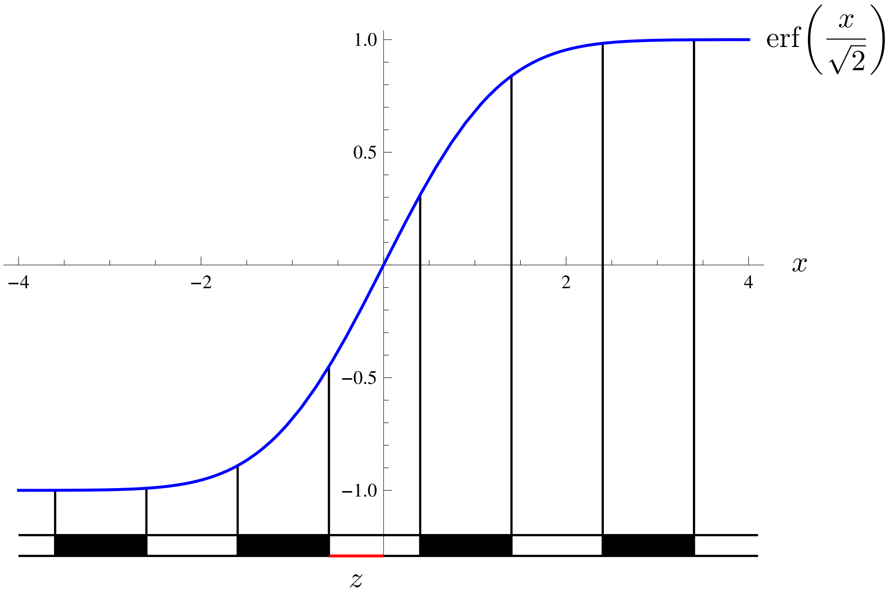


Figure 4.9: The generator signal of a spatial grating (black-white bars) projected onto a Gaussian receptive field profile is an infinite sum of erf error functions, evaluated at the grating boundaries (see equation (4.3)).

Equation (4.2) allows us to calculate the generator signal by the black-white grating stimulus, as illustrated in figure 4.9, since each (black or white) stripe is the difference between two half-planes. For simplicity and without loss of generality we set the grating width to 1 and assume that the white stripes extend from $d = 2n$ to $d = 2n + 1$ ($n \in \mathbb{Z}$). Conversely for the black stripes, the borders are $d = 2n - 1$ and $2n$. Summation over all stripes leads to the generator signal

$$
\begin{aligned}
G_{\text{grating}}(z,\sigma) = \\
c \sum_{n=-\infty}^{\infty} \left[\operatorname{erf}\!\left(\frac{2n+1-z}{\sqrt{2}\,\sigma}\right) - 2\operatorname{erf}\!\left(\frac{2n-z}{\sqrt{2}\,\sigma}\right) + \operatorname{erf}\!\left(\frac{2n-1-z}{\sqrt{2}\,\sigma}\right)\right] \\
= -2\,c \sum_{n=-\infty}^{\infty} (-1)^n \operatorname{erf}\!\left(\frac{n-z}{\sqrt{2}\,\sigma}\right) \qquad (4.3)
\end{aligned}
$$

where c is a normalization factor such that $\max_z |G_{\text{grating}}(z,\sigma)| = 1$. To further analyze this infinite sum, we first write

$$
-2\operatorname{erf}\!\left(\frac{n-z}{\sqrt{2}\,\sigma}\right) = \int \frac{1}{\sqrt{2\pi}} \frac{4}{\sigma}\, \mathrm{e}^{-\frac{(n-z)^2}{2\sigma^2}}\, \mathrm{d}z
$$

and interchange the order of integration and summation (which is justified due to the exponential decay). Together with $(-1)^n = \mathrm{e}^{\mathbb{i}\pi n}$, we arrive at

$$
G_{\text{grating}}(z,\sigma) = c \int_0^z \frac{1}{\sqrt{2\pi}} \frac{4}{\sigma}\, \mathrm{e}^{-\frac{z'^2}{2\sigma^2}}\, \vartheta\Big(\pi/2 - \mathbb{i}z'/(2\sigma^2), \mathrm{e}^{-\frac{1}{2\sigma^2}}\Big)\, \mathrm{d}z' \qquad (4.4)
$$

with the Jacobi theta function defined as

$$
\vartheta(u,q) := \sum_{n=-\infty}^{\infty} q^{n^2} \mathrm{e}^{2\mathbb{i}nu} \quad \text{for all} \quad u, q \in \mathbb{C} \quad \text{with} \quad |q| < 1.
$$

Applying the Jacobi identity

$$
\sqrt{-\mathbb{i}\tau}\; \mathrm{e}^{\frac{\mathbb{i}u^2}{\pi\tau}}\, \vartheta\big(u, \mathrm{e}^{\pi\mathbb{i}\tau}\big) = \vartheta\Big(u/\tau, \mathrm{e}^{-\pi\mathbb{i}/\tau}\Big)
$$

to the integrand in (4.4) leads to

$$
4\,\mathrm{e}^{\mathbb{i}\pi z' - \frac{1}{2}(\pi\sigma)^2}\, \vartheta\Big(-\pi z' - \mathbb{i}(\pi\sigma)^2, \mathrm{e}^{-2(\pi\sigma)^2}\Big) = 4 \sum_{\substack{n=-\infty \\ n \text{ odd}}}^{\infty} \mathrm{e}^{-\frac{1}{2}(\pi n\sigma)^2} \mathrm{e}^{\pi\mathbb{i}nz'}.
$$

We can now again interchange the order of summation and integration and explicitly evaluate the integral over z', giving

$$
G_{\text{grating}}(z,\sigma) = 4\,c \sum_{\substack{n=-\infty \\ n \text{ odd}}}^{\infty} \frac{\mathrm{e}^{-\frac{1}{2}(\pi n\sigma)^2}}{\mathbb{i}\pi n}\, \mathrm{e}^{\mathbb{i}\pi nz} = 8\,c \sum_{\substack{n=1 \\ n \text{ odd}}}^{\infty} \frac{\mathrm{e}^{-\frac{1}{2}(\pi n\sigma)^2}}{\pi n} \sin(\pi nz). \qquad (4.5)
$$

That is, we have obtained the Fourier decomposition of $G_{\text{grating}}(z,\sigma)$. The normalization factor $c \equiv c(\sigma)$ results from plugging $z = 1/2$ into the above expression:

$$
c \equiv c(\sigma) = 1 \Big/ 8 \sum_{\substack{n=1 \\ n \text{ odd}}}^{\infty} (-1)^{(n-1)/2}\, \frac{\mathrm{e}^{-\frac{1}{2}(\pi n\sigma)^2}}{\pi n}.
$$

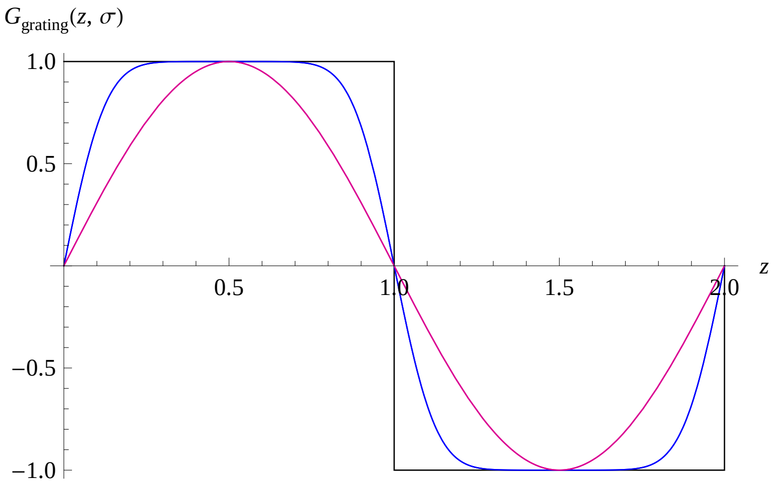


Figure 4.10: Pronounced deviation of the exact grating generator signal (see equations (4.3) and (4.5)) from the sine approximation (magenta) for small receptive field size σ. This example illustrates the transition from a point-like receptive field ($\sigma = 0$, black curve) via a small receptive field ($\sigma = 1/10$, blue) to a size similar to the grating width ($\sigma = 1$, magenta), for which the sine approximation is valid. In the point-size limit, the generator signal is a rectangular-shaped function, which is 1 or -1 depending on whether the receptive field point lies inside a white or black stripe.

Figure 4.10 plots $G_{\mathrm{grating}}(z, \sigma)$ for various values of σ and illustrates the transition from $\sigma = 0$ (i.e., a point-like receptive field) to $\sigma = 1$, in which case the generator signal seems indistinguishable from a sine curve. In other words, we obtain

$$G_{\mathrm{grating}}(z, \sigma) \simeq \sin(\pi z) \quad \text{for} \quad \sigma \geq 1. \tag{4.6}$$

Equation (4.5) rigorously proves this observation (since the Fourier coefficients decay exponentially), and thus verifies our above claim that $G_{\mathrm{grating}}(z, \sigma)$ is (almost) independent of the effective receptive field size σ, as long as σ is larger than the grating width. Note that the quotient between two subsequent Fourier coefficients equals

$$\frac{n}{n+2}\, \mathrm{e}^{-2(n+1)(\pi\sigma)^2} \leq \frac{1}{3}\, \mathrm{e}^{-(2\pi\sigma)^2} \ll 1 \quad \text{for} \quad \sigma \geq 1.$$

Thus, the first sine function $\sin(\pi z)$ almost exclusively contributes to the sum.

For a quantitative error analysis of equation (4.6), we employ the L^2 norm and define

$$s_{\mathrm{err}}(\sigma) := \int_0^2 \left| G_{\mathrm{grating}}(z, \sigma) - \sin(\pi z) \right|^2 \, \mathrm{d}z. \tag{4.7}$$

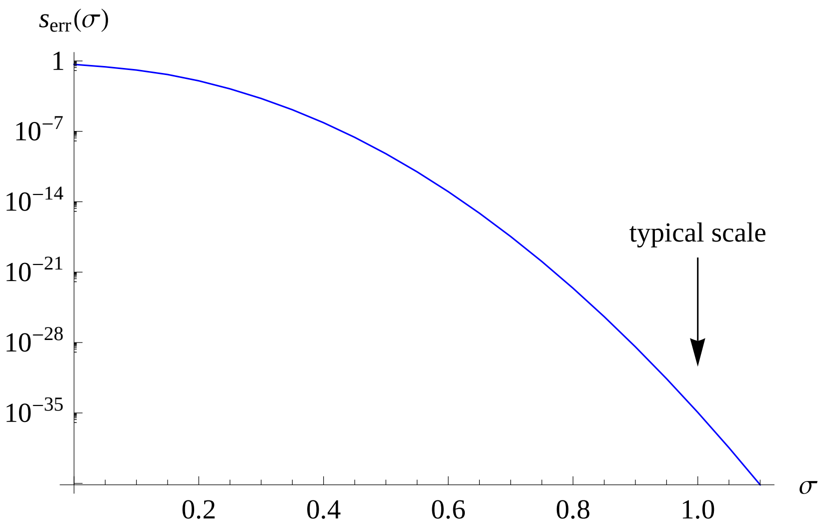


Figure 4.11: L^2 approximation error of the grating generator signal by a sine function (see equation (4.7)), dependent on the effective receptive field size σ. Note the exponential convergence as the receptive field size reaches the grating width. The arrow indicates the typical scale in experiments, for which the approximation error is clearly negligible.

From (4.5) and the orthogonality properties of the sine functions $\sin(\pi n z)$, we obtain

$$s_{\mathrm{err}}(\sigma) = \sum_{\substack{n=1 \\ n \text{ odd}}}^{\infty} \left(8\, c(\sigma)\, \frac{\mathrm{e}^{-\frac{1}{2}(\pi n \sigma)^2}}{\pi n} - \delta_{n,1} \right)^2 .$$

The logarithmic plot of $s_{\mathrm{err}}(\sigma)$ in figure 4.11 nicely illustrates the exponential convergence. As a remark, we can analytically calculate $s_{\mathrm{err}}(0)$ since $G_{\mathrm{grating}}(z, 0)$ is a rectangular function (see figure 4.10), and obtain $s_{\mathrm{err}}(0) = -8/\pi + 3$.

To sum this section up, we have derived an important simplification from a modeling perspective. Namely – for our stimulus parameters – the spatial generator signal is (to very good approximation) the sine of the receptive field center position, without influence from the precise elliptical shape or size of the receptive field (as long as it is not much smaller than the grating width). Thus, regarding the spatial generator signal, we will omit the reference to the receptive field size σ in the following sections.

The sine approximation lends itself to the notation $\varphi \equiv \pi z$, where we interpret φ as the sine curve phase offset corresponding to the receptive field position z. This identification finally justifies the labeling in figure 4.3.

4.4 Latency and Spike Count Response versus Receptive Field Position

In the present section, we test the following hypothesis: the brightness change between the receptive field positions *before* and *after* the image shift mostly determines the cell's latency and spike count response, at least for shift amplitudes equal to the grating stripe width. In more detail, we hypothesize that an OFF ganglion cell will respond fastest (i.e., with shortest latency) for maximally decreasing brightness inside the receptive field, and will also exhibit the largest number of spikes. Correspondingly, it will respond slowest (largest latency) for maximally increasing brightness, with the least number of spikes. The reverse model holds for ON ganglion cells, showing the shortest latency for maximally increasing brightness. We consider the case where the shift amplitude precisely equals the grating width of 100 μm. Note that this simple model does not take the temporal dynamic during the shift into account, which we will analyze in the later sections.

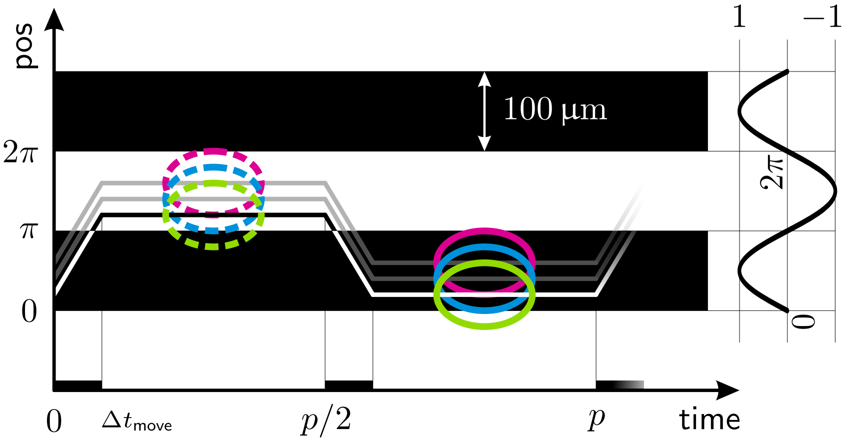


Figure 4.12: Stimulus scheme to imitate (micro-)saccades by rapid shifts (black-white trajectory), with amplitude equal to the grating stripe width. The green ellipse represents the receptive field of an individual retinal ganglion cell. After 10 repeated trials, the whole grating relocates as indicated by the blue ellipse, and so forth (compare with figure 4.3). The spatial generator signals of the dashed and respective continuous ellipses are sign-inverted due to the particular shift amplitude.

Stimulation method To test our hypothesis, we repeat 10 saccadic image shifts and then relocate the whole visual image to a slightly different position, such that any ganglion cell sees a different brightness change at the offset position. Figure 4.12 includes an illustration of 3 image offsets (colored ellipses) applied after 10 trials of the upward and downward saccadic shifts.

Expected outcome Due to the particular shift amplitude, a cell which starts from the middle of a white stripe will reach the middle of a black stripe and vice versa. Similarly, starting from the border between a black and white grating stripe means that the cell will also target such a border. In mathematical terms,

$$\varphi_{\text{target}} = \varphi_{\text{start}} + \pi \quad (\text{mod } 2\pi). \tag{4.8}$$

In equations (4.5) and (4.6) from the last section we have shown that the spatial generator signal is basically the sine of the receptive field position relative to the grating. Making use of these equations and applying the relation (4.8) to the generator signal difference (i.e., brightness change), we obtain

$$\Delta G\left(\varphi_{\text{target}}\right) := G_{\text{grating}}\left(\varphi_{\text{target}}\right) - G_{\text{grating}}\left(\varphi_{\text{start}}\right) = 2\, G_{\text{grating}}\left(\varphi_{\text{target}}\right). \tag{4.9}$$

In other words, the generator signal *difference* (dependent on the target position) is the same as the generator signal at that position, up to a constant prefactor. Thus, the target position precisely indicates the brightness change. In particular, the maximal brightness decrease inside the receptive field occurs when the cell reaches the middle of a black grating stripe. In this case, we expect the shortest latency response for an OFF ganglion cell. However note that equation (4.9) holds only true for shift amplitudes equal to the grating stripe width.

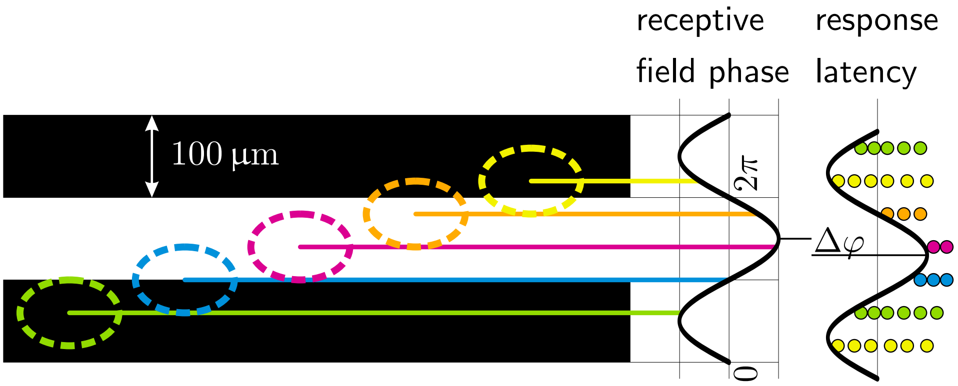


Figure 4.13: Schematic illustration of the phase difference $\Delta\varphi$ between a cell's vertical receptive field position on the grating (colored ellipses analogous to figure 4.3) and the corresponding latency response (exemplified in figure 4.4). The receptive fields refer to the position *after* the rapid image shift. Assuming that the cell is an OFF ganglion cell, it should tend to respond earlier (i.e., with shorter latency) when it reaches a black stripe (green and yellow ellipses) since the overall brightness has decreased, and correspondingly later when targeting a white stripe (magenta ellipse) since the brightness has increased.

Via the sinusoidal latency response tuning curve (see figures 4.6a and 4.7a), we can now define the *phase difference* $\Delta\varphi$ between the receptive field position and the corresponding latency response, as illustrated in figure 4.13. Thus, according to our hypothesis, this phase difference $\Delta\varphi$ should be zero.

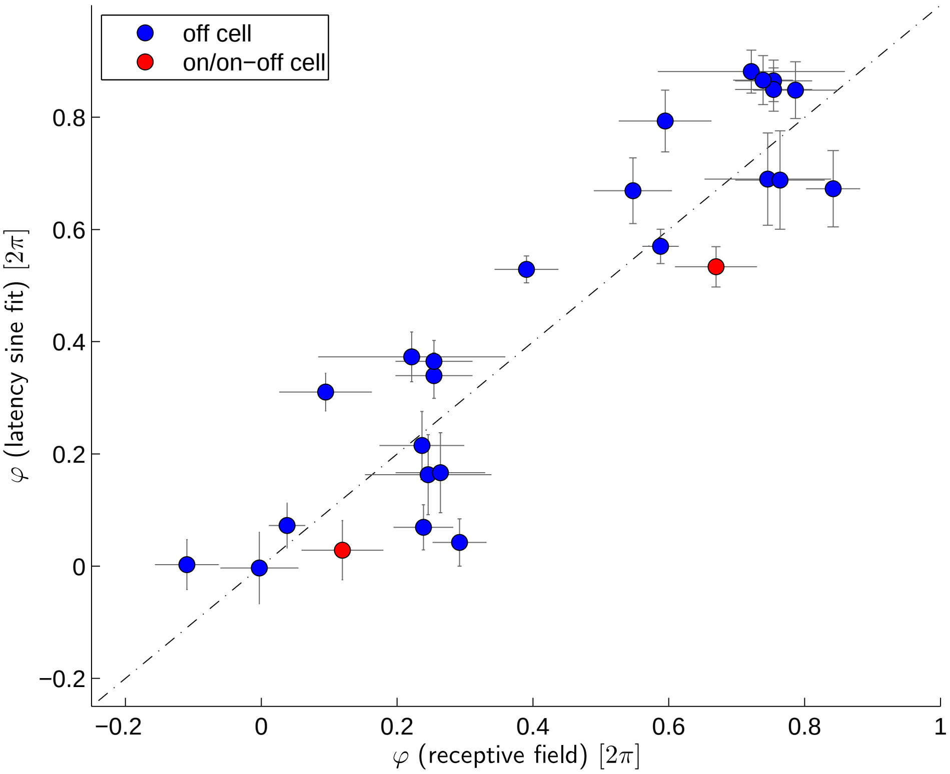


Figure 4.14: Predicting the experimental latency response to the stimulus in figure 4.12 from the phase of the receptive field position. The identity line corresponds to $\Delta\varphi = 0$ in figure 4.13. For ON and ON-OFF cells, we have incremented the receptive field phase by π (mod 2π) to account for the reversed black-white preference (see also figure 4.16).

Experimental results Figures 4.14 and 4.15 contain the experimental data to test our hypothesis, namely, the receptive field position (x-axis) predicting the phase offset of the latency/spike count tuning curve (y-axis), for several retinal ganglion cells. Precise accordance with our hypothesis corresponds to the dashed identity line. Note that every cell has two tuning curves (upward and downward shift), yielding two data points per cell. For ON (or ON-OFF) cells, the receptive field phase is incremented by π (mod 2π) since the black-white preference is reversed.

We determine the phase offset φ of the tuning curves for each cell using nonlinear regression (Matlab Statistics Toolbox function `nlinfit`). Specifi-

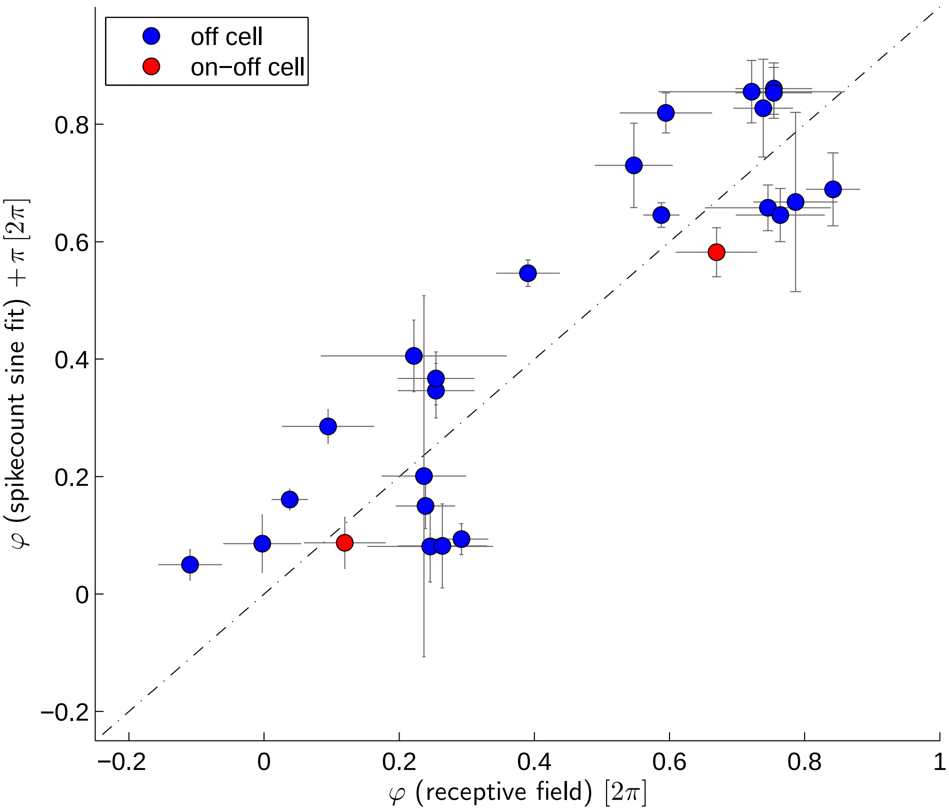


Figure 4.15: Phase of receptive field position versus spike *count* tuning curve, for the same cells as in figure 4.14. Adding π to the phase respects the sign inversion of the spike count curves as compared to the latency curves, since a minimum in the latency curve (i.e., fast response) typically results in a high spike count.

cally, φ and c are the coefficients of the "latency model"

$$\hat{\mathrm{lat}} = c\,\sin(2\pi x + \varphi) \tag{4.10}$$

with the "predictors" $x_k = k/10$ for $k = 0, 1, \ldots 9$ and the experimental (mean subtracted) latency responses lat_k (as in figures 4.6 and 4.7), and similarly for the spike count. In other words, we minimize the least squares error

$$\sum_{k=0}^{9} \left| c \sin\left(2\pi\, k/10 + \varphi\right) - \mathrm{lat}_k \right|^2 .$$

The error bars in figures 4.14 and 4.15 show the 95% confidence interval of φ, calculated by the Matlab function `nlparci`.

Similarly for each receptive field, we fit a two-dimensional Gaussian distribution to the STA (spike-triggered average) introduced in section 3.4.

This yields the receptive field center $\mu \in \mathbb{R}^2$, including 95% confidence intervals for μ_1 and μ_2 (again calculated via the Matlab function `nlparci`). From that, we obtain the receptive field phase

$$\varphi_{\text{recept}} = \frac{\pi \, \mu_2}{w} \pmod{2\pi},$$

where $w = 100\,\mu\text{m}$ denotes the grating width.

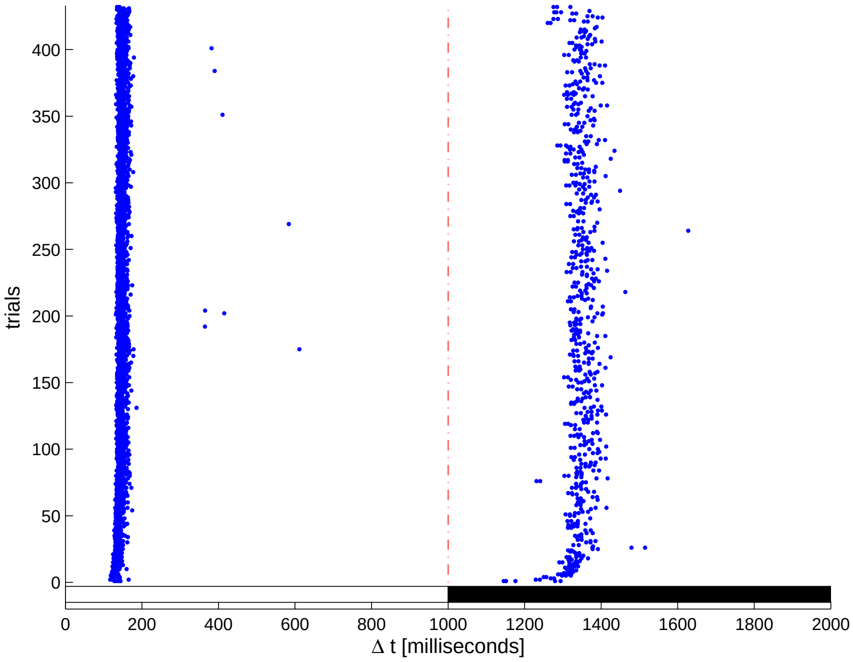


Figure 4.16: Spike scatter plot of the red ("ON-OFF") cell from figure 4.14, presented with a spatially uniform, alternating bright-dark stimulus, as indicated by the black-white bar. The cell responds faster to the ON-step as compared to the OFF-stimulus. Additionally, the variance of the latency (time difference between stimulus onset and first spike) is much smaller for the ON-stimulus (approximately 10 ms) than for the OFF-stimulus (approximately 60 ms). This justifies treating the cell effectively as an ON-cell in figure 4.14.

Interpreting ON-OFF cells as ON without prior knowledge requires justification: figure 4.16 shows the spike scatter plot of the single ON-OFF cell in figure 4.14, when flashing a spatially uniform, alternating bright-dark stimulus onto the retina in order to determine its ON/OFF preference. For this particular cell, we observe that it responds both to the ON and OFF step (hence the characterization as ON-OFF); nevertheless, the ON-response sets in earlier and is more precise as compared to OFF. This justifies that we regard the cell as an ON cell for the purpose of the present section.

Regarding our hypothesis, we clearly observe the tendency of the latency tuning to follow the generator signal difference ΔG defined in (4.9).

However, the agreement is not precise; according to figure 4.14, there is a relatively large variation of latency tuning phases given the receptive field position. This variation cannot solely be explained by the fitting errors of the latency tuning or receptive field positions. Thus, we will refine our model in the subsequent sections.

Considering the spike *count*, figure 4.15 shows the corresponding experimental data for the same cells as in figure 4.14. Since a fast response (i.e., short latency) typically results in a high spike count (see figure 4.4a), the latency and spike count tuning curves are sign-inverted; we account for this fact in figure 4.15 by adding π (mod 2π) to the spike count tuning phase. The spike count data resembles the latency data quite well, and indicates that the receptive field position and the spike count tuning curves are closely related. Thus, the values confirm our initial observation that the spike count might also transmit information about the receptive field position on the grating.

Discussion Concluding, the schematic figure 4.13 (for small $\Delta\varphi$) captures the general tendency of retinal ganglion cells to respond according to the receptive field position on the grating after the rapid movement shift. This is plausible since the position on the grating directly translates to the brightness change ΔG elicited by the shift via equation (4.9) (as long as the movement shift is equal to the grating width). Thus, we can obtain information about the position based on the latency response tuning curve. Interestingly, this result holds true for both the latency and spike count; nevertheless, higher brain areas can – inherently – read out the latency information earlier; in this respect, latency coding is advantageous for *fast* information transmission.

We have to keep in mind, however, that movement amplitudes which are smaller or greater than the grating width might require alternative explanations of the measured responses, since (4.9) no longer holds true. We will investigate these issues in the subsequent sections.

4.5 Dynamics of Movement Shift Leads to Temporal Bias

In the present section, we investigate the effect of the temporal dynamics of the movement shift. The idea is to refine the picture painted in the previous section. Here, we test specifically whether the cell's response indeed depends only on the brightness difference between shift start and target position. We therefore directly compare the responses to two different time-dependent stimulus patterns *during* the movement shift, while the stationary visual patterns before and after the shift are the same in both scenarios. Thus the response should also be the same according to the previous section. This statement defines the null-hypothesis of the current section.

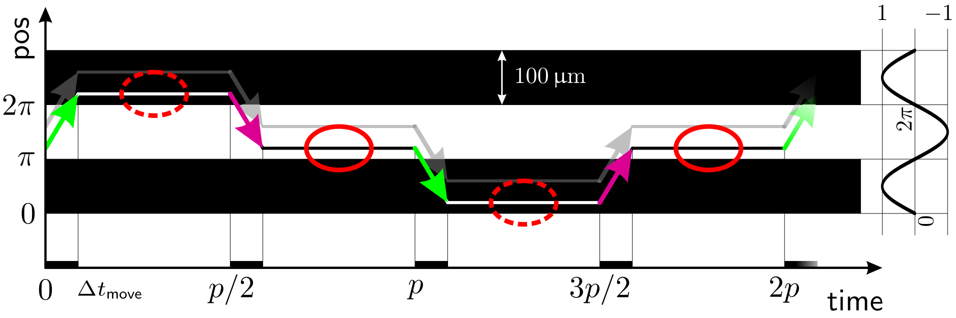


Figure 4.17: "Inverted movement direction" stimulus presentation scheme. The green and purple arrows show the trajectory of a cell's receptive field center relative to the grating. The continuous red ellipses illustrate the cell's receptive field before the movement shift (reference position). Accordingly, the dashed ellipses show the receptive field after movement shift (target position). After 10 repeated trials, the whole grating is offset as illustrated in figure 4.3. The faint gray trajectory indicates the first of these offsets.

Stimulation method The precise time course of the visual stimulus is shown in figure 4.17. The two different time-dependent visual patterns result from the opposite movement directions, which have nevertheless the same starting and target position on the grating (modulo periodicity). The trajectory in the time interval $[0, p]$ is equal to the curve in figure 4.1, imitating (micro)saccades. In other words, the stimulus consists of two copies of figure 4.1, except that the movement directions are inverted in the second copy. This inversion on a single trial basis assures that we avoid adaptation effects.

An essential difference from the previous section 4.4 is as follows: we directly compare phases of latency tuning curves without actually requiring the receptive field position, thus avoiding the associated uncertainty. Nevertheless, the "inverse movement" data can also be obtained from the previous scheme in section 4.4, except for the adaptation effects. For example, in figure 4.3, we can compare movement shifts between the red ellipse pair ($\varphi = 0$) to shifts between the yellow pair ($\varphi = \pi$). The dashed red ellipse "sees" exactly the same stimulus as the continuous yellow one, except that the grating moves in the opposite direction. Indeed, pursuing this idea leads to a plot quite similar to figure 4.18 below (data not shown).

We should emphasize that our focus is not on intrinsic direction selectivity of retinal ganglion cells (as reported in mice by Kim et al. (2008)), but on the time-dependent visual pattern resulting from the trajectory on the grating.

Expected outcome If the response of a retinal ganglion cell actually depends only on the stationary visual stimulus before and after the shift, the latency tuning curve for both green movement shifts in figure 4.17 should be the same. In particular, the phase difference between the tuning curves should be zero. The same holds true for the purple shifts in figure 4.17.

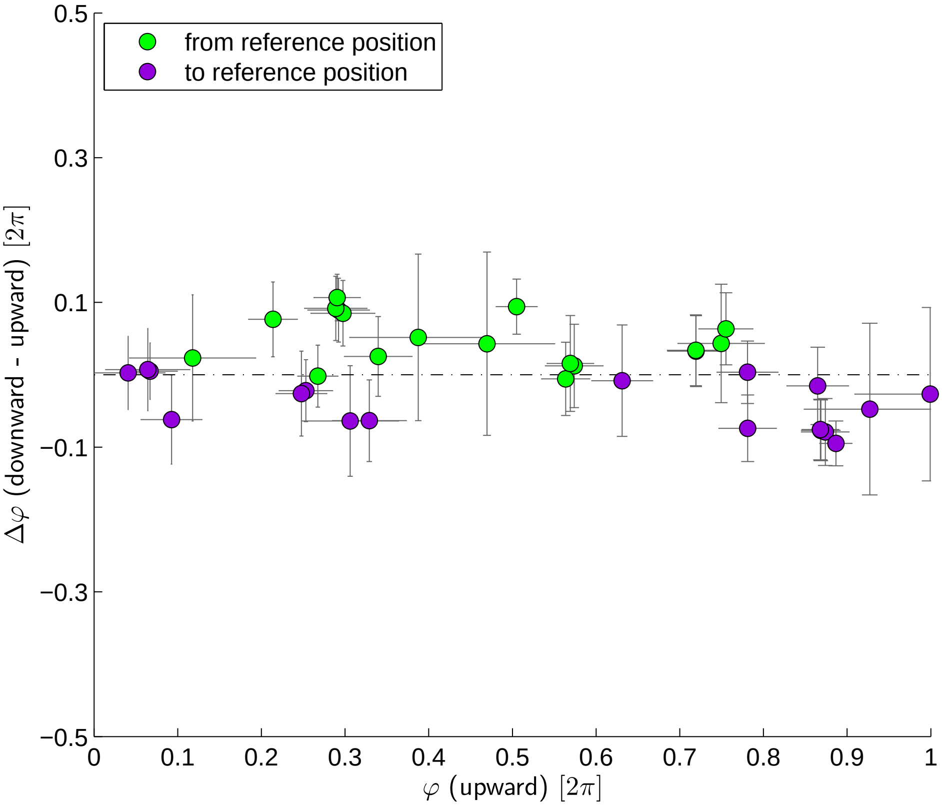


Figure 4.18: Experimental data corresponding to the stimulus scheme in figure 4.17: phase differences between latency tuning curves (separately for the green and purple arrows). The actual receptive field of the cell is not required for the shown data. Green data points are the latency phases corresponding to shifts at time p minus shifts at time 0 in figure 4.17, and purple at time $3p/2$ minus time $p/2$. Note the systematic bias of the green points above and the purple points below the identity line, respectively.

Experimental results The experimental data is shown in figure 4.18, where we try to answer the question whether the temporal movement dynamics has a major influence on the response. The experimental data points show the phase differences between latency tuning curves corresponding to upward and downward rapid shifts. The green data points visualize the latency tuning phase of the second green rapid shift minus the first green shift (see figure 4.17), and correspondingly for the purple data points. Error bars in figure 4.18 are 95% confidence intervals as described above (see equation (4.10)). The plot indicates that the temporal movement dynamics has only a small role, i.e., the data points are close to the dashed identity line at $\Delta\varphi = 0$. Nevertheless, we observe a systematic bias of the green and purple data points.

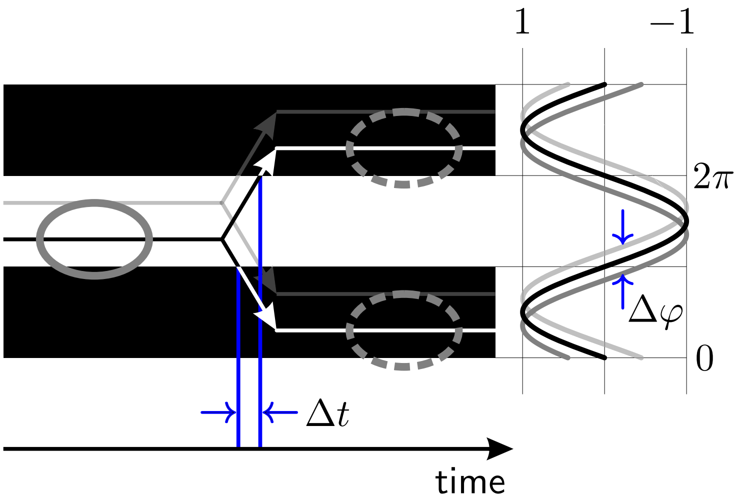


Figure 4.19: Heuristic explanation of the latency bias Δt and corresponding phase difference $\Delta\varphi$ induced by the timing of the black-white border crossing. The figure shows the hypothetical shortest latency scenario for downward movement. Although the starting position of the receptive field (continuous ellipse) should be in the center of the white stripe to maximize the brightness change, the earlier crossing of the black-white border compensates for this sub-optimality. The faint gray trajectories show the mirror-symmetric optimal scenario for the upward shift. In comparison, the latency tuning sine curves are offset by $\Delta\varphi$.

Interpretation Figure 4.19 provides a hypothetical explanation of this bias, by taking the *timing* of the black-white border crossing into account. It illustrates the hypothetical shortest latency scenario for downward movement shifts. The centered black sine tuning curve on the right represents the model where a cell's latency depends solely on the starting and target position, with shortest latency for maximal brightness decrease between start and target (i.e., shift from center of a white stripe to the center of a black stripe, see section 4.4). Figure 4.19 adds a bias to this model by taking the timing of the border crossing into account. Thus, the optimal starting position (continuous ellipse) eliciting the fastest response is now located slightly below the center of the white stripe. This leads to a corresponding phase shift of the latency tuning curve (lower gray sine curve), which we can parametrize as $\sin(x + \varphi)$ with a small $\varphi > 0$. The optimal scenario for *upward* shifts is mirror-symmetric and shown as faint gray trajectories in figure 4.19, with corresponding latency tuning curve $\sin(x - \varphi)$. The resulting phase difference $\Delta\varphi = 2\varphi$ is precisely the distance of the green points in figure 4.18 from the dashed identity line. In particular, our model reproduces the correct sign of the bias. The inverted movement directions of the respective green and purple arrows in figure 4.17 explains the opposite sign of the purple data points. We remark that the systematic bias is not visible in figure 4.14 from the previous section, presumably due to the relatively large uncertainty in the receptive field position.

Discussion Concluding, we have extended the conceptual model from the previous section by a heuristic understanding (figure 4.19) how the temporal dynamic of the movement shift might contribute to the response latency tuning. The critical idea consists of comparing two stimulus scenarios which should lead to exactly the same response according to the simpler model, but show small deviations in experiments. These deviations (i.e., the bias in figure 4.18) is relatively small, so we infer that the general idea of relating response latency to brightness changes still holds true to first approximation, but requires further refinement for a detailed prediction of the cells' responses.

4.6 Contribution of the Visual Image Before and After a Saccade

After each saccade, a different visual image appears on the retina under natural conditions. The visual system has to process this image within the fixation time window, until the next saccade occurs. However, one could ask how the previous image (before the saccade) influences the response to the new image, when assuming that ganglion cells are mostly sensitive to *changes* in activation. In other words, do retinal ganglion cells actually report the differential image change to higher brain regions (instead of the new image, independent of the previous one)?

Thus, the present section aims at a detailed analysis to which degree the pre- and post-shift images contribute to the response latency tuning curves.

Stimulation method We employ the stimulus modality illustrated in figure 4.20, presenting two different shift amplitudes (a_1 and a_2) in alternating order to the retina. The stimulus is similar to figure 4.27 from the previous section 4.7, but with the decisive distinction that the dashed ellipses now "see" different parts of the grating. This means that we directly compare situations with different starting or target visual images. As before, the different movement amplitudes a_1 and a_2 lead to different velocities

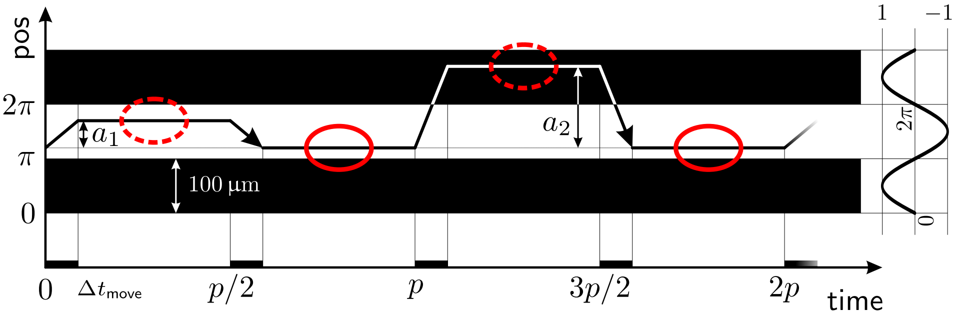


Figure 4.20: Interleaved stimulus scheme presenting two different shift amplitudes a_1 and a_2. The black-white curve visualizes the time-dependent trajectory induced by the shifts, and the two arrows on the trajectory are the shifts with the same target position on the grating (continuous ellipses).

$a_i/\Delta t_\text{move}$ since we fix the shift time interval $\Delta t_\text{move} = 100\,\text{ms}$. The remaining parameters are $p = 1s$, as well as either $(a_1, a_2) = (75\,\mu\text{m}, 125\,\mu\text{m})$ or $(a_1, a_2) = (50\,\mu\text{m}, 150\,\mu\text{m})$.

Expected outcome First, we investigate the hypothesis that the visual image before or after the saccade solely determines the response latency, as exemplified in figure 4.21. In the simplest case, if we assume that the cell is

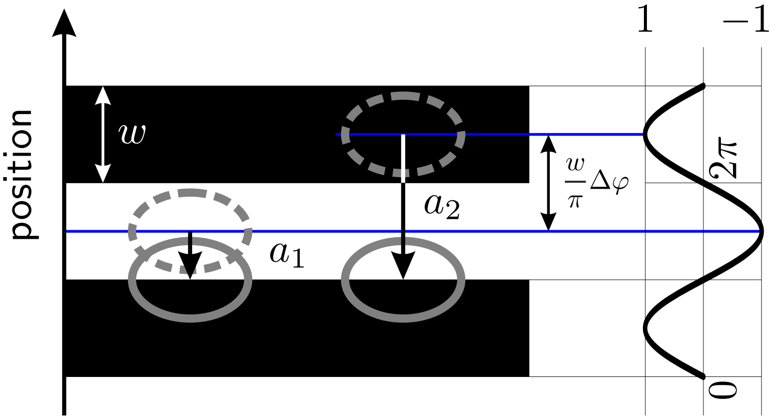


Figure 4.21: Rapid image shifts targeting the same position on the grating, as shown in figure 4.20 (for movement amplitudes $(a_1, a_2) = (50\,\mu\text{m}, 150\,\mu\text{m})$ and grating width $w = 100\,\mu\text{m}$). Assuming that the receptive field position on the grating *before* the shift (dashed ellipses) solely determines the response latency tuning curve, one obtains the phase difference $\Delta\varphi = \pi\,(a_2 - a_1)\,/w$. For the current illustration, this gives $\Delta\varphi = \pi$, and for $a_1 = 75\,\mu\text{m}$, $a_2 = 125\,\mu\text{m}$, we obtain $\Delta\varphi = \pi/2$. On the other hand, if the target position after the shift (continuous ellipse) solely determined the latency tuning, the phase difference $\Delta\varphi$ would be zero.

mostly sensitive to its activation after the shift, it should show the fastest response when targeting the middle of a black stripe (for OFF cells); in particular, the phase differences between the responses to the shifts in 4.21 should be zero. Conversely, under the assumption that the cell cares most about the starting position, it should spike fastest when starting from the middle of a white stripe.

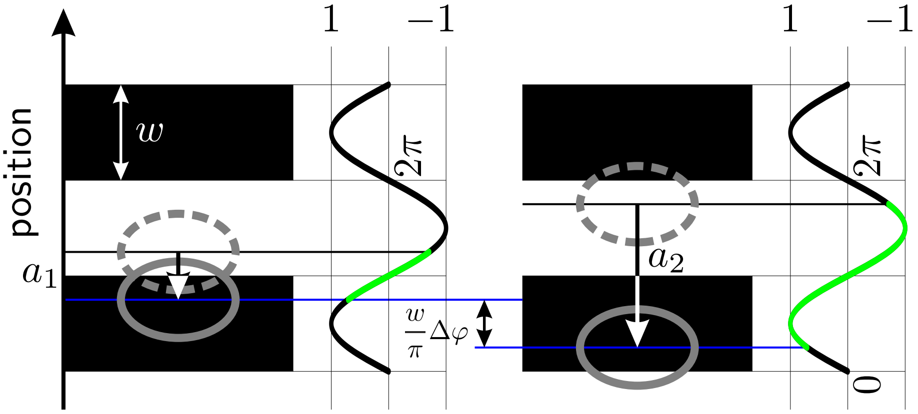


Figure 4.22: Maximal activation difference ΔG between start and target image after a rapid image shift, for the stimulus modality in figure 4.20. The sinusoidal curves shows the activation G dependent on the current receptive field position on the grating, and the green sections the traversed interval during the shift. The movement amplitude on the left side equals $a_1 = 50\,\mu\mathrm{m}$, and on the right side $a_2 = 150\,\mu\mathrm{m}$. In this case, the corresponding phase difference between the continuous ellipses is $\Delta\varphi = \pi/2$.

Second, we investigate the hypothesis that the activation *difference* ΔG between starting and target image of the saccadic shift determines the response tuning. Figure 4.22 solves the task of maximizing this difference for movement amplitudes $(a_1, a_2) = (50\,\mu\mathrm{m}, 150\,\mu\mathrm{m})$, obtaining $\Delta\varphi = \pi/2$ for the resulting phase difference. Note that the traversed green section of the activation curve in figure 4.22 is symmetric around zero. As can be seen from the figure, the general formula for $\Delta\varphi$ corresponding to the greatest activation difference ΔG reads

$$\Delta\varphi = \frac{\pi\,(a_2 - a_1)}{2\,w} \tag{4.11}$$

where a is the movement amplitude and w the grating stripe width. Accordingly, the other set of movement amplitudes $(a_1, a_2) = (75\,\mu\mathrm{m}, 125\,\mu\mathrm{m})$ results in the phase difference $\Delta\varphi = \pi/4$.

Comparing both hypothesis, we observe that the resulting phase difference $\Delta\varphi$ in figure 4.21 (receptive field position before the shift determines the latency response) is twice as large as compared to figure 4.22, which optimizes the activation *difference*.

Experimental results Figure 4.23 shows the experimental phase differences of the latency tuning curves for shifts to the same *target* position (continuous ellipses in figure 4.20). Notably, the blue dots (50 µm, 150 µm shift amplitudes) show phase differences between $\pi/2$ and π (up to one exception), whereas the yellow ones (75 µm, 125 µm shift amplitudes) show phase differences between 0 and $\pi/2$. Quantitatively, the mean values confirm these observations:

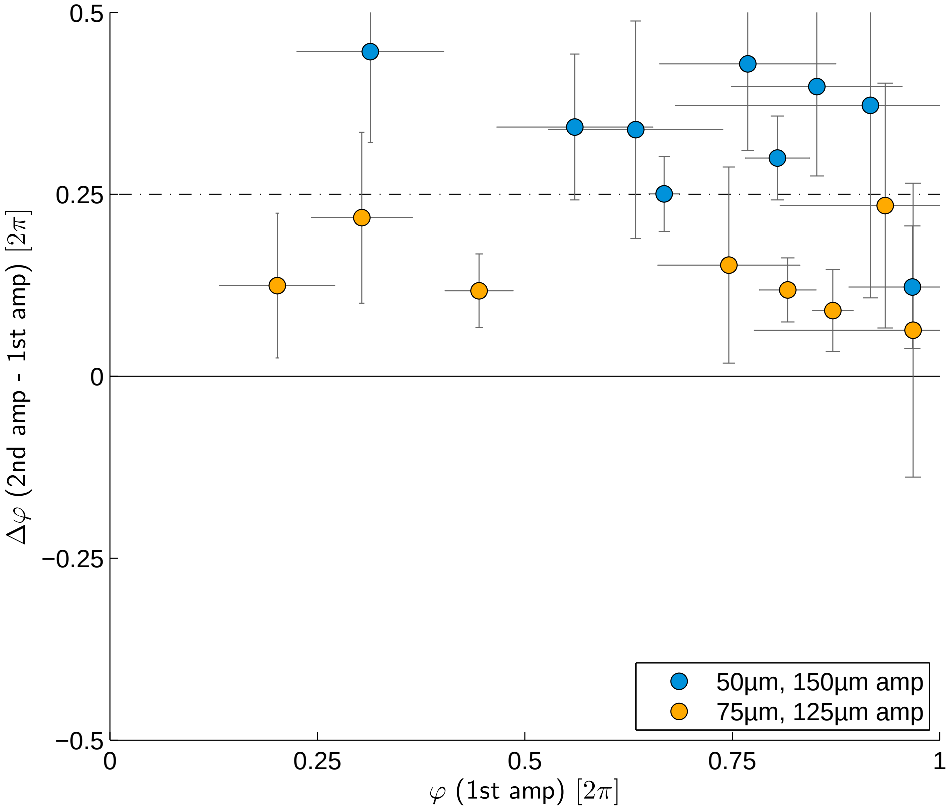


Figure 4.23: Experimental latency tuning phase differences between the two arrows in figure 4.20 (movement towards the same target position), for several ganglion cells. Note that the systematic offset $\Delta\varphi > 0$ (above the dashed identity line) is more pronounced for the blue dots.

$$\langle \Delta\varphi_{\text{same target}} \rangle = 2\pi \cdot (0.33 \pm 0.1) \qquad \text{for } 50, 150\,\mu\text{m}, \quad \text{and} \tag{4.12}$$

$$\langle \Delta\varphi_{\text{same target}} \rangle = 2\pi \cdot (0.14 \pm 0.06) \qquad \text{for } 75, 125\,\mu\text{m}. \tag{4.13}$$

The systematic offset $\Delta\varphi > 0$ (above the identity line) means that the target position does not solely determine the latency; otherwise we would expect $\Delta\varphi = 0$ according to figure 4.21.

Conversely, figure 4.24 contains the experimental data for saccadic shifts *starting* from the same reference position (continuous ellipses in the illustra-

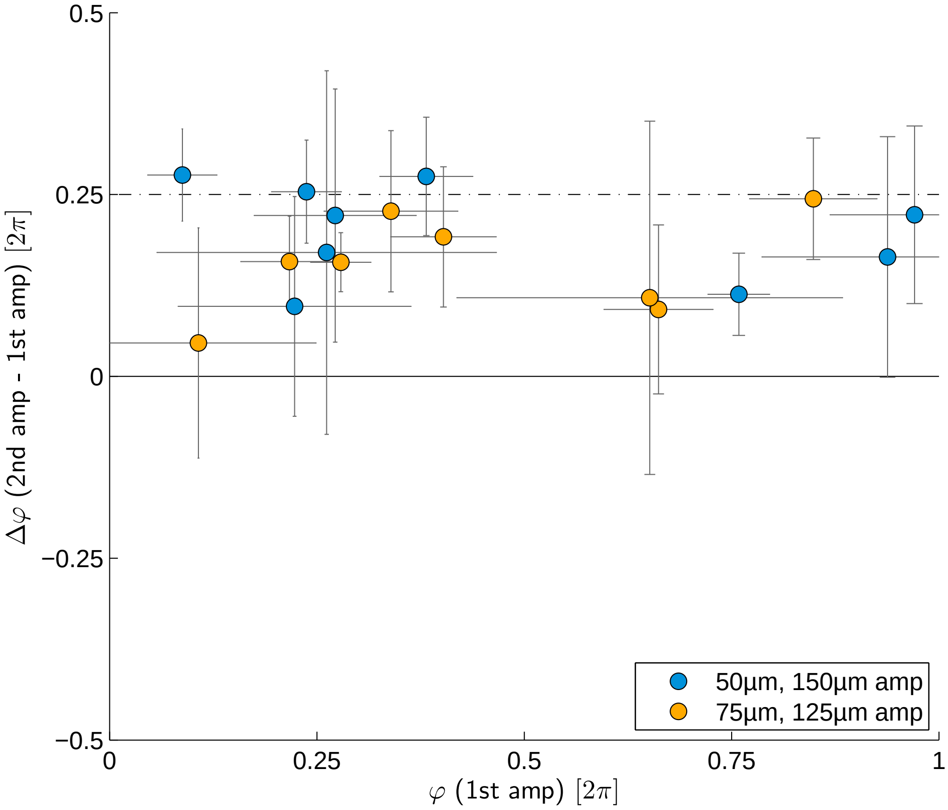


Figure 4.24: Experimental phase differences between latency tuning curves, showing the same cells as in figure 4.23, but for coinciding movement starting positions (stimulus illustrated in figure 4.20). Again, there is a systematic positive phase difference $\Delta\varphi > 0$. In contrast to figure 4.23, both the blue and orange dots are between 0 and $\pi/2$ (up to small deviations).

tion 4.20), but targeting different positions (dashed ellipses). Here, all phase differences are between 0 and (approximately) $\pi/2$, for both amplitude sets. Quantitatively, the mean values are

$$\langle \Delta\varphi_{\text{same start}} \rangle = 2\pi \cdot (0.20 \pm 0.07) \qquad \text{for } 50, 150\,\mu\text{m}, \quad \text{and} \tag{4.14}$$

$$\langle \Delta\varphi_{\text{same start}} \rangle = 2\pi \cdot (0.15 \pm 0.07) \qquad \text{for } 75, 125\,\mu\text{m}. \tag{4.15}$$

Similarly to above, the systematic positive phase difference $\Delta\varphi > 0$ means that the starting position (before the shift) does not solely determine the latency response.

For comparison, the hypothesis that the activation *difference* ΔG between starting and target image determines the latency response (see figure 4.22) predicts $\Delta\varphi = 2\pi \cdot 0.25$ for the blue dots ($50\,\mu\text{m}, 150\,\mu\text{m}$) in both figures 4.23 and 4.24 (since the continuous ellipses serve as reference), which is compatible with the experimental values of equations (4.12) and (4.14).

Similarly, the same hypothesis predicts $\Delta\varphi = 2\pi \cdot 0.125$ for the yellow dots ($75\,\mu\text{m}, 125\,\mu\text{m}$), which approximately agrees with the experimental values of equations (4.13) and (4.15). Furthermore, the experimental data in figure 4.23 complies with the prediction of this hypothesis in the sense that the blue values are approximately twice as large as the yellow ones; however, the data points tend to overshoot the predicted value. In any case, we have captured the correct sign and range of the phase difference $\Delta\varphi$.

Interestingly, the experimental outcome in the second figure 4.24 (coinciding starting position) differs from the first figure 4.23 (coinciding target position) since the range of the blue dots matches the range of the yellow ones in figure 4.24, but not in figure 4.23. Together with the illustration 4.21, this observation might point to the conclusion that the starting position of the saccadic shift is actually more important than the target. Note that the relatively large variation in the experimental values renders a clear distinction between models difficult.

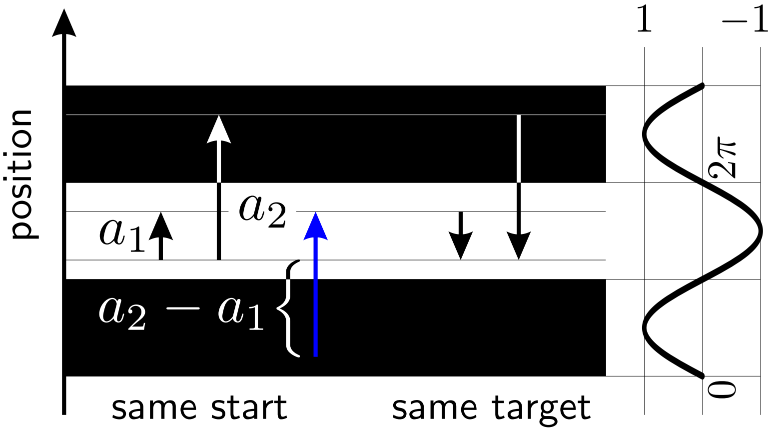
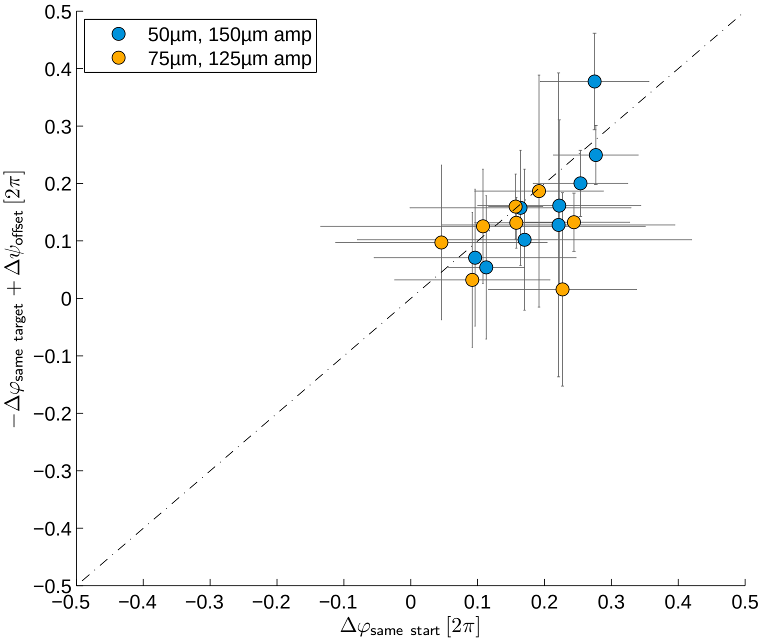
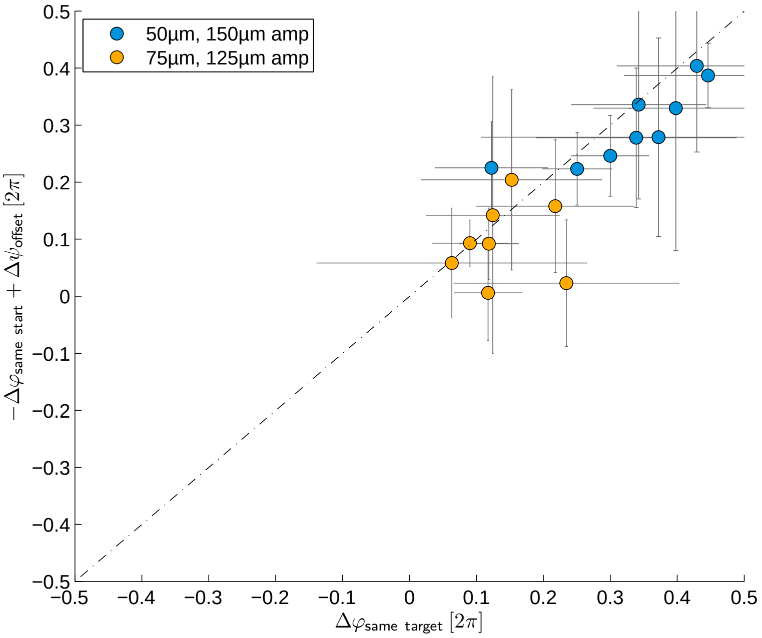


Figure 4.25: Movement shifts starting from the same position (left) or targeting the same position (right), for two different amplitudes a_1 and a_2, as illustrated in figures 4.20 and 4.21. Assuming that cell is not intrinsically direction selective, an offset of the downward-pointing (a_2-amplitude) arrow towards the blue arrow adopts the spatio-temporal generator signal on the left side to the situation on the right side. Additionally, we have to switch signs since the arrow directions on the left and right side are inverted.

Finally, the following idea allows us to perform a consistency check and test the reliability of the experimental results. Under the assumption that the ganglion cells are not intrinsically direction selective, we can transform the stimulus modality "movement to the same target position" to the modality "movement from the same starting position". Concretely, figure 4.25 illustrates how to apply a phase shift offset such that the experimental data points in figure 4.24 are supposedly mapped to the data points in figure 4.23.



(a) x-axis: coinciding starting position (same as y-axis in figure 4.24)



(b) x-axis: coinciding target position (same as y-axis in figure 4.23)

Figure 4.26: Testing the relation $\Delta\varphi_{\text{same start}} = -\Delta\varphi_{\text{same target}} + \Delta\psi_{\text{offset}}$ (figure 4.25) applied to the experimental data points from figures 4.23 and 4.24. The identity line corresponds to exact agreement. Note that in **(a)**, the average latency tuning phase difference $\Delta\varphi_{\text{same start}}$ is similar for both amplitude sets.

Mathematically,

$$\Delta\varphi_{\text{same start}} = -\Delta\varphi_{\text{same target}} + \Delta\psi_{\text{offset}}, \tag{4.16}$$

where $\Delta\psi_{\text{offset}}$ quantifies the "offset" towards the blue arrow in figure 4.25. The phase difference $\Delta\varphi$ switches its sign since the arrow directions are inverted. In mathematical terms, the formula for the offset reads $\Delta\psi_{\text{offset}} = \pi\left(a_2 - a_1\right)/w$, where a_i denotes the shift amplitude and w the grating width. In our case, we obtain π and $\pi/2$ for the two sets of amplitudes, respectively.

We find good agreement with the experimental values, as shown in figure 4.26, when applying equation (4.16) to the data points from figures 4.23 and 4.24. The dashed identity line marks precise agreement. Specifically, subfigure 4.26a plots $\Delta\varphi_{\text{same start}}$ from the left hand side of equation (4.16) on the x-axis and the right hand side on the y-axis. Subfigure 4.26b is analogous after reordering terms, with $\Delta\varphi_{\text{same target}}$ on the x-axis. Since intrinsic direction selectivity would add a correction term to equation (4.16), we have indirectly confirmed that such a selectivity seems to be negligible for our stimulus parameters.

Discussion To summarize our findings, the activation difference ΔG correctly estimates the range and signs of the experimental latency phase differences, but fails to explain why the blue and yellow dots in figure 4.24 coincide while differing by a factor of two in figure 4.23. However, we can match the experimental data by weighting the relative importance of the shift starting and target position. Namely, the about twice as large blue phase differences ($50\,\mu\text{m}, 150\,\mu\text{m}$ amplitude data points) in figure 4.23 point to the conclusion that these ganglion cells are more sensitive to the starting position (or at least the time-dependent activation close to the start of the image shift) instead of the shift target.

4.7 Effect of the Saccadic Shift Amplitude

In the previous section 4.6, we have learned that the saccadic shift *amplitude* seems to play an important role. This is in line with experimental observations that salamanders actually exhibit saccades of various amplitudes (Werner and Himstedt 1985). Thus, our goal in the present section is a simple model of how this amplitude influences the latency response. We directly compare shifts with different amplitudes, but with the same starting and target position on the grating (different from the previous section). Again, our null-hypothesis states that only the stationary visual pattern before and after the saccadic shift determines the response, independent of the pattern during the shift.

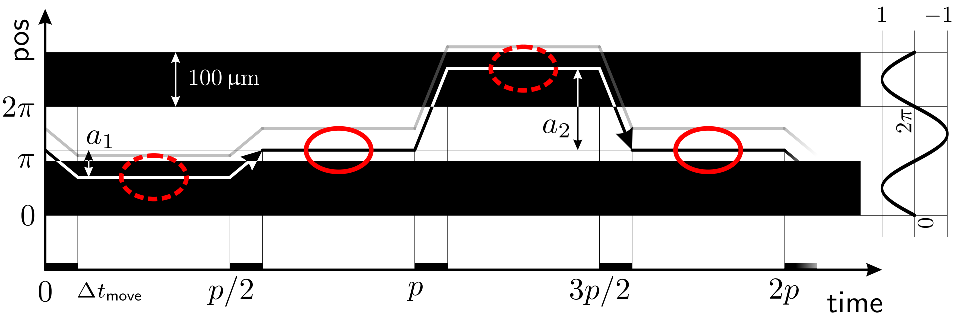
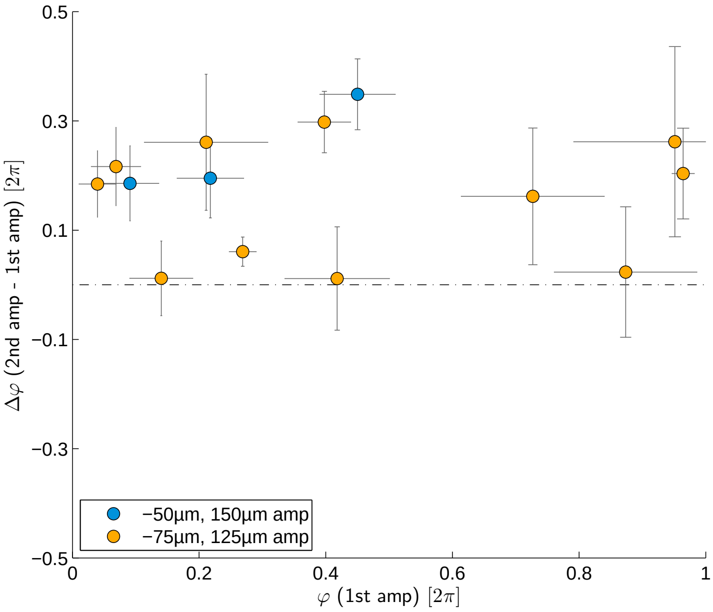


Figure 4.27: Stimulus presentation scheme imitating saccadic image shifts with different movement amplitudes a_1, a_2. The black-white curve visualizes the time-dependent trajectory induced by the shifts. The ellipses represent the receptive field of a single retinal ganglion cell relative to the grating, before and after the rapid shifts. After each 10 repeated trials, the whole grating is offset (see figure 4.3; the faint gray trajectory indicates the first of these offsets).

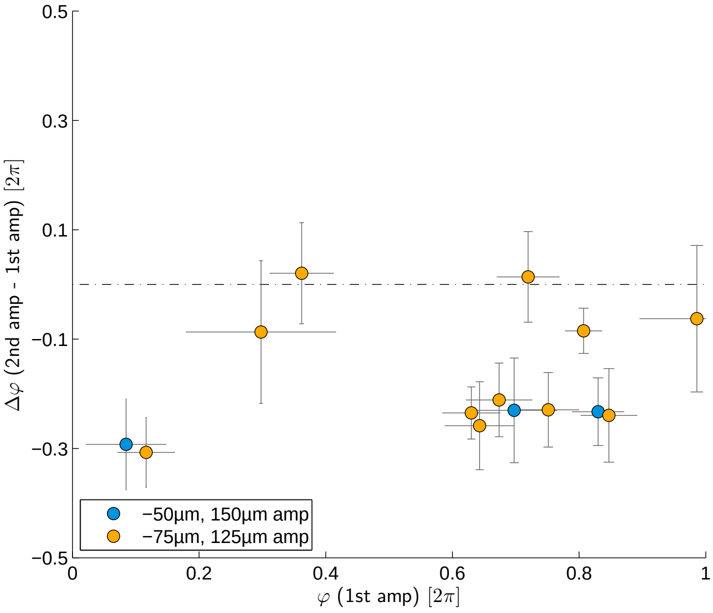
Stimulation method Figure 4.27 illustrates the stimulus for investigating shift amplitudes in detail. Similar to section 4.5, we compare two stimulus modalities presented in alternating order, where the visual stimulus before and after the shift is the same for both modalities. In contrast to section 4.5, we employ two sets of shift amplitudes, denoted a_1 and a_2. The two dashed ellipses in figure 4.27 (movement target) are at the same position on the grating modulo periodicity; thus, the *stationary* generator signal difference is the same for both scenarios (between $[0, p]$ and $[p, 2p]$). In our experiments, the concrete parameters are $p = 1s$, $\Delta t_{\text{move}} = 100\,\text{ms}$ as well as either $(a_1, a_2) = (-75\,\mu\text{m}, 125\,\mu\text{m})$ or $(a_1, a_2) = (-50\,\mu\text{m}, 150\,\mu\text{m})$. As a remark, the different amplitudes also result in different shift velocities $a_i/\Delta t_{\text{move}}$ since we fix the movement time interval $\Delta t_{\text{move}} = 100\,\text{ms}$.

Expected outcome Similar to section 4.5, the spatial generator signal at the target location minus starting location matches in both scenarios. Thus, according to the null-hypothesis, the response should also be the same in both cases. In particular, we expect that the latency tuning phases of the image shifts at $3p/2$ and $p/2$ (figure 4.27) should match, as well as the shifts at p and 0.

Experimental results Figure 4.28 compares the phases of the resulting response latency tuning curves, for several ganglion cells. More specifically, $\Delta\varphi$ on the y-axis of figure 4.28a shows the tuning phase of the response to the image shift at $3p/2$ (see figure 4.27) minus the response to the shift at $p/2$. Correspondingly, the second figure 4.28b compares the responses to the shifts starting at p and 0, respectively. Since the dashed ellipses are at the same position on the grating (modulo periodicity), the only difference



(a) movement towards continuous ellipses in figure 4.27: $\Delta\varphi$ is the difference between latency phases of saccadic shifts at $3p/2$ and $p/2$



(b) movement towards dashed ellipses in figure 4.27: $\Delta\varphi$ is the difference between latency phases of saccadic shifts at p and 0

Figure 4.28: Experimental phase differences between latency tuning curves for the stimulus in figure 4.27. Yellow and blue colors encode two different sets of movement amplitudes. Interestingly, all encountered phase differences are positive ($\Delta\varphi > 0$) in **(a)** and negative in **(b)** (up to small deviations).

between the figures (from the perspective of the ganglion cell) is the opposite saccadic shift direction. Thus, we expect that the experimental data in both figures should match, except that the sign of $\Delta\varphi$ should flip. Indeed, after the sign flip the ranges of $\Delta\varphi$ are similar in both figures.

Remarkably, we observe a systematic offset $\Delta\varphi > 0$ away from the identity line in figure 4.28a, and correspondingly $\Delta\varphi < 0$ in figure 4.28b (up to small deviations), with comparable ranges of values. Quantitatively, the average values combined from both figures (after the sign flip in figure 4.28b) read

$$\langle \Delta\varphi_{\text{latency}} \rangle = 2\pi \cdot (0.25 \pm 0.06) \qquad \text{for } -50, 150\,\mu\text{m}, \qquad \text{and} \tag{4.17}$$

$$\langle \Delta\varphi_{\text{latency}} \rangle = 2\pi \cdot (0.15 \pm 0.11) \qquad \text{for } -75, 125\,\mu\text{m}. \tag{4.18}$$

This contradicts our null-hypothesis, according to which $\Delta\varphi$ should be zero (since the stationary spatial generator signal at the target location minus starting location matches for the compared saccadic shifts).

We have already noted in section 4.4 that the spike *count* may serve as alternative to the latency response for providing information about the visual stimulus. It is conceivable that the latency mostly depends on the transient activation during the beginning of the saccadic shifts, whereas the spike count reflects the brightness difference between starting and target location after the shift. To study this hypothesis experimentally, figure 4.29 compares the spike *count* tuning curve phases, analogous to the latency tuning curves. Again, the data points are significantly offset from the dashed identity line, which means that the hypothesis (brightness change between start and target solely determines tuning phase) does not hold true. Nevertheless, the quantification of the data (analogous to the latency, after a sign flip in figure 4.29b)

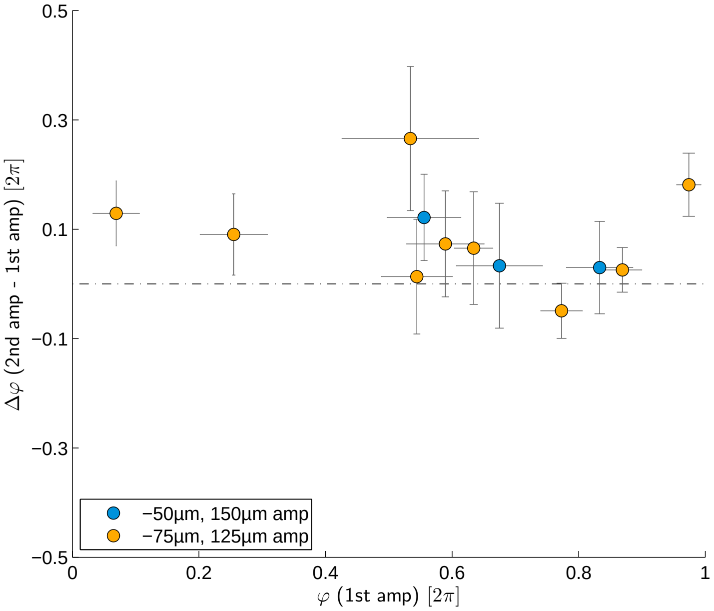
$$\langle \Delta\varphi_{\text{spike count}} \rangle = 2\pi \cdot (0.04 \pm 0.14) \qquad \text{for } -50, 150\,\mu\text{m}, \qquad \text{and} \tag{4.19}$$

$$\langle \Delta\varphi_{\text{spike count}} \rangle = 2\pi \cdot (0.09 \pm 0.10) \qquad \text{for } -75, 125\,\mu\text{m} \tag{4.20}$$

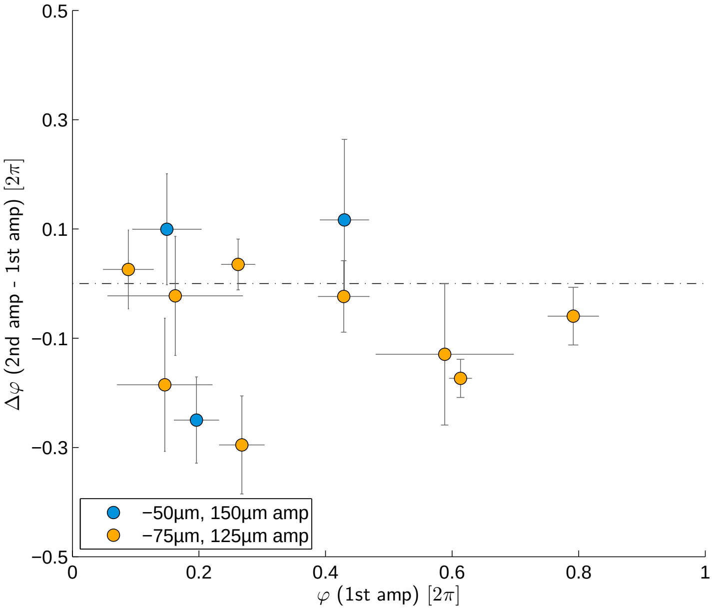
shows that the average values are closer to zero as compared to the latency tuning phases (4.17) and (4.18). This could indicate that the spike count indeed reflects the brightness difference between starting and target location more faithfully than the latency.

Interpretation To find a heuristic explanation of the encountered latency phase differences $\Delta\varphi$, we try to maximize a combination of the following two measures:

i) the activation difference ΔG between start and target position, i.e., the brightness change between the stationary location on the grating before and after the shift

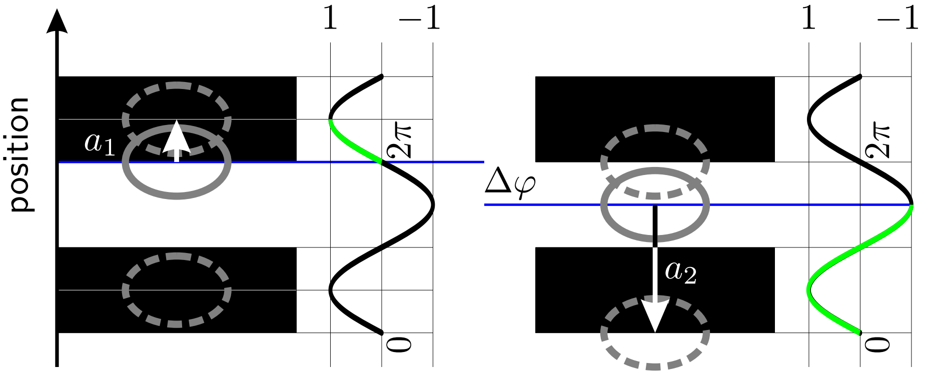


(a) $\Delta\varphi$ is the difference between spike *count* tuning phases of shifts at $3p/2$ and $p/2$ towards continuous ellipses in figure 4.27

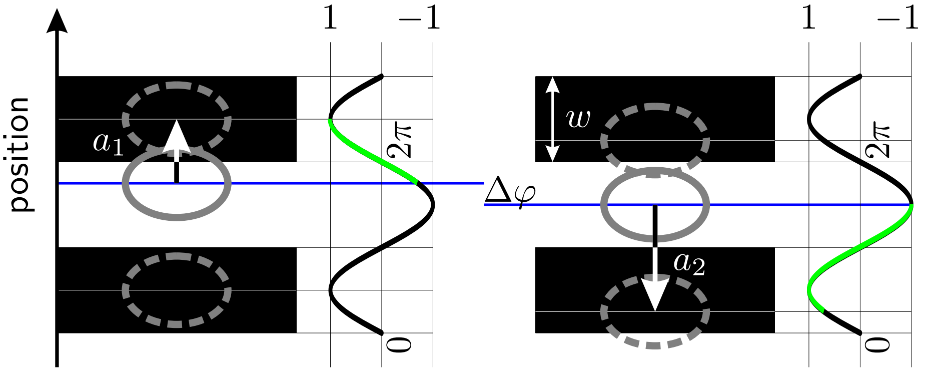


(b) $\Delta\varphi$ is the difference between spike *count* tuning phases of shifts at p and 0 towards dashed ellipses in figure 4.27

Figure 4.29: Experimental phase differences between spike *count* tuning curves, analogous to figure 4.28, for the stimulus illustrated in figure 4.27.



(a) movement amplitudes $(a_1, a_2) = (-50\,\mu\text{m}, 150\,\mu\text{m})$, grating width $w = 100\,\mu\text{m}$ and resulting phase difference $\Delta\varphi = \pi/2$



(b) movement amplitudes $(a_1, a_2) = (-75\,\mu\text{m}, 125\,\mu\text{m})$, grating width $w = 100\,\mu\text{m}$ and resulting phase difference $\Delta\varphi = \pi/4$

Figure 4.30: Heuristic model of the observed latency tuning phase differences $\Delta\varphi$ in figure 4.28, employing the putative explanation that the overall brightness change during the shift as well as the *transient* activation at the beginning of a shift contribute to the latency. Ellipses in each subplot represent the receptive field of a single ganglion cell relative to the grating, with rapid movement shifts starting from the same position (continuous ellipse) and also ending up at the same position modulo periodicity (dashed ellipses), but with different movement amplitudes. Left side: putative scenario leading to the shortest response latency for *upward* motion; right side: correspondingly for *downward* motion.

ii) the transient, instantaneous activation of the cell during the beginning of the saccadic shift.

Point ii) becomes plausible by noting that the first spike is presumably elicited by the fastest depolarizing pathway of the cell; in particular, slow inhibitory pathways leave the latency unaffected if their efficacy sets in after the first spike. In contrast to that, we expect that the total spike *count* following the shift is actually affected by inhibitory pathways.

Figure 4.30 tries to implement i) and ii) for both amplitude sets. Each sine curve in the figure is the location-dependent generator signal induced by the grating. Green subcurves denote the sections traversed during the saccadic shifts. The left side shows the putative scenarios yielding the shortest response latency for upward motion, and – conversely – the right side for downward motion. Specifically in the left subfigure 4.30a, the generator signal crosses the black-white border directly at the beginning of the shift. Similarly, in the left subfigure 4.30b, the generator signal change ΔG between start and final position is close to the maximum and the cell crosses the white-black border early. In the right subfigures, the generator signal ranges from completely bright (at the shift start) to completely dark during the time course of the shift. The resulting phase difference in subfigure 4.30a reads $\Delta\varphi = \pi/2$, which matches the experimental value in equation (4.17); in subfigure 4.30b, we obtain $\Delta\varphi = \pi/4$, which is again within the range of the experimental value in equation (4.18).

Discussion In summary, the experimental latency and spike count phase differences in figures 4.28 and 4.29 have invalidated the null-hypothesis that the activation difference ΔG between starting and target position solely determines the response. Nevertheless, the averages for the spike count in equations (4.19) and (4.20) point to the conclusion that the spike count integrates the net activation ΔG more reliably than the latency. Concerning the latency response, we have constructed a heuristic model (figure 4.30) which qualitatively reproduces the experimental latency data by taking the transient activation during the beginning of the saccadic shifts into account. Interestingly, when applying this model to the stimulus scheme in figure 4.20 from the previous section 4.6, we obtain precisely the same predictions for $\Delta\varphi$ as in equation (4.11). Namely, $\Delta\varphi = \pi/2$ for movement amplitudes $(a_1, a_2) = (50\,\mu\text{m}, 150\,\mu\text{m})$ and $\Delta\varphi = \pi/4$ for amplitudes $(a_1, a_2) = (75\,\mu\text{m}, 125\,\mu\text{m})$, both for coinciding movement starting positions and coinciding target positions. In particular, repeating the analysis from section 4.6, the model complies with the average experimental values but cannot account for the differences between figures 4.23 and 4.24. Thus our sole hypothetical explanation remains the relative importance of the shift starting and target position as discussed in section 4.6.

4.8 Pharmacology Experiments

Pharmacological studies are an effective and widely used tool to investigate the internal workings of neuronal circuits. There is a huge range of how chemicals can influence the biochemical processes at the basis of spiking behavior and signal transmission of neurons; examples include completely silencing neurons (e.g., tetrodotoxin blocks sodium channels), modulating the receptors which uptake neurotransmitters, influencing the synaptic transmission by binding to a neurotransmitter, or actively exciting neurons. Generally speaking, when influencing a specific neuronal subcircuit, we can estimate the role of this subcircuit by comparing it to the control condition (without any drugs).

For our studies, we employ two types of pharmacological experiments: selectively blocking the ON-bipolar pathway, and blocking inhibition. Chapter 3 lists the precise concentrations of the chemicals added to the Ringer's solution which superflows the retina.

4.8.1 Blocking the ON Bipolar Pathway

The broadest-level (and also one of the earliest) classification of retinal ganglion cells partitions them into ON- and OFF-cells, depending on whether they primarily respond to increasing or decreasing light intensity, respectively. As a remark, OFF-cells are more abundant in salamanders (Burkhardt et al. 1998). Physiologically, the retina contains parallel ON and OFF pathways (see figure 2.3 above), which originate at the bipolar cell level; note that these pathways intermingle at the ganglion cell level; for example, ON ganglion cells can receive signals from OFF bipolar cells via an AII amacrine cell and a sign-inverting glycinergic synapse between this amacrine cell and the bipolar cell. Moreover, the dendrites of ON-OFF ganglion cells stratify in at least two sub-layers of the inner plexiform layer (see figure 2.2).

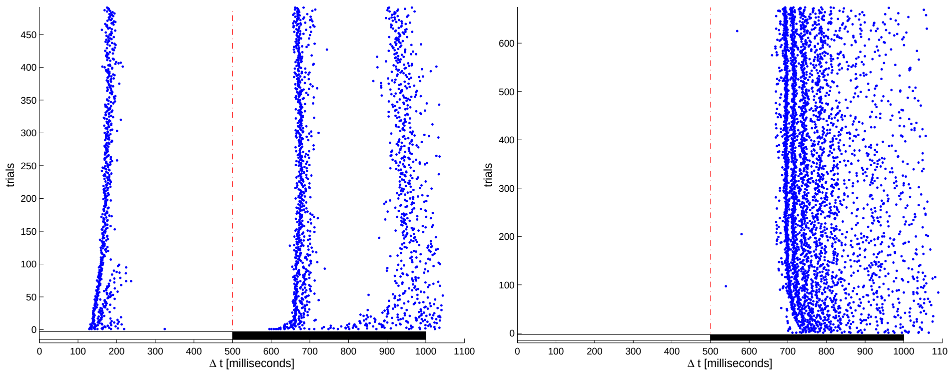
The detailed ON and OFF pathway dynamics are of particular interest for our study since the respective first excitation determines the latency response of a ganglion cell.

Experimental method For our studies, we selectively block the signal transmission from photoreceptors to ON bipolar cells via 200 μM of 2-amino-4-phosphonobutyrate (APB) added to the Ringer's solution (Slaughter and Miller 1981). APB is a metabotropic glutamate receptor agonist (denoted mGluR receptors) (Nakajima et al. 1993; Conn and Pin 1997; Yang 2004). That is, in order to signal increasing brightness, photoreceptors stop emitting the glutamate neurotransmitter at the synaptic cleft to ON bipolar cells. APB effectively interferes with this mechanism since it continuously drives the mGluRs receptors, thus imitating constant darkness from the point of view of the bipolar cell.

The visual stimulus is the same as in section 4.4 above, shown in figure 4.12. That is, we imitate saccades by rapid shifts of the grating, with shift amplitude equal to the grating stripe width $w = 100\,\mu\mathrm{m}$.

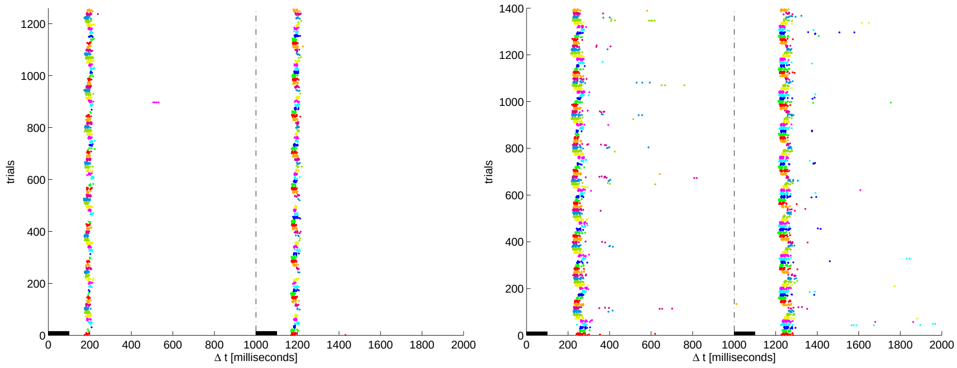
Expected outcome We discuss two scenarios: in the first, the ganglion cell's receptive field has a much smaller extent than the width of a grating stripe. Then all upstream photoreceptors see a brightness increase for the phase offset where the receptive field moves from the center of a black grating stripe to a white stripe. However, since we have blocked the ON pathway, these photoreceptors cannot forward the information to ON bipolar cells. Thus we expect that the ganglion cell remains silent for this particular phase offset, such that we cannot define a latency tuning curve. This is in line with the study by Gollisch and Meister (2008), who report that the ON-pathway is indeed relevant.

In the second scenario, the receptive field has at least the size of a grating stripe. This means that the ganglion cell pools from a spatially widespread area of bipolar cells, such that for every phase offset, at least some of these bipolar cells see a brightness *decrease*. Thus the ganglion cell might respond even if the net brightness level increases. Since our grating stripe width is smaller than in the study by Gollisch and Meister (2008), the second scenario might indeed prevail in our case.

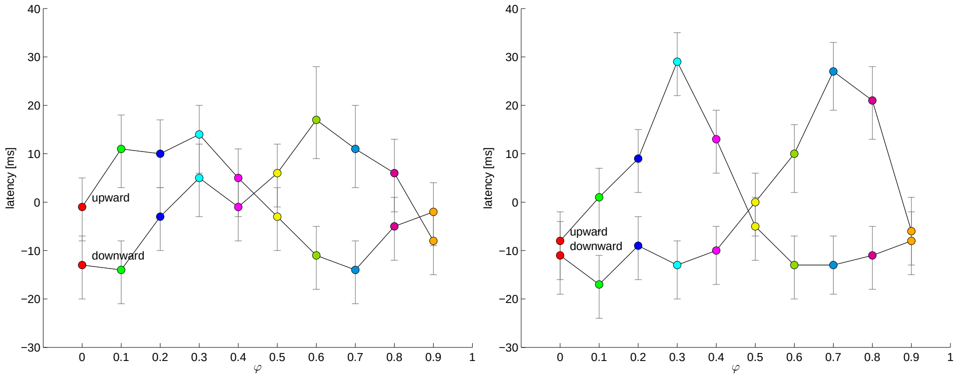


(a) ON-OFF stimulus, control condition **(b)** after blocking th ON pathway

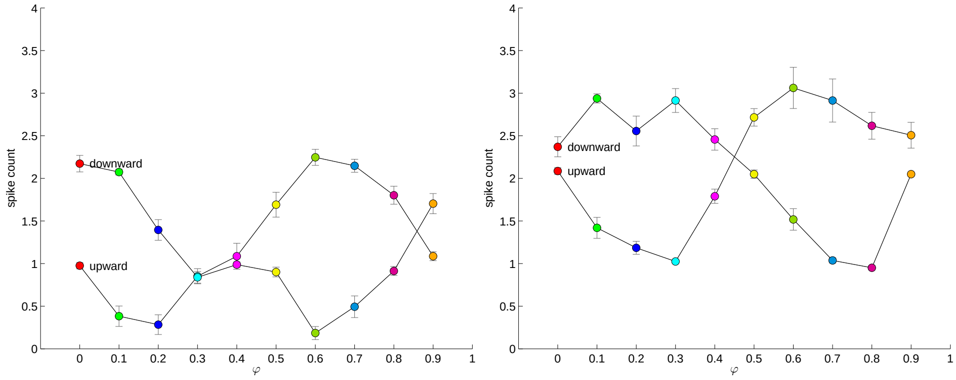
Figure 4.31: Spike raster plot of a single retinal ganglion cell, showing the effects of APB (right side). We employ a spatially uniform, alternating bright-dark stimulus (as indicated by the black-white bars) to test the cell's response to rapidly increasing or decreasing illumination. Left plot shows the control condition, before using pharmacology. As expected, the ON responses disappear, since APB blocks the bipolar ON-pathway. However, note that the temporal pattern of the OFF response changes, too. This effect re-establishes that the cell receives signals from both the ON and OFF pathways.



(a) spike rasters (rapid shifts), control **(b)** spike rasters, APB



(c) latency tuning curve, control condition **(d)** latency tuning curve, APB



(e) spike count, control condition **(f)** spike count, APB

Figure 4.32: Spike raster plots of an individual retinal ganglion cell presented with the rapid-shifting grating stimulus, comparing the effects of APB (right side) to the control without pharmacology (left side). The cell spikes at all phase offsets (i.e., colors) even though the receptive field moves from a black to a white grating stripe at some (unknown) phase offset. The overall shape of the latency and spike count tuning curves hardly changes; only the average spike count slightly increases.

Experimental results First, we test whether the APB treatment actually works. For that purpose, we flash a uniform, alternating bright-dark stimulus onto the retina, as indicated by the white-black bars in figure 4.31. The plot shows the spike response rasters of a single ganglion cell to several hundred trials of this stimulus. We directly compare the control condition (left side) with the effects of administering APB (right side). Clearly, the ON-responses completely disappear, as expected. This confirms that the APB treatment is successful.

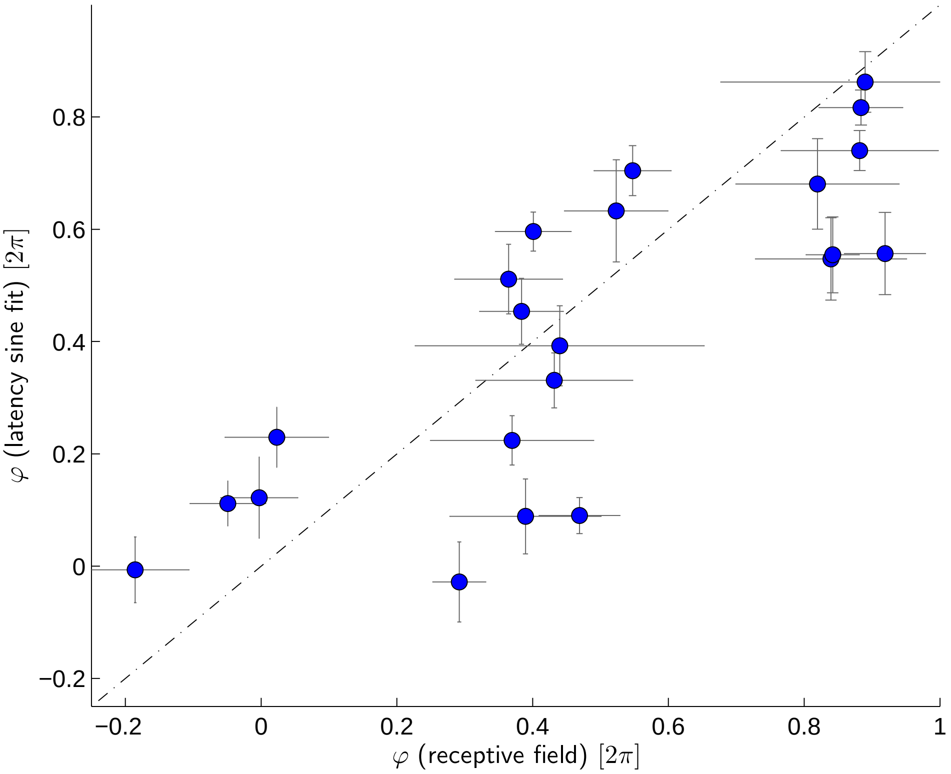


Figure 4.33: Phases of latency tuning curves versus receptive field position on the grating (as in figure 4.14), after blocking the bipolar ON-pathway via APB. Note that ganglion cells which primarily exhibited ON responses before the APB treatment become OFF cells. The figure largely resembles the control condition in figure 4.14. This means that the OFF pathway can suffice for the latency tuning, providing (almost) the same amount of information about the receptive field position on the grating as with the ON pathway enabled.

In the context of rapid image shifts to imitate saccadic eye movements, the results are surprising: according to figure 4.32, the ganglion cell spikes for all phase offsets, even though the cell "sees" a net brightness increase when moving from a black to a white stripe for some phase offset (see figure 4.3). This complies with the second scenario above, meaning that the ganglion cell likely pools from several OFF bipolar cells in a spatially extended area. In any case, figure 4.32 shows that latency tuning curves still exist in the presence of APB.

With this observation, we can ask whether the latency tuning phase still provides information about the receptive field position, as we have investigated in section 4.4. Figure 4.33 tries to answers this question; it is equivalent to figure 4.14 above, except that we have administered APB. (Unfortunately, the sets of ganglion cells in the two plots are not identical, since several cells exclusively respond either in the control condition or after applying APB.) The two figures look quite similar, i.e., even after blocking the ON pathway, the receptive field position and latency tuning are clearly correlated.

Discussion Concluding, we have shown that the OFF pathway in the retina suffices to generate latency and spike count tuning curves, without any contribution from the ON pathway. Moreover, the correlation between the receptive field position on the grating and the tuning phase remains largely unaffected as compared to section 4.4. In particular, there are retinal ganglion cells which fire spikes at any grating location before the saccadic shift, even if the receptive field center moves from a black to a white grating stripe. This is in line with the second scenario above, i.e., each ganglion cell which exhibits such a tuning curve pools from a spatially widespread area of bipolar cells, such that at least some of these bipolar cells see a brightness decrease.

4.8.2 Blocking Inhibition

Neuronal inhibition is a fundamental property of neuronal circuits. For example, the number of inhibitory synapses in the human brain is of the same order of magnitude as the excitatory synapses. Thus, inhibitory mechanisms could play an essential role in our context of latency coding. On the other hand, inhibition putative delays the timing of the first spike (i.e., increases the latency), which counteracts particularly *fast* information transfer. Thus, after investigating whether the latency (and spike count) coding still work at all, we will study the intriguing possibility that blocking inhibition might lead to even faster responses while still carrying information.

Experimental method We employ a cocktail of strychnine, bicuculline and picrotoxin, which act as glycine and $\mathrm{GABA_A}$ receptor blockers. Section 3.6 lists the precise concentrations of these drugs, which we add to the Ringer's solution superflowing the retina.

The visual stimulus is the same as in the previous section 4.8.1 and illustrated in figure 4.12 above, imitating saccades by rapid shifts of the grating.

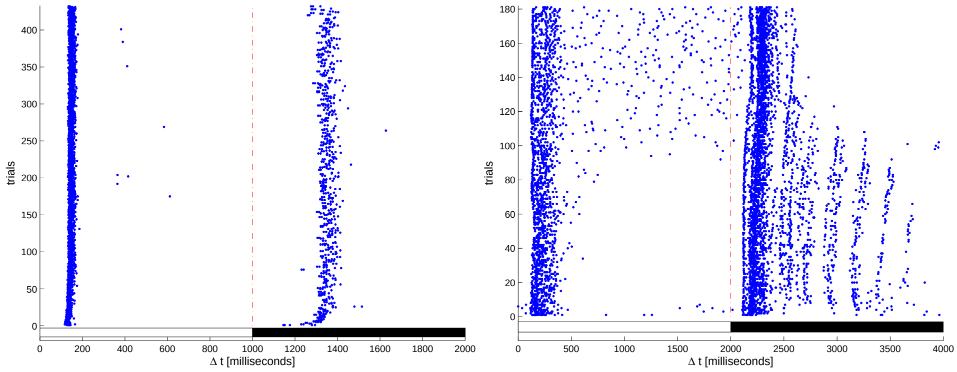
Expected outcome As mentioned, we expect faster latency responses when blocking inhibition as compared to the control condition without pharmacology. Additionally, the overall spike rate should increase.

Experimental results Figure 4.34 visualizes the effects of blocking inhibition, for the same retinal ganglion cell as in figure 4.4b above. We observe the following: first, the overall spike rate increases, as expected. Second, the cell changes from primarily ON (4.34a) to primarily OFF (4.34b), i.e., the flashed dark screen now elicits many more spikes with a shorter latency. Note that the stimulus period in 4.34b is twice as large as compared to 4.34a in order to capture the quite long-lasting responses. Quite interestingly, the cell exhibits an oscillating, sustained spiking pattern to the OFF step after blocking inhibition. The middle row of the figure contains the spike responses to the phase-shifted oscillating grating stimulus. This particular cell shows a less pronounced latency tuning curve as compared to the control condition, i.e., the range of latency values for different phase offsets (colors) is reduced (4.34e and 4.34f). Heuristically, this means that the cell conveys less information about the receptive field position. Note that 4.34c and 4.34e are the same plots as figures 4.4b and 4.7a, respectively, which we have re-plotted to allow for direct comparison. Quite surprisingly, the cell's response sets in at the same time or even later than in the control condition, which contradicts our initial guess that the overall latency should decrease. Summarizing, blocking inhibition seems to degrade the latency tuning of this particular cell.

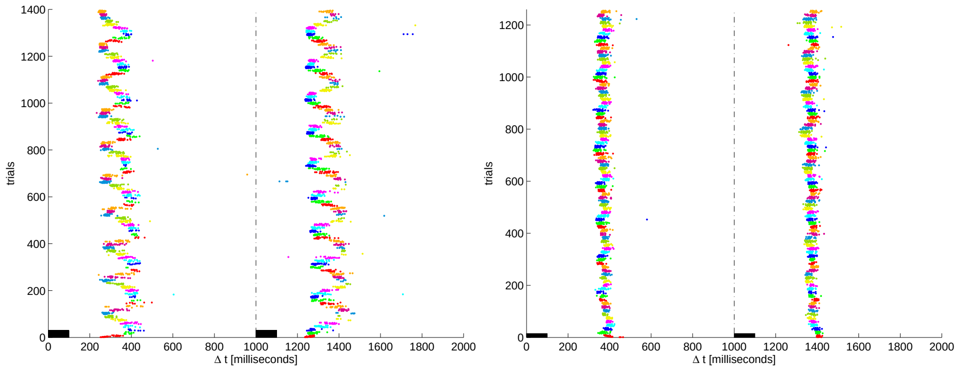
Nevertheless, there are examples of enhanced latency tuning, as demonstrated in figure 4.35 with a different ganglion cell. The cell behaves like an ON cell in the control condition (4.35a), but surprisingly shows a delayed, additional OFF response after blocking inhibition (4.35b). This OFF response drifts over time, different from the ON response. Thus, these features presumably trace back to two different internal mechanisms. As mentioned, the cell exhibits a much broader range of latency values after blocking inhibition, which could stem from the amplified OFF pathway. Different from what we hypothesized, the response sets in about 200 ms *later* than in the control condition. Thus, under normal conditions without pharmacology, higher brain areas can possibly process the visual information much earlier.

Figure 4.36 shows the spike counts of the same ganglion cell for the various phase offsets. The overall number of spikes increases; additionally, the spike counts form a clear tuning curve after blocking inhibition, similar to the latency.

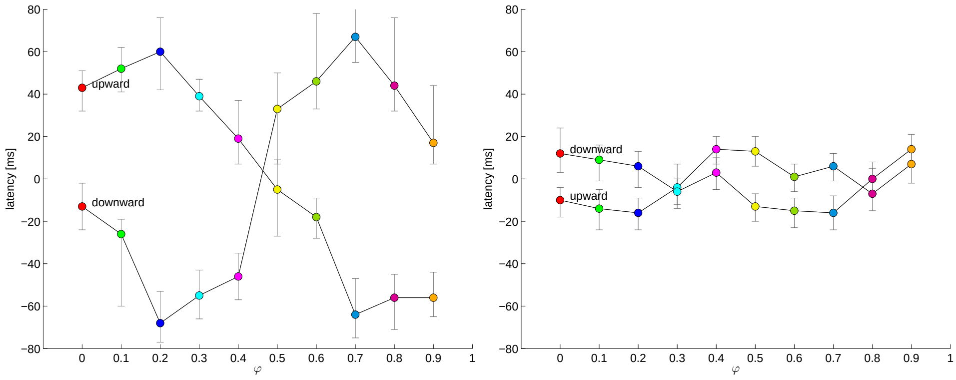
Summarizing, blocking inhibition increases the amount of information about the receptive field position for this cell due to the clearer latency and spike count tuning curves, but on the other hand delays the availability of visual information in higher brain areas.



(a) ON-OFF stimulus, control condition **(b)** after blocking inhibition

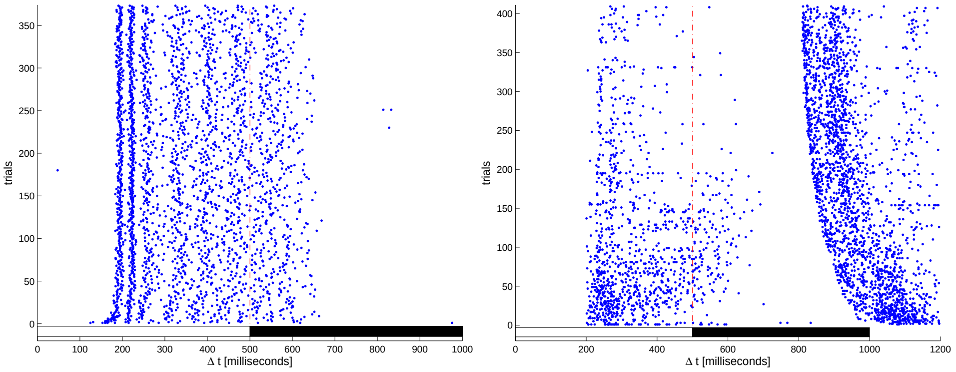


(c) spike rasters (rapid shifts), control **(d)** spike rasters, after blocking inhibition

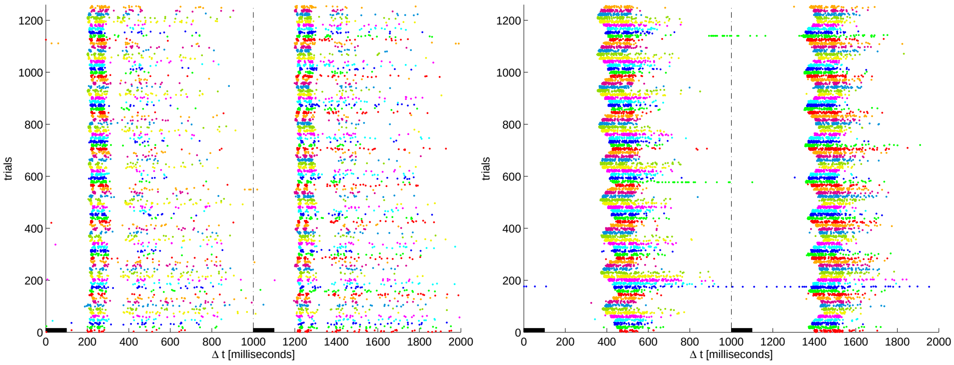


(e) latency tuning curve **(f)** latency tuning, after blocking inhibition

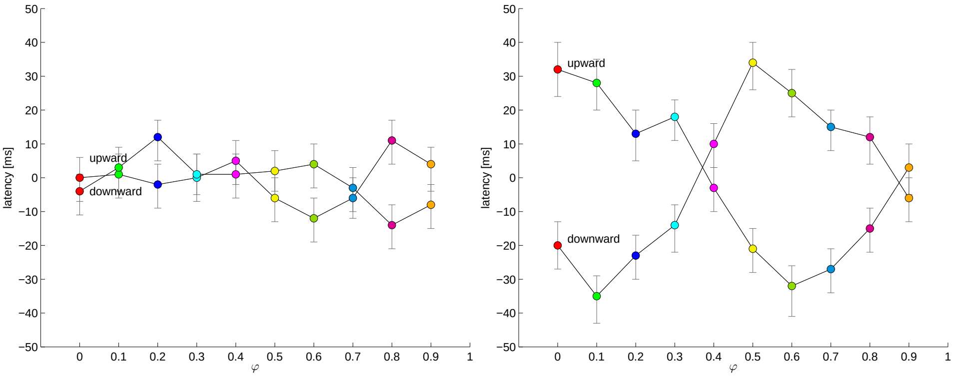
Figure 4.34: Effects of blocking inhibition (right side) exemplified by spike raster plots of the ganglion cell from figure 4.4b. Top row: responses to the spatially uniform, alternating bright-dark stimulus indicated by the black-white bars. The cell changes from primarily ON **(a)** to primarily OFF **(b)**. Middle row: responses to the phase-shifted oscillating grating stimulus. The tuning becomes seemingly less pronounced after blocking inhibition. Bottom row: corresponding latency tuning curves. **(c)** and **(e)** are the same plots as figures 4.4b and 4.7a, respectively.



(a) ON-OFF stimulus, control condition **(b)** after blocking inhibition

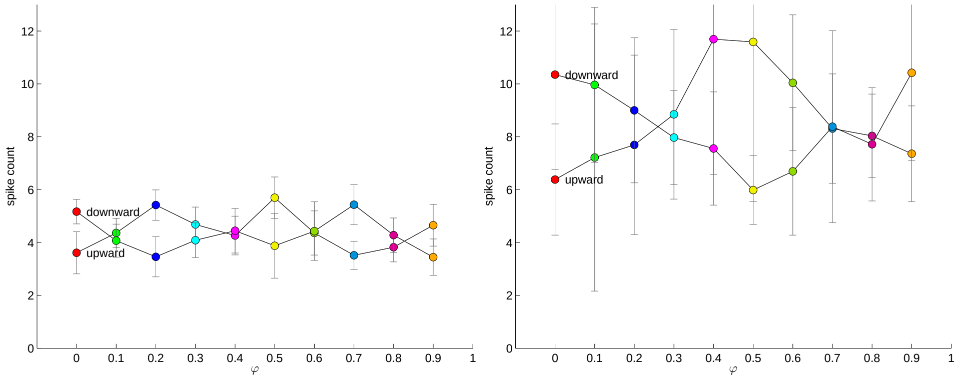


(c) spike rasters (rapid shifts), control **(d)** spike rasters, after blocking inhibition



(e) latency tuning curve **(f)** latency tuning, after blocking inhibition

Figure 4.35: Effects of inhibition blockers (right) demonstrated by a single retinal ganglion cell (as in figure 4.34). The cell is clearly ON type by its exclusive response to the flashed bright stimulus **(a)**, but starts exhibiting an additional, delayed OFF response **(b)** after blocking inhibition. For the phase-shifted grating stimulus, the spike rate largely increases (as expected). Different from figure 4.34, the latency tuning curve becomes more pronounced, and retains a sinusoidal tuning pattern. Notably, blocking inhibition delays the latency response by ≈ 200 ms in **(d)**.



(a) spike counts **(b)** spike counts, after blocking inhibition

Figure 4.36: Spike count tuning curve of the retinal ganglion cell from figure 4.35, showing the effect of blocking inhibition. As expected, the overall number of spikes increases. Moreover, we observe a much larger range of values and a clearer tuning curve.

For statistical quantification, figure 4.37 shows the effects of blocking inhibition as compared to the control in figure 4.14 from section 4.4. As first remark, the latency tuning mechanism still seems to function after blocking inhibition, since the data points show a correlation between receptive field position and latency tuning. On the other hand, the deviation from exact agreement (dashed identity line) is larger than for the control. This means that inhibition probably plays a role for the precise tuning of the responses, as long as the receptive fields have not changed due to the pharmacology. The results point to the conclusion that inhibition increases the precision of the tuning curve, but is not strictly required for the tuning mechanism to work at all.

Discussion We have no evidence that blocking inhibition leads to faster responses, but – to the contrary – have found examples where the response is even delayed. Interestingly, these cells show a pronounced OFF response, which was absent before blocking inhibition. In other words, the pharmacology largely alters the dynamics of the retinal network, which might lead to the observed delay of the response.

Importantly, the phase of the receptive field position on the grating remains an approximate predictor of the latency tuning phase (see figure 4.37) in the presence of inhibition blockers. We have even found an example (in figure 4.35) where the cell's latency tuning becomes clearer pronounced after blocking inhibition (as compared to the control condition), which might enable easier readout in higher brain areas.

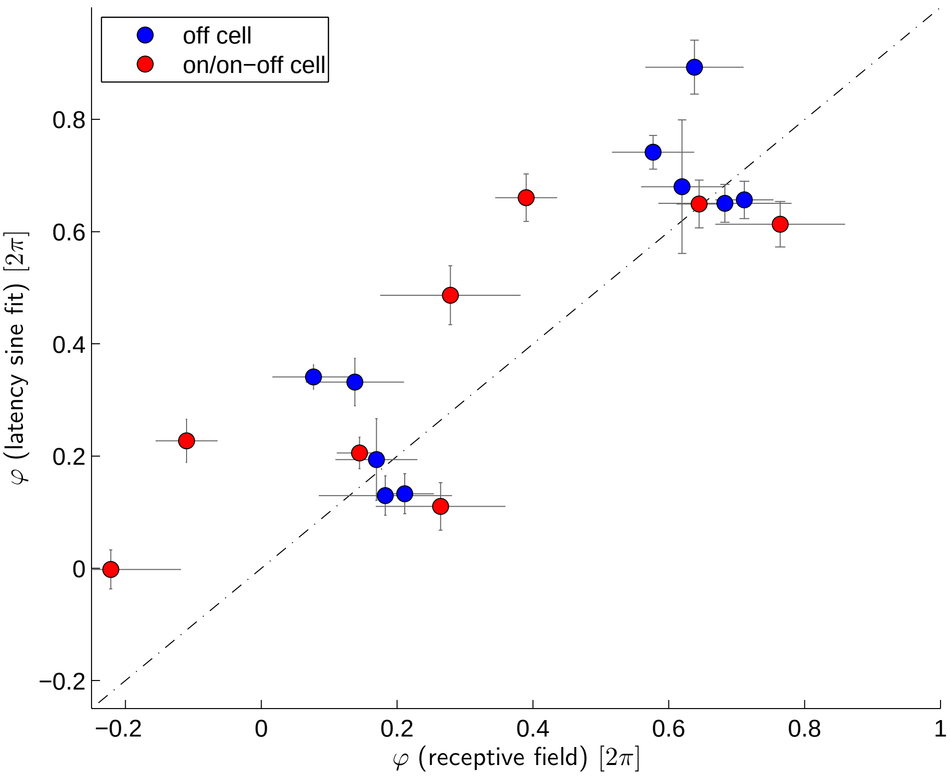


Figure 4.37: Phases of receptive field position on the grating versus latency tuning curve for the rapidly-shifted grating stimulus, showing the effects of blocking inhibition. Refer to figure 4.14 above for the control (without pharmacology).

Chapter 5

Review: Information Theory for Neuronal Networks

Concepts from statistical physics and information theory applied to neuronal networks have attracted growing interest in the neuroscience community. The present chapter is concerned with a review of these concepts.

Functional aspects of complex neuronal systems critically depend on the delicate interplay of neurons. Capturing this interplay requires *parallel* recordings of neuronal populations, as enabled with the advent of multi-electrode arrays (see also sections 3.2 and 3.3 in chapter 3). Nowadays it is possible to record from hundreds of individual neurons (Nicolelis and Lebedev 2009; Fitzsimmons et al. 2009). Thus, the question arises how to analyze, quantify and interpret the recorded data (consisting of spike *timings*). In this respect, statistical physics and information theory provide the required quantitative framework; for example, the Shannon entropy is the canonical measure of the uncertainty of a probability distribution.

5.1 Basic Concepts: Entropy and Mutual Information

Consider a probability distribution $P(x)$ on a discrete set of values. Then, the *Shannon entropy* is defined as

$$H(P) := -\sum_x P(x) \log P(x), \tag{5.1}$$

where we understand that the logarithm log is taken to base 2, and $0 \log 0 := \lim_{p \to 0} p \log p = 0$.

Consider the following setting: $X_1, X_2, \ldots$ are i.i.d. (identically, independently distributed) random variables, which serve as source for random sequences $x = (x_1, x_2, \ldots, x_n)$. By a *compression scheme of rate R* for such a sequence we understand a mapping of x to a bit string of length nR.

Shannon asked for the optimal (i.e., smallest possible) rate R of a reliable compression scheme:

Theorem 1 (**Shannon's noiseless channel coding theorem** (Shannon and Weaver 1949; Nielsen and Chuang 2000))**.** *Suppose* $\{X_i\}$ *is an i.i.d. information source with entropy rate* $H(X)$*. Suppose* $R > H(X)$*. Then there exists a reliable scheme of rate* R *for the source. Conversely, if* $R < H(X)$*, then any compression scheme will not be reliable.*

Given two probability distributions P and Q, the *relative entropy* or *Kullback-Leibler divergence* of P to Q is defined by

$$H_{\mathrm{KL}}(P \parallel Q) := \sum_x P(x) \log \frac{P(x)}{Q(x)} = -H(P) - \sum_x P(x) \log Q(x).$$

Darroch and Ratcliff (1972) denote it as "discriminatory information function" and propose an iterative algorithm to arrive at the "maximum entropy" Boltzmann distribution (see section 5.2).

See e.g. Nielsen and Chuang (2000) for a short proof (using $\ln x \leq x - 1$ for $x > 0$) of the following

Theorem 2 (**Non-negativity of the relative entropy**)**.** *The relative entropy is non-negative,* $H_{\mathrm{KL}}(P \parallel Q) \geq 0$*, with equality if and only if* $P = Q$*.*

Let X, Y be two random variables with joint probability distribution $P(x, y)$. The entropy of X conditional on knowing Y is given by the *conditional entropy* defined as

$$H_{\mathrm{cond}}(X|Y) := H(X,Y) - H(Y).$$

Note that from the above definition of the Shannon entropy, the joint entropy equals

$$H(X,Y) = -\sum_{x,y} P(x,y) \log P(x,y).$$

The *mutual information* between two random variables X and Y is defined as

$$I_{\mathrm{m}}(X:Y) := H(X) + H(Y) - H(X,Y). \tag{5.2}$$

In terms of probability distributions: denote the joint probability distribution by $P(x, y)$ and the marginal distributions by $P(x) \equiv \sum_y P(x, y)$ and $P(y) \equiv \sum_x P(x, y)$. Then, according to equation (5.2), the mutual information equals

$$I_{\mathrm{m}}(X:Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}.$$

Note that this is precisely the Kullback-Leibler divergence between the joint probability distribution $P(x, y)$ and the product of the marginal distributions, $P(x)P(y)$. In other words, the mutual information is a special form of the Kullback-Leibler divergence.

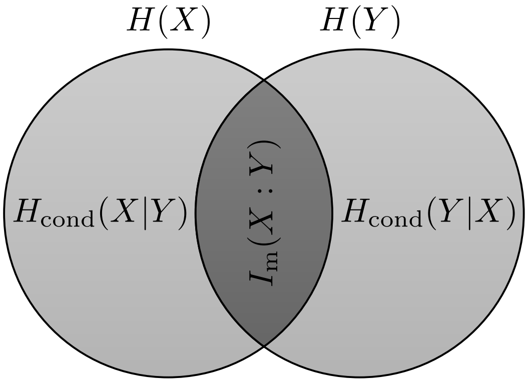


Figure 5.1: "Entropy Venn diagram": illustration of the various relationships between entropies. Adapted from Nielsen and Chuang (2000).

Figure 5.1 depicts the relations between various entropies.

In the following, we follow the discussion in Dayan and Abbott (2005) and interpret the various information theory quantities in terms of the stimulus – response framework. The random variable S captures the statistical distribution of a collection of stimuli denoted s, and R the corresponding response statistics.

For a *fixed* stimulus s, the entropy of the response distribution $P(r|s)$ reads

$$H(R|s) = -\sum_r P(r|s)\log P(r|s).$$

Averaging over all stimuli yields the *noise entropy* of the response (Dayan and Abbott 2005), which we identify as conditional entropy (response conditional on knowing the stimulus):

$$\begin{aligned}\langle H(R|s)\rangle_s &= -\sum_{r,s} P(r,s)\log P(r|s) \\ &= H(S,R) - H(S) = H_{\text{cond}}(R|S).\end{aligned} \tag{5.3}$$

We can interpret the noise entropy as intrinsic response variability, which does not originate from the changes of the stimulus (Dayan and Abbott 2005).

This leads to an alternative interpretation of the mutual information between stimulus and response, namely, the entropy of the response decremented by the noise entropy (5.3):

$$I_{\text{m}}(S:R) = H(R) - \langle H(R|s)\rangle_s\,. \tag{5.4}$$

Since both terms are nonnegative, $H(R)$ is an upper bound on the mutual information (see also theorem 3 below).

Finally, we list some relationships between the various entropies defined above. We will need the concept of a *Markov chain* $X_1 \to X_2 \to \cdots$ of random variables, defined via the requirement that X_{n+1} is independent of $X_1, \ldots, X_{n-1}$, given X_n (Nielsen and Chuang 2000):

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_1 = x_1) = P(X_{n+1} = x_{n+1} | X_n = x_n).$$

As an interesting remark, one can show directly via the definition that if $X \to Y \to Z$ is a Markov chain, then also $Z \to Y \to X$.

Theorem 3 (**Basic properties of Shannon entropy** (Nielsen and Chuang 2000))**.** *The following relationships between the various entropies hold true:*

- $H_{\mathrm{cond}}(Y|X) \geq 0$ *and thus* $I_{\mathrm{m}}(X : Y) \leq H(Y)$*, with equality if and only if* Y *is a function of* X*,* $Y = f(X)$*.*
- Subadditivity*:* $H(X,Y) \leq H(X) + H(Y)$ *with equality if and only if* X *and* Y *are independent random variables.*
- Strong subadditivity*:* $H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$*, with equality if and only if* $Z \to Y \to X$ *forms a Markov chain.*

The proof of subadditivity uses $\ln x \leq x - 1$ for all positive x, with equality if and only if $x = 1$.

5.2 The "Maximum Entropy" Principle

A fundamental physical concept originating from thermodynamics is the maximum entropy ("maxent") principle first proposed by Jaynes (1957a,b). In abstract terms, our goal is to find a statistical model for a physical quantity X (i.e., the probability distribution $P(x)$) subject to given expectation values $\bar{f}_i$ of some functions $f_1, \ldots, f_K$,

$$\bar{f}_i \overset{!}{=} \langle f_i(x) \rangle_P = \sum_x P(x) f_i(x), \quad i = 1, \ldots, K \tag{5.5}$$

and the normalization condition

$$\sum_x P(x) = 1. \tag{5.6}$$

At first sight, P is under-determined in general, i.e., there are several possible solutions to (5.5), (5.6). But there is another criterion which leads to a unique solution. This is precisely the "maximum entropy" principle proposed by Jaynes (1957a): we look for the probability distribution P which is "maximally noncommittal with regard to missing information" given the constraints (5.5), (5.6), i.e., which maximizes the entropy

$$H(P) := -\sum_x P(x) \log P(x). \tag{5.7}$$

A short calculation leads to the well-known Boltzmann distribution

$$P_\theta(x) = \frac{1}{Z(\theta)}\exp\left[-\sum_{i=1}^{K}\theta_i f_i(x)\right] \tag{5.8}$$

with Lagrange multipliers θ_i and the partition function

$$Z(\theta) := \sum_x \exp\left[-\sum_{i=1}^{K}\theta_i f_i(x)\right].$$

These Lagrange multipliers θ_i are chosen to solve the following concise formula for the expectation values,

$$\bar{f}_i \overset{!}{=} \langle f_i(x)\rangle_{P_\theta} = -\frac{\partial}{\partial\theta_i}\ln Z(\theta).$$

In a typical experimental setting, we repeat m trials of a stimulus, obtaining m samples $\{x_1,\ldots,x_m\}$. Thus, the empirical probability distribution (i.e., the relative number of occurrences) reads

$$\tilde{P}(x) := \frac{1}{m}\sum_{n=1}^{m}\delta_{x,x_n}.$$

A short calculation (Broderick et al. 2007) shows that matching the experimental expectation values

$$\bar{f}_i \equiv \langle f_i(x)\rangle_{\tilde{P}} = \frac{1}{m}\sum_{n=1}^{m} f_i(x_n)$$

by the “maximum entropy” probability distribution results in a minimum (over θ) of

$$L_{\tilde{P}}(\theta) := -\frac{1}{m}\sum_{n=1}^{m}\ln P_\theta(x_n) = \langle -\ln P_\theta(x)\rangle_{\tilde{P}}. \tag{5.9}$$

We can exchange the logarithm and the sum in the second term, yielding

$$L_{\tilde{P}}(\theta) = -\frac{1}{m}\ln P_{\text{total}} \quad \text{with} \quad P_{\text{total}} := \prod_{n=1}^{m} P_\theta(x_n).$$

As noted by Broderick et al. (2007), P_{total} equals the probability of drawing the independent experimental samples x_n given the probability distribution P_θ.

5.3 Ising Models of Pairwise Interactions

The following paragraph is concerned with neuronal networks and their statistical spiking patterns. We introduce the approach initiated by Schneidman et al. (2006); Tkacik et al. (2006) and Shlens et al. (2006), which is still being extended and refined (Tkacik et al. 2009; Ganmor et al. 2009, 2011a,b; Shlens et al. 2009; Roudi et al. 2009). The authors investigate whether the spiking statistics of neuronal networks complies with *Ising models* from statistical mechanics and thermodynamics, which we will explain in detail.

Within a small time window (on the order of 10 ms), an individual neuron hardly fires more than one spike (due to the refractory period). Thus, we can capture the instantaneous spiking activity via a binary variable σ_i for each neuron, such that $\sigma_i = 1$ indicates that neuron i fires a spike, and likewise $\sigma_i = -1$ that neuron i remains silent. The spiking activity of the full network thus corresponds to the configuration $\sigma \equiv (\sigma_1, \dots, \sigma_N)$, where N is the number of neurons.

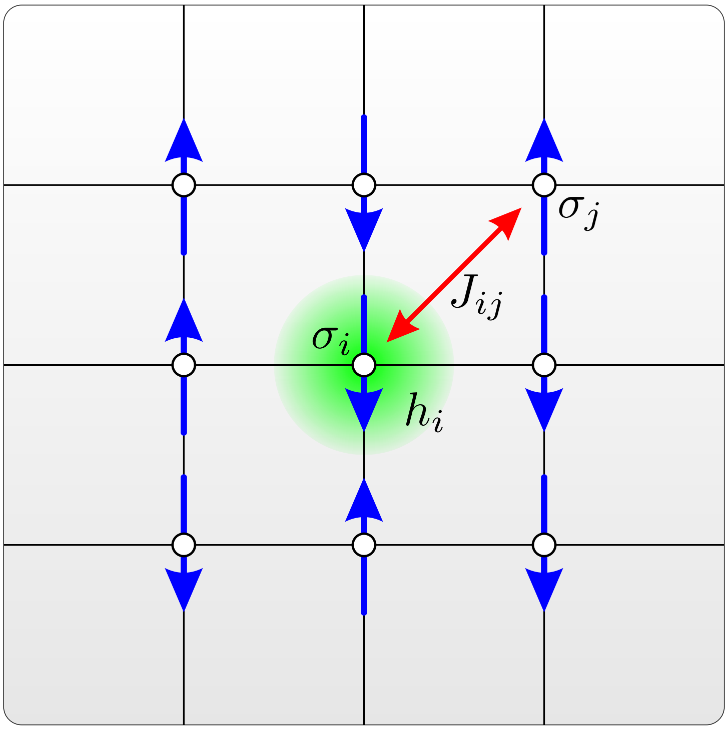


Figure 5.2: Illustration of the (square lattice) Ising model. In physical terms, each lattice site i contains an electron endowed with a "classical" spin variable σ_i, which assumes two possible states ↑ or ↓. Each local magnetic field h_i influences the respective spin σ_i and contributes $h_i\,\sigma_i$ to the total energy, and each pairwise interaction J_{ij} adds $J_{ij}\,\sigma_i\,\sigma_j$ to the total energy.

From a statistical mechanics and thermodynamics perspective, *Ising models* (Ising 1924; Onsager 1944) successfully describe diverse collective effects (like phase transitions) emerging from interacting *pairs*. The physical setup of this model consists of electronic spins σ_i arranged on a regular lattice (see figure 5.2). In the simplest "classical" case, each spin points either up ($\sigma_i = 1$) or down ($\sigma_i = -1$). The overall energy of the network results from local magnetic fields h_i influencing electron i, as well as interactions J_{ij} between all pairs of electrons. That is, the Hamiltonian of the system reads

$$H(\sigma) = \sum_i h_i \, \sigma_i + \sum_{i<j} J_{ij} \, \sigma_i \, \sigma_j .$$

Assume we know the average energy E, i.e.,

$$\langle H(\sigma) \rangle_P \overset{!}{=} E \tag{5.10}$$

equivalent to equation (5.5) above. Then, according to the maximum entropy principle (see previous section 5.2), the probability distribution P consistent with the constraint (5.10) is the Boltzmann distribution $P(\sigma) = \exp[-H(\sigma)/k_B T]/Z$, where – in terms of thermodynamics – k_B is the Boltzmann constant and T the temperature in Kelvin. Comparing with (5.8), we identify $\theta := 1/k_B T$ as Lagrange multiplier which ensures that (5.10) holds true. Note that in the limit of zero temperature $T \to 0$, the spins assume the optimal configuration such that the Hamiltonian reaches its minimum (denoted ground state energy).

From a slightly different perspective, instead of fixing the average energy, we now assume that we know the average of each spin variable, $\langle \sigma_i \rangle$, as well as the correlations among all pairs, $\langle \sigma_i \, \sigma_j \rangle$ for $1 \leq i < j \leq N$. Then, again according to the maximum entropy principle, the Boltzmann distribution (5.8) respecting these constraints reads

$$P_{\text{Ising}}(\sigma_1, \ldots, \sigma_N) = \frac{1}{Z} \exp \left[\sum_i h_i \, \sigma_i + \sum_{i<j} J_{ij} \, \sigma_i \, \sigma_j \right] . \tag{5.11}$$

This is precisely the Ising model to be applied to neuronal population activity. Interestingly, the probability distribution has the same form as above for fixed energy; however, we now interpret the to-be determined parameters h_i and J_{ij} as Lagrange multipliers to resolve the constraints on $\langle \sigma_i \rangle$ and $\langle \sigma_i \, \sigma_j \rangle$.

As central question, one asks whether the Ising model faithfully describes the spike pattern distribution of the neuronal population, i.e., if it suffices to take only *pairwise* interactions into account – a premise which would greatly simplify the overall analysis.

Schneidman et al. (2006) answer this question to the affirmative for networks up to ~ 40 neurons, concluding that the Ising model "captures more than 90% of the structure in the detailed patterns of spikes and silence in the network". On the other hand, according to the authors, the independent model (each neurons fires independently of the others) fails to described the experimentally observed spiking probabilities. We will apply the Ising model to our own experimental spiking data in chapter 6 and investigate whether we can confirm these results. Note that there are two issues in the study by Schneidman et al. (2006): first, the approach cannot account for time-dependent interactions since the model considers the instantaneous activity only. Second, the authors show movies of natural environments, but do not discriminate between the stimulus-induced and intrinsic correlations. Interestingly, a recent study by the same research group (Ganmor et al. 2011b) asserts that pairwise models are not sufficient any more when dealing even larger ($\sim$100) groups of neurons.

The computational challenge in obtaining the probability distribution (5.11) stems from the exponential increase in the number of states $\sigma \equiv (\sigma_1, \dots, \sigma_N)$ with N. To alleviate this difficulty, Broderick et al. (2007) propose a "histogram Monte Carlo" algorithm to calculate the minimum of (5.9) and thus obtain the maximum entropy distribution (5.11).

In a theoretical paper, Tkacik et al. (2010) employ a modified Ising model

$$P(\sigma_1, \dots, \sigma_N | s) = \frac{1}{Z} \exp\left[\beta \Big(\sum_i \left(h_i^0 + h_i(s)\right) \sigma_i + \sum_{i<j} J_{ij}\, \sigma_i\, \sigma_j \Big)\right],$$

with s describing the stimulus, and h_i^0 the basic, stimulus independent firing rate. The "temperature" β encodes neural reliability (or likewise, noise). Namely, neurons become deterministic for $\beta \to \infty$ and completely noisy for $\beta \to 0$ (random firing, independent of the stimulus and interactions J_{ij}). Given β and the stimulus-dependent firing bias $h_i(s)$, the goal consists of optimizing $\left\{h_i^0, J_{ij}\right\}$ to maximize the mutual information $I(\{\sigma_i\}; h(s))$ between $h(s)$ and the response $\{\sigma_i\}$. The authors report that "the optimal population code interpolates smoothly between redundant error-correcting codes, independent codes, and decorrelating codes, depending on the strength of stimulus correlations and neuronal reliability."

Finally, we remark that one may extend the construction of probability distributions as in (5.11) to include correlations of any order k. For example, analogous to (5.11), we obtain for $k = 3$

$$P^{(3)}\left(\sigma_1, \dots, \sigma_N\right) = \frac{1}{Z} \exp\left[\sum_i h_i\, \sigma_i + \sum_{i<j} J_{ij}\, \sigma_i\, \sigma_j + \sum_{i<j<m} K_{ijm}\, \sigma_i\, \sigma_j\, \sigma_m\right]. \tag{5.12}$$

In this model, all combinations of expectation values $\langle \sigma_i \rangle$, $\langle \sigma_i\, \sigma_j \rangle$, $\langle \sigma_i\, \sigma_j\, \sigma_m \rangle$ are assumed to be known, which then fixes the Lagrange multipliers h_i, J_{ij} and K_{ijm}. In this way we arrive at a hierarchy $P^{(k)}, k = 1, 2, \dots, N$ of probability distributions, with $P^{(1)}(\sigma_1, \ldots, \sigma_N) = \frac{1}{Z} \prod_i \exp[h_i\, \sigma_i]$ describing independent random variables, $P^{(2)} \equiv P_{\text{Ising}}$ the Ising model, and $P^{(N)}$ the full probability distribution of N variables (see also Schneidman et al. (2003b)). Note that in the latter case, there are precisely

$$\sum_{i=1}^{N} \binom{N}{i} = 2^N - 1 \tag{5.13}$$

Lagrange multipliers, which is equal to the degrees of freedom (respecting the normalization constraint) of an arbitrary probability distribution for binary variables $\sigma \equiv (\sigma_1, \ldots, \sigma_N)$, as expected.

5.4 Connected Information

In what follows, we introduce the idea of "connected correlations" by Schneidman et al. (2003b). Similar to the previous sections, the approach rests on the maximum entropy principle. Here, we construct probability distributions consistent with k^{th} order marginals of a known (experimentally obtained) probability distribution.

In detail, consider a set of N random variables $X_1, \ldots, X_N$ with joint probability distribution $P(x_1, \ldots, x_N)$. Then, according to the definition (5.1) above, the joint entropy equals

$$H(X_1, \ldots, X_N) = -\sum_{x_1, \ldots, x_N} P(x_1, \ldots, x_N) \log P(x_1, \ldots, x_N).$$

Iterated application of subadditivity (see theorem 3) tells us that the joint entropy is smaller or equal to the sum of the individual entropies,

$$H(X_1, \ldots, X_N) \le \sum_i H(X_i), \tag{5.14}$$

with strict inequality if the random variables are correlated. In analogy to the mutual information (5.2) between *two* variables, the entropy difference between several ones is called *multi-information*:

$$\begin{aligned} I_{\text{m}}(X_1 : X_2 \cdots X_N) &:= \sum_i H(X_i) - H(X_1, \ldots, X_N) \\ &= \sum_{x_1, \ldots, x_N} P(x_1, \ldots, x_N) \log \frac{P(x_1, \ldots, x_N)}{\prod_i P(x_i)} \end{aligned} \tag{5.15}$$

with $P(x_i)$ the marginal probability distributions. According to (5.14), the multi-information is nonnegative, and equal to zero precisely if the random variables are independent.

However, as noted by the authors Schneidman et al. (2003b), the multi-information is not sufficient to resolve the question whether pairwise, triple, or higher order interactions govern the random variables. For a quantitative characterization, the authors employ the maximum entropy distribution, denoted

$$\tilde{P}^{(k)}(x_1, \dots, x_N), \tag{5.16}$$

which is – by definition – consistent with all k^{th} order marginals of the given full distribution $P(x_1, \dots, x_N)$. For example, for $N = 3$ and $k = 2$,

$$\sum_{x_3} \tilde{P}^{(2)}(x_1, x_2, x_3) \overset{!}{=} \sum_{x_3} P(x_1, x_2, x_3) \quad \forall\, x_1, x_2 \tag{5.17}$$

and equivalently for permutations of the x_i. Formally, in terms of equation (5.5) above, these marginal constraints lead to a family of delta distributions

$$f_{x_1,x_2}(y_1, \dots, y_N) := \delta_{x_1,y_1} \delta_{x_2,y_2} \quad \forall\, x_1, x_2$$

(again equivalently for permutations of the x_i) with

$$\bar{f}_{x_1,x_2} := \sum_{x_3} P(x_1, x_2, x_3).$$

Schneidman et al. (2003b) concisely write the Boltzmann distribution (5.8) as

$$\tilde{P}^{(2)}(x_1, x_2, x_3) = \frac{1}{Z} \exp\left[-\lambda_{12}(x_1, x_2) - \lambda_{23}(x_2, x_3) - \lambda_{31}(x_3, x_1)\right]$$

with to-be determined functions $\lambda_{ij}(x_i, x_j)$ such that (5.17) holds. We can justify this step by combining the functions f_{x_i,x_j} with the corresponding Lagrange multipliers θ_{x_i,x_j}, arriving at

$$\lambda_{ij}(y_i, y_j) := \sum_{x_i,x_j} \theta_{x_i,x_j} f_{x_i,x_j}(y_i, y_j) = \theta_{y_i,y_j}.$$

That is, the λ_{ij} are precisely the Lagrange multipliers θ_{x_i,x_j}.

Note that $\tilde{P}^{(1)}(x_1, \dots, x_N) = \prod_i P(x_i)$, i.e., the distribution with all variables independent. On the other hand, we recover the full, true probability distribution at the latest when k reaches N:

$$\tilde{P}^{(N)}(x_1, \dots, x_N) = P(x_1, \dots, x_N).$$

Importantly, this equality might also hold true even if $k < N$.

With the above definitions, Schneidman et al. (2003b) decompose the multi-information (5.15) as

$$I_{\mathrm{m}}(X_1 : X_2 \cdots X_N) = \sum_{k=2}^{N} I_C^{(k)}(X_1, \dots, X_N) \tag{5.18}$$

with the *connected information of order k* defined by

$$I_C^{(k)}(X_1,\dots,X_N) := H\Big(\tilde{P}^{(k-1)}(x_1,\dots,x_N)\Big) - H\Big(\tilde{P}^{(k)}(x_1,\dots,x_N)\Big)\,. \tag{5.19}$$

Schneidman et al. (2006) employ the concept of “connected information” to determine the maximal order k relevant in real neuronal systems, i.e., whether neurons are statistically independent or whether pairwise or even triple interactions play a role. This question is a central topic in chapter 6 of the present thesis.

For the special case of *binary* variables $x_i \equiv \sigma_i \in \{-1, 1\}$, the maximum entropy distribution $\tilde{P}^{(k)}$ respecting all k^{th} order marginals is the same as the maxent distribution for *correlations* up to order k (see section 5.3). In particular for $k = 2$, we arrive at the Ising model, which is plausible since the averages $\langle\sigma_i\rangle$, $\langle\sigma_j\rangle$ and $\langle\sigma_i\,\sigma_j\rangle$ (together with the normalization constraint) completely determine any probability distribution in σ_i and σ_j.

5.5 Synergy, Redundancy and Independence

This section focuses on the following question: Do pairs of neurons convey information – for example, by synchronous activity – that is not present when observing individual neurons only? This questions directly translates to an information theoretic quantity called synergy (Schneidman et al. 2003a), which measures the additional information in the concerted activity about the applied stimulus. Synergy can originate from synchronous spiking, but there are also more complex codes which can elicit synergy. On the other end of the spectrum, it is possible that neurons provide the same, i.e., redundant information (Puchalla et al. 2005). In information theory terms, redundancy equals negative synergy.

The discussion about the interpretation of information theoretic concepts applied to neuroscience is still ongoing (Latham and Nirenberg 2005), sometimes with contradictory experimental results. For example, Nirenberg et al. (2001) put forward the view that neurons are largely independent, which contradicts the findings by Schneidman et al. (2006). Independent of experimental outcomes, we focus on the theoretical concepts from information theory which can be employed to analyze the data. We follow the exposition by Schneidman et al. (2003a).

To cast synergy into mathematical formulas, first recall from equations (5.2) and (5.4) that the mutual information between the stimulus S and the response R_1 and R_2 of two neurons equals

$$I_{\mathrm{m}}(S : R_1, R_2) = \sum_{s,r_1,r_2} P(s, r_1, r_2) \log \frac{P(s, r_1, r_2)}{P(s)P(r_1, r_2)}. \tag{5.20}$$

The *synergy* of these two neurons is defined as

$$\mathrm{Syn}(R_1 : R_2; S) := I_{\mathrm{m}}(S : R_1, R_2) - I_{\mathrm{m}}(S : R_1) - I_{\mathrm{m}}(S : R_2) \tag{5.21}$$

(Reich et al. 2001; Schneidman et al. 2003a). It captures the information difference between the joint response and the individual responses. We call the pair synergistic if the joint response conveys more information about the stimulus than the sum of the individual responses, and – conversely – redundant if the joint response conveys less. As a simple illustration of a completely synergistic cell pair, consider the following (binary) stimulus-response probability distribution (with the remaining combinations of s, r_1 and r_2 having probability zero):

s	r_1	r_2	P
0	0	0	1/4
0	1	1	1/4
1	0	1	1/4
1	1	0	1/4

The individual neurons convey no information at all about the stimulus, $I_{\mathrm{m}}(S : R_i) = 0$, since the conditional distribution $P(r_i|s) \equiv 1/2$ for $i = 1, 2$. That is, no matter what the stimulus, each individual neuron responds with probability 1/2. On the other hand, we can infer with certainty that $s = 0$ if both neurons exhibit the same response, and conversely $s = 1$ if the neurons show different responses. Thus, the synergy turns out to be 1 bit in this example.

At the other extreme, a cell pair is completely redundant if there is a bijective mapping between the responses of the two neurons. Thus, we gain no additional information about the stimulus when observing the second cell if we know the response of the first one already.

As noted by Schneidman et al. (2003a), the synergy divided by the joint mutual information obeys

$$\frac{\mathrm{Syn}(R_1 : R_2; S)}{I_{\mathrm{m}}(S : R_1, R_2)} \in [-1, 1],$$

serving as a normalized version of the synergy.

When investigating the correlations between the two cells, we have to distinguish between stimulus-induced correlations (for example, when the neurons fire simultaneously since they respond to the same stimulus features, possibly without any intrinsic connection between the neurons), and correlations which do originate from internal wiring (denoted "noise correlations"). To quantify these noise correlations, Schneidman et al. (2003a) measure the *conditional independence* of the two neurons via the mutual information for a fixed stimulus s:

$$I_{\mathrm{m}}(R_1 : R_2|s) = \sum_{r_1, r_2} P(r_1, r_2|s) \log \frac{P(r_1, r_2|s)}{P(r_1|s)P(r_2|s)}.$$

This quantity excludes any correlations due to the stimulation since we resort to a single stimulus. Taking the average over all stimuli yields

$$\begin{aligned}\langle I_{\mathrm{m}}(R_1 : R_2|s)\rangle_s &= H_{\mathrm{cond}}(R_1|S) + H_{\mathrm{cond}}(R_2|S) - H_{\mathrm{cond}}(R_1, R_2|S)\\ &= I_{\mathrm{m}}(S : R_1, R_2) - I_{\mathrm{m}}(S : R_1) - I_{\mathrm{m}}(S : R_2) + I_{\mathrm{m}}(R_1 : R_2)\end{aligned}$$

with the joint noise entropy $H_{\mathrm{cond}}(R_1, R_2|S)$ from (5.3). Comparing with (5.21), we arrive at the following alternative definition of synergy:

$$\mathrm{Syn}(R_1 : R_2;\, S) = \langle I_{\mathrm{m}}(R_1 : R_2|s)\rangle_s - I_{\mathrm{m}}(R_1 : R_2). \tag{5.22}$$

Schneidman et al. (2003a) denote the last term, the mutual information between the neurons $I_{\mathrm{m}}(R_1 : R_2)$, as *activity independence*. Namely, it quantifies how much the activity of one neuron conveys about the activity of the other. As noted by the authors, both terms on the right hand side in (5.22) are non-negative, they provide upper and lower bounds on the synergy, respectively.

From an alternative perspective, we can imitate the situation where the cells are recorded non-simultaneously by setting

$$P_{\mathrm{shuffle}}(s, r_1, r_2) := P(s)P(r_1|s)P(r_2|s), \tag{5.23}$$

i.e., assuming “conditional independence” (Schneidman et al. 2003a). As noted by the authors, this is precisely the probability distribution resulting from combining different (“shuffled”) response trials. Note that the marginal distribution of the responses derived from (5.23) reads

$$P_{\mathrm{shuffle}}(r_1, r_2) = \sum_s P(s)P(r_1|s)P(r_2|s).$$

Plugging (5.23) into the definition of the mutual information (5.20) between stimulus and response, we obtain

$$I_{\mathrm{shuffle}}(S : R_1, R_2) := \sum_{s, r_1, r_2} P(s)P(r_1|s)P(r_2|s) \log \frac{P(r_1|s)P(r_2|s)}{\sum_{s'} P(s')P(r_1|s')P(r_2|s')}.$$

Schneidman et al. (2003a) denote the corresponding difference between the standard and shuffled mutual information as

$$\Delta I_{\mathrm{noise}}(S : R_1, R_2) := I_{\mathrm{m}}(S : R_1, R_2) - I_{\mathrm{shuffle}}(S : R_1, R_2).$$

It quantifies the internal, “noise”-induced correlations between the neurons. Similarly, Schneidman et al. (2003a) argue that

$$\Delta I_{\mathrm{signal}}(S : R_1, R_2) := I_{\mathrm{m}}(S : R_1) + I_{\mathrm{m}}(S : R_2) - I_{\mathrm{shuffle}}(S : R_1, R_2). \tag{5.24}$$

“measures the effect of signal-induced correlations on the encoded information”. The authors show that the quantity (5.24) is non-negative. Finally, we arrive at still another definition of synergy

$$\mathrm{Syn}(R_1 : R_2;\, S) = \Delta I_{\mathrm{noise}}(S : R_1, R_2) - \Delta I_{\mathrm{signal}}(S : R_1, R_2), \tag{5.25}$$

which follows directly from the relevant definitions.

5.6 Entropy Estimation

The following section is centered around the approach by Wolpert and Wolf (1995) and Nemenman et al. (2002, 2004); Nemenman (2002) to arrive at an unbiased entropy estimation starting from a (possibly undersampled) probability distribution. For example, imagine we investigate the spike pattern probabilities of a neuronal population, aiming at the entropy of this distribution (with the unknown probabilities denoted as $\boldsymbol{q} \equiv (q_1, \ldots, q_K)$). If pattern i occurs n_i times, then the simplest idea would be to plug the value $\hat{q}_i := n_i/M$ for pattern i into the formula (5.1) of the entropy, where $M := \sum_i n_i$ is the total number of observations. However, this can lead to a severe bias of the resulting entropy, if a pattern i with small probability q_i is not observed at all ($n_i = 0$), for example. The problem typically arises when the limited recording time of the experiments restricts the number of observations M, and the number of possibilities K is large.

A more consistent approach comes from statistical interference, which employs “probabilities of probability distributions”. Namely, we start with a uniform *a priori* distribution on the space of probability distributions,

$$P_{\mathrm{u}}^{\mathrm{prior}}(\boldsymbol{q}) := \frac{1}{Z_{\mathrm{u}}} \delta\left(1 - \sum_{i=1}^{K} q_i\right), \quad Z_{\mathrm{u}} := \int_{[0,1]^K} \delta\left(1 - \sum_{i=1}^{K} q_i\right) \mathrm{d}^K \boldsymbol{q} \tag{5.26}$$

(Wolpert and Wolf 1995). In other words, a priori, all probability distributions $\boldsymbol{q}$ are equally likely as long as they obey the normalization constraint $\sum q_i = 1$. Note that the *likelihood* of the observations $\boldsymbol{n}$ given the probability distribution $\boldsymbol{q}$ equals

$$P(\boldsymbol{n}|\boldsymbol{q}) = \frac{M!}{\prod_i n_i!} \prod_i q_i^{n_i} \tag{5.27}$$

(Wolpert and Wolf 1995). This allows us – conversely – to quantify the probability of the distribution $\boldsymbol{q}$ given the observations $\boldsymbol{n}$ via Bayes’ rule,

$$P(\boldsymbol{q}|\boldsymbol{n}) = \frac{P(\boldsymbol{n}|\boldsymbol{q}) P_{\mathrm{u}}^{\mathrm{prior}}(\boldsymbol{q})}{P(\boldsymbol{n})}, \quad P(\boldsymbol{n}) := \int_{\mathbb{R}_+^K} P(\boldsymbol{n}|\boldsymbol{q})\, P_{\mathrm{u}}^{\mathrm{prior}}(\boldsymbol{q})\, \mathrm{d}^K \boldsymbol{q}.$$

Together with (5.26) and (5.27), the formula leads (up to a normalization factor) to the following expectation value of the entropy $H(\boldsymbol{q}) = -\sum_i q_i \log q_i$:

$$J(H|\boldsymbol{n}) := \int_{\mathbb{R}_+^K} H(\boldsymbol{q})\, \delta\Big(1 - \sum q_i\Big) \prod_{i=1}^{K} q_i^{n_i}\, \mathrm{d}^K \boldsymbol{q}. \tag{5.28}$$

In order to calculate (5.28), Wolpert and Wolf (1995) first rewrite the following integral via a change of variables (for given functions h_i) in terms of the convolution operator $\otimes$:

$$\int_{\mathbb{R}_+^K} \delta\Big(1 - \sum q_i\Big) \prod_{i=1}^{K} h_i(q_i)\, \mathrm{d}^K \boldsymbol{q} = \left(\bigotimes_{i=1}^{K} h_i\right)(\tau)\Big|_{\tau=1}.$$

Together with the Laplace transform, this leads to

$$J(1|\boldsymbol{n}) \equiv \int_{\mathbb{R}_+^K} \delta\Big(1 - \sum q_i\Big) \prod_{i=1}^{K} q_i^{n_i}\, \mathrm{d}^K \boldsymbol{q} = \frac{\prod_{i=1}^{K} \Gamma\left(n_i + 1\right)}{\Gamma\left(M + K\right)} \tag{5.29}$$

(Wolpert and Wolf 1995). To obtain $J(H|\boldsymbol{n})$, the authors make use of the formula $\frac{\mathrm{d}^n}{\mathrm{d}x^n} a^x = a^x \ln(a)^n$ to calculate

$$J(\ln(q_i)|\boldsymbol{n}) = \frac{\partial}{\partial n_i} J(1|\boldsymbol{n}) = J(1|\boldsymbol{n})\left(\psi_0\left(n_i + 1\right) - \psi_0\left(M + K\right)\right)$$

where ψ_m are the polygamma functions defined as

$$\psi_m(z) := \frac{\mathrm{d}^{m+1}}{\mathrm{d}z^{m+1}} \ln \Gamma(z).$$

Note that $J(q_i \ln(q_i)|\boldsymbol{n}) = J(\ln(q_i)|\boldsymbol{n} + e_i)$. Thus, the expectation value of the entropy under the uniform prior (5.26) and for observations $\boldsymbol{n}$ reads

$$\langle H|\boldsymbol{n}\rangle_{\mathrm{u}} = J(H|\boldsymbol{n})/J(1|\boldsymbol{n}) = -\sum_{i=1}^{K} \frac{n_i + 1}{M + K}\left(\psi_0(n_i + 2) - \psi_0(M + K + 1)\right). \tag{5.30}$$

With the same technique, one arrives at explicit formulas for all higher moments $\langle H^m|\boldsymbol{n}\rangle_{\mathrm{u}}$, and thus also for the variance of the entropy,

$$\Delta H = \left(\langle H^2|\boldsymbol{n}\rangle_{\mathrm{u}} - \langle H|\boldsymbol{n}\rangle_{\mathrm{u}}^2\right)^{1/2}.$$

Nevertheless, as noted by the authors, even when observing no data at all ($n_i = 0$ for all i), we obtain the fixed expectation value

$$\langle H|\boldsymbol{0}\rangle_{\mathrm{u}} = -\left(\psi_0(2) - \psi_0(K + 1)\right) = \sum_{i=2}^{K} \frac{1}{i}.$$

Nemenman et al. (2002, 2004) generalize the hitherto analysis and replace – as first step – the uniform prior from (5.26) by a Dirichlet family (in the parameter β) of priors,

$$P_\beta^{\mathrm{prior}}(\boldsymbol{q}) := \frac{1}{Z(\beta)} \delta\left(1 - \sum_{i=1}^{K} q_i\right) \prod_{i=1}^{K} q_i^{\beta - 1}, \tag{5.31}$$

where the δ-term again ensures proper normalization of the probability distribution $\boldsymbol{q}$. Note that we recover (5.26) by setting $\beta = 1$. The normalization factor $Z(\beta)$ is just (5.29) with the n_i replaced by $\beta - 1$, i.e., $Z(\beta) = \Gamma(\beta)^K / \Gamma(K\beta)$.

The Dirichlet family of priors (5.31) effectively maps $n_i \to n_i + \beta - 1$ as compared to the above analysis for the uniform prior. Thus, the generalization of the entropy expectation value in (5.30) reads

$$\langle H|\boldsymbol{n}\rangle_\beta = -\sum_i \frac{n_i + \beta}{M + K\beta} \left(\psi_0(n_i + \beta + 1) - \psi_0(M + K\beta + 1)\right). \tag{5.32}$$

In particular, for the case of no observations ($n_i = 0$ for all i), this reduces to

$$\xi_K(\beta) := \langle H|\mathbf{0}\rangle_\beta = \psi_0(K\beta + 1) - \psi_0(\beta + 1) \tag{5.33}$$

and variance

$$\sigma_K(\beta) := \left\langle (\Delta H)^2 | \mathbf{0} \right\rangle_\beta = \frac{\beta + 1}{K\beta + 1} \psi_1(\beta + 1) - \psi_1(K\beta + 1)$$

(Nemenman et al. 2002). These formulas reveal the main issue with entropy estimation using the Dirichlet priors. Namely, as the authors emphasize, the variance scales like $\sigma_K(\beta) \simeq 1/\sqrt{K}$ for fixed β and large K; in other words, the *a priori* distribution of the entropy is sharply peaked. In particular, the choice of β mostly determines the entropy in the undersampled regime ($M \lesssim K$), irrespective of the observations $\boldsymbol{n}$.

To remedy the situation, Nemenman et al. (2002, 2004) design a prior (denoted "NSB" prior according to the initials of the authors) simply by averaging the Dirichlet family (5.31) over β, such that all mean entropies with respect to this prior are equally likely. Namely,

$$P^{\text{prior}}_{\text{NSB}}(\boldsymbol{q}) := \frac{1}{Z_{\text{NSB}}} \int_0^\infty \frac{\mathrm{d}\xi_K(\beta)}{\mathrm{d}\beta} P^{\text{prior}}_\beta(\boldsymbol{q})\, \mathrm{d}\beta, \tag{5.34}$$

with the average a priori entropy $\xi_K(\beta)$ at fixed β given in (5.33). Bayesian inference follows the same scheme as above, resulting in

$$\langle H^m|\boldsymbol{n}\rangle_{\text{NSB}} = \frac{\int H(\boldsymbol{q})^m P(\boldsymbol{n}|\boldsymbol{q}) P^{\text{prior}}_{\text{NSB}}(\boldsymbol{q})\, \mathrm{d}^K\boldsymbol{q}}{\int P(\boldsymbol{n}|\boldsymbol{q}) P^{\text{prior}}_{\text{NSB}}(\boldsymbol{q})\, \mathrm{d}^K\boldsymbol{q}}.$$

Using the definition (5.34), the numerator (times the constant Z_{NSB}) equals

$$\int_0^\infty \int_{\mathbb{R}^K_+} H(\boldsymbol{q})^m P(\boldsymbol{n}|\boldsymbol{q}) P^{\text{prior}}_\beta(\boldsymbol{q})\, \mathrm{d}^K\boldsymbol{q}\; \xi'_K(\beta)\, \mathrm{d}\beta$$

$$= \int_0^\infty \rho(\beta, \boldsymbol{n})\, \langle H^m|\boldsymbol{n}\rangle_\beta\, \xi'_K(\beta)\, \mathrm{d}\beta$$

with $\langle H^m|\boldsymbol{n}\rangle_\beta$ for $m=1$ explicitly defined in (5.32), and

$$\begin{aligned}\rho(\beta,\boldsymbol{n}) &:= \int_{\mathbb{R}_+^K} P(\boldsymbol{n}|\boldsymbol{q}) P_\beta^{\text{prior}}(\boldsymbol{q})\,\mathrm{d}^K\boldsymbol{q} \\ &= \frac{\Gamma(K\beta)}{\Gamma(M+K\beta)}\prod_{i=1}^{K}\frac{\Gamma(n_i+\beta)}{\Gamma(\beta)} = \frac{1}{(K\beta)_M}\prod_{i=1}^{K}(\beta)_{n_i}.\end{aligned} \tag{5.35}$$

For the last equation, we have employed the *Pochhammer* symbol

$$(x)_n \equiv \frac{\Gamma(x+n)}{\Gamma(x)} = x(x+1)\cdots(x+n-1).$$

Assembling everything, we reproduce the formula by Nemenman et al. (2002),

$$\langle H^m|\boldsymbol{n}\rangle_{\text{NSB}} = \frac{\int_0^\infty \rho(\beta,\boldsymbol{n})\,\langle H^m|\boldsymbol{n}\rangle_\beta\,\xi_K'(\beta)\,\mathrm{d}\beta}{\int_0^\infty \rho(\beta,\boldsymbol{n})\xi_K'(\beta)\,\mathrm{d}\beta}. \tag{5.36}$$

The authors call $\rho(\beta,\boldsymbol{n})$ in (5.35) the experimental *evidence* for a particular value of β since it weights the integral over the Dirichlet expectation value $\langle H^m|\boldsymbol{n}\rangle_\beta$ based on the observations $\boldsymbol{n}$. The last equation completes the analytic part of our NSB entropy estimation survey; in the following, we are concerned with the numerical implementation of the algorithm.

Considering the numerical evaluation of $\rho(\beta,\boldsymbol{n})$, a major difficulty stems from the quickly growing Gamma functions. To avoid numerical over- or underflow, we first rewrite the evidence as

$$\rho(\beta,\boldsymbol{n}) = \mathrm{e}^{-\mathcal{L}(\beta,\boldsymbol{n})} \tag{5.37}$$

with

$$\mathcal{L}(\beta,\boldsymbol{n}) := -\sum_{i=1}^{K}\ln\Gamma(n_i+\beta) + K\ln\Gamma(\beta) - \ln\Gamma(K\beta) + \ln\Gamma(M+K\beta)$$

(Nemenman 2002). Note that the log-Gamma functions $\ln\Gamma$ are already built into numerical software tools like Mathematica or Matlab such that no numerical overflow occurs (as if first evaluating the Gamma function and then taking the logarithm). An essential observation regarding $\rho(\beta,\boldsymbol{n})$ is as follows: for any constant $\mathcal{L}_0$, the re-scaled version

$$\tilde{\rho}(\beta,\boldsymbol{n}) := \mathrm{e}^{-\mathcal{L}(\beta,\boldsymbol{n})+\mathcal{L}_0} \tag{5.38}$$

leads to the same entropy estimate since the scaling factor $\mathrm{e}^{\mathcal{L}_0}$ cancels in the formula (5.36). Thus, a proper choice of $\mathcal{L}_0$ (close to the minimum of $\mathcal{L}(\beta,\boldsymbol{n})$) allows us to escape the numerical overflow problem as if evaluating equation (5.37) directly.

Nemenman (2002) devises a saddle point method to calculate the integrals in (5.36). For that purpose, one needs the (unique) maximizer β^* of $\rho(\beta, \boldsymbol{n})$, or – equivalently – the minimizer of $\mathcal{L}(\beta, \boldsymbol{n})$. Nemenman (2002) derives the following equation (valid to zeroth order in $1/K$):

$$\frac{K_1}{\kappa^*} \overset{!}{=} \psi_0(\kappa^* + M) - \psi_0(\kappa^*), \qquad \kappa^* \equiv K\beta^*. \tag{5.39}$$

Here, K_1 is number of possibilities (or bins) with at least one observed occurrence, i.e., the number of i for which $n_i \geq 1$.

In the remainder of the current section, we share our own ideas regarding the implementation of the NSB algorithm. While Nemenman (2002) derives higher order corrections of (5.39) to determine β^* as precisely as possible, we have found (5.39) sufficient, in the sense that

$$\mathcal{L}_0 := \mathcal{L}(\beta^*, \boldsymbol{n})$$

plugged into the formula (5.38) escapes numerical over- or underflow problem.

Considering the integration over β in (5.36), we employ the substitution

$$\beta \to \left(\frac{\omega}{1-\omega}\right)^2, \quad \omega \in [0, 1), \tag{5.40}$$

which maps the integration domain from $\beta \in [0, \infty)$ to $\omega \in [0, 1)$, and has the added benefit of broadening the support of the integrand (see figure 5.3). The alternative approach taken by Nemenman (2002) is the canonical substitution

$$\int_0^\infty \rho(\beta, \boldsymbol{n}) \left\langle H^m | \boldsymbol{n} \right\rangle_\beta \xi_K'(\beta) \, \mathrm{d}\beta = \int_0^{\ln K} \rho(\beta(\xi), \boldsymbol{n}) \left\langle H^m | \boldsymbol{n} \right\rangle_{\beta(\xi)} \mathrm{d}\xi,$$

where $\beta(\xi)$ is the inverse of the function $\xi_K(\beta)$ defined in (5.33). Thus, our approach saves the effort of actually computing this inverse.

The curves in figure 5.3 stems from a toy model with $K = 9$ possibilities (or bins), observations $\boldsymbol{n} = (4, 2, 3, 0, 2, 4, 0, 0, 2)$, and thus $M = \sum n_i = 17$. Using our implementation, we obtain

$$\left\langle H | \boldsymbol{n} \right\rangle_{\mathrm{NSB}} = 2.79976 \pm 0.225 \,\text{bits},$$

in exact agreement (8 digits) with the Matlab implementation by Nemenman (2011). For comparison, the "naive" calculation of the entropy (plugging the observed frequencies n_i/M into the formula (5.1)) gives $H_{\mathrm{direct}} = 2.5136\,\text{bits}$, which is noticeable different form the NSB result.

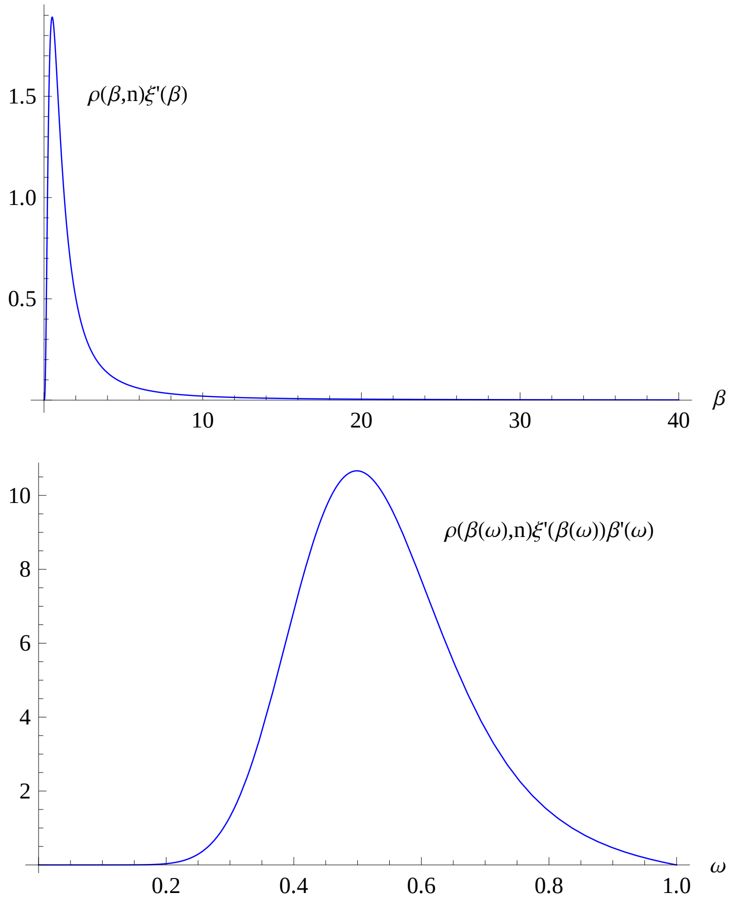


Figure 5.3: The substitution (5.40) maps the domain of the integrals in (5.36) to the interval $[0, 1)$ and broadens the integrand, thus facilitating easier numerical integration. The underlying toy model for the shown curves consists of $K = 9$ possibilities (or bins) and observations $\boldsymbol{n} = (4, 2, 3, 0, 2, 4, 0, 0, 2)$. The evidence $\rho(\beta, \boldsymbol{n})$ is the rescaled version in (5.38).

Chapter 6

Neuronal Network Interactions

In this chapter, we apply the information theory concepts described in chapter 5 to the experimental spike timings of a neuronal population. We ask whether pairwise interactions are sufficient to describe the spiking statistics (in a similar vein as Schneidman et al. (2006), but for *spontaneous* spiking activity in the absence of any visual stimulus). Finding an appropriate framework is important since it would greatly simplify the analysis of neuronal networks. Namely, the number of possible spiking patterns increases exponentially with the number of neurons N, such that the full probability distribution is typically undersampled in experiments. Since some patterns occur very rarely, a probability estimation based on the occurrence frequency might be biased and not reliable (see also section 5.6). On the other hand, if we knew that pairwise interactions suffice to capture the spiking statistics, we would only require the experimental firing rates and pairwise correlations to obtain a faithful spiking probability model of the full neuronal network.

Nevertheless, even if higher-order (≥ 3) interactions are weak, our analysis indicates that they might still play a role. For example, it is conceivable that third-order interactions result from the signal transmission between two neurons modulated by another neuron. Thus we investigate these interactions in more detail in section 6.3.

We illustrate that retinal ganglion cells can indeed be correlated by a concrete example: the curve in figure 6.1 shows the cross-correlogram of the spontaneous spiking activity of two retinal ganglion cells in a frog retina, in the absence of any visual stimulus. Thus all observed correlations must stem from internal neuronal interactions. The peaks at ±4 ms (see inset) likely result from gap junctions, i.e., direct connections of the cell membranes allowing ions to flow from one cell to the other. Another explanation could be a shared bipolar cell which excites both ganglion cells.

In the rest of the current chapter, we employ information theory mea-

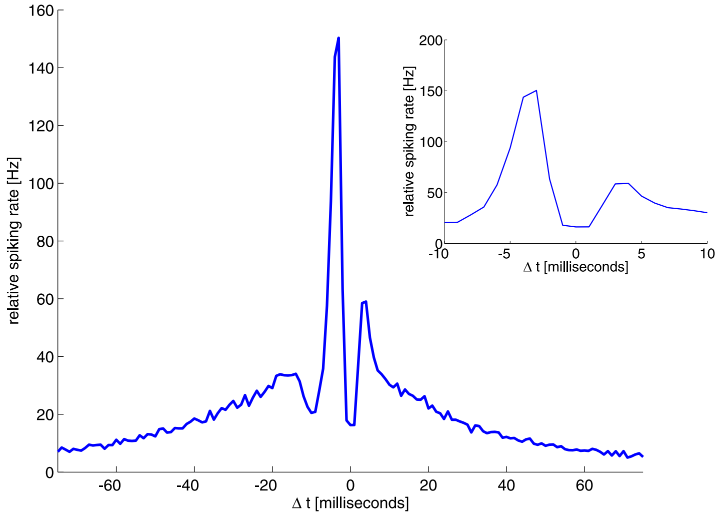


Figure 6.1: Cross-correlogram of the spontaneous spiking activity of two retinal ganglion cells. The curve shows the histogram of the relative spike timings (bin size 1 ms), i.e., the average spike rate of one cell relative to the spike timings of the other cell. Inset is a close-up of the time interval between ±10 ms. Note that the peaks are not symmetric.

sures (see chapter 5) to quantify the interaction strength of larger groups of neurons.

6.1 Experimental and Numerical Methods

We briefly outline the experimental and numerical methods specific for the present chapter.

Simultaneous recording of the joint neuronal activity of several ganglion cells is a prerequisite to capture neuronal network statistics. This is facilitated by the multi-electrode array (MEA) setup described in section 3.2. Recall that in our case the electrode array consists of 60 electrodes arranged on a regular 8×8 square grid (with the 4 corner electrodes missing), and 100 µm distance between the electrodes.

Visual stimulation should ideally be avoided completely (i.e., monitor switched off) since we are interested in the *intrinsic* neuronal interactions, not the spiking correlations induced by the visual stimulus. In other words, the neurons should fire spontaneously (as in figure 6.1 above). Indeed, we observe spontaneous activity in frog retinas, which we will analyze in the present chapter. This saves us from explicitly subtracting the stimulus-induced correlations (see equation (5.25) in section 5.5). In this respect, our analysis differs from the study by Schneidman et al. (2006), who presented visual stimuli imitating natural scenes to the retina.

In axolotl, however, we observe that the retinal ganglion cells remain silent without visual stimulation. This necessitates a visual stimulus which elicits spikes, but at the same time introduces hardly any correlations. In our case, we present a spatially uniform, linearly increasing brightness change from black to white, followed by a corresponding decrease from white to black, and then repeating in an infinite loop. In order to avoid stimulus-induced temporal correlations, both the increase and decrease between black and white are relatively slow (they take 1 s in our case).

Extraction of spiking patterns refers to the discretization the individual spike trains to arrive at a finite number of possible discrete patterns. In our case, we follow the binning strategy approach by Schneidman et al. (2006), as illustrated in figure 6.2. The current neuronal network activity is described by a binary word of length N, where N is the number of neurons. Bit i of this pattern is 1 if ganglion cell i fires (at least) one spike within the time window Δt_{bin}, and 0 if it remains silent. We set $\Delta t_{\text{bin}} = 10\,\text{ms}$, such that the cells' refractory period and relatively low spiking rate typically result in at most one spike within each time window.

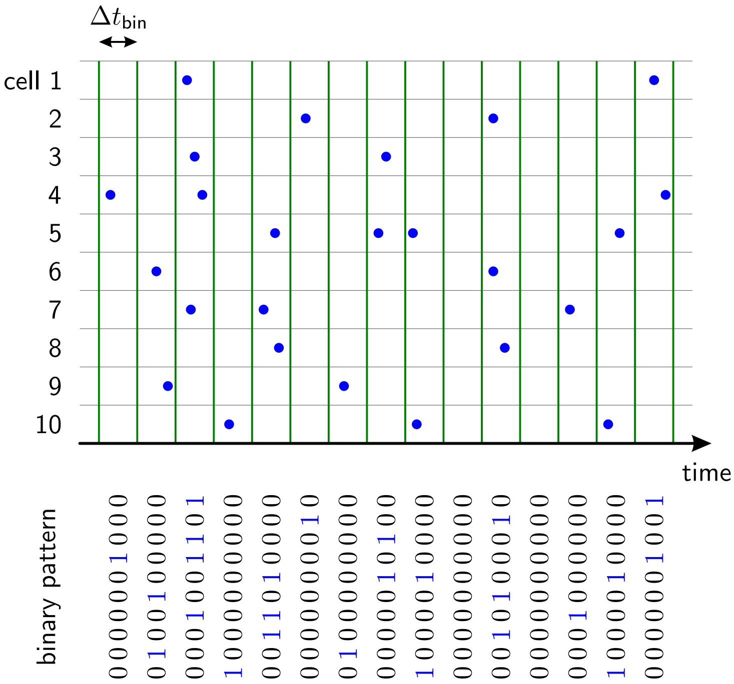


Figure 6.2: Discretization of spike trains into binary patterns, with bin size Δt_{bin} (schematic illustration). Each blue dot represents a single spike and is encoded as 1 in the corresponding binary word. In our case, the time window $\Delta t_{\text{bin}} = 10\,\text{ms}$.

Analysis of the spiking patterns starts from the experimental probability distribution $P_{\mathrm{exp}}(\sigma_1, \ldots, \sigma_N)$ derived from the observed number of occurrences of each binary pattern $(\sigma_1, \ldots, \sigma_N)$ with $\sigma_i \in \{0, 1\}$. Without bias correction (which we will treat below), the probability of each pattern is simply its occurrence frequency. In order to reflect the notation of the Ising model, we relabel $0 \to -1$ such that $\sigma_i \in \{-1, 1\}$.

Specifically for the Ising model, we can now use $P_{\mathrm{exp}}(\sigma_1, \ldots, \sigma_N)$ to calculate the experimental average values $\langle \sigma_i \rangle_{\mathrm{exp}}$ and the correlations among all pairs, $\langle \sigma_i \, \sigma_j \rangle_{\mathrm{exp}}$ for $1 \leq i < j \leq N$. Recall from chapter 5 that the Ising model probability distribution (5.11),

$$P_{\mathrm{Ising}}(\sigma_1, \ldots, \sigma_N) = \frac{1}{Z} \exp\left[\sum_i h_i \, \sigma_i + \sum_{i<j} J_{ij} \, \sigma_i \, \sigma_j \right] ,$$

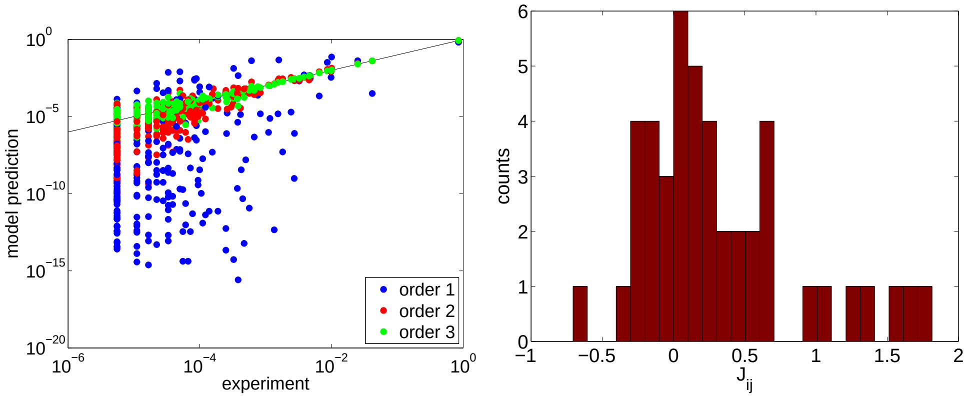
is precisely the maximum entropy distribution (5.8) which reproduces all $\langle \sigma_i \rangle_{\mathrm{exp}}$ and $\langle \sigma_i \, \sigma_j \rangle_{\mathrm{exp}}$, after we have calculated the appropriate Lagrange multipliers h_i and J_{ij} (see below). The third-order model (5.12) is analogous but additionally includes the experimental triple correlations $\langle \sigma_i \, \sigma_j \, \sigma_m \rangle_{\mathrm{exp}}$.

Numeric calculation of maximum entropy probability distributions is intricate in our case due to the large number of to-be fitted Lagrange multipliers involved. For example, for the distribution which reproduces the experimental correlations up to order k (Ising model for $k = 2$), the number of Lagrange multipliers equals $\sum_{i=1}^{k} \binom{N}{i}$ (recovering equation (5.13) for $k = N$). In our case ($N = 10$), this results in 10, 55 and 175 parameters for $k = 1, 2, 3$, respectively.

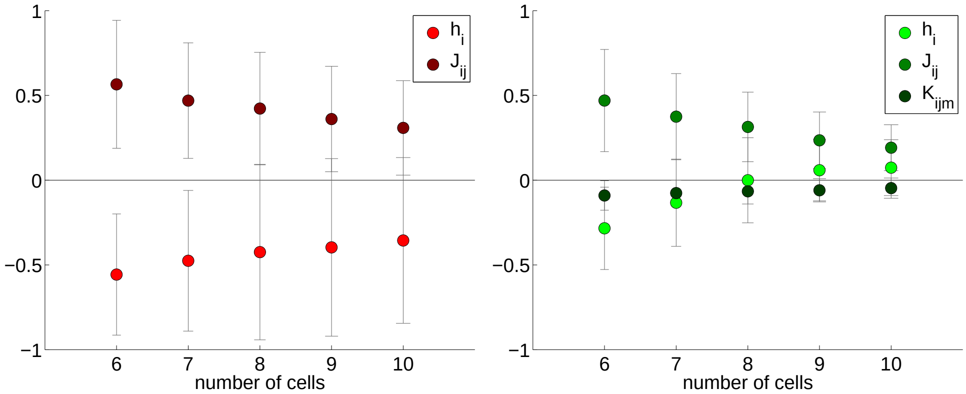
We employ the iterative algorithm devised by Darroch and Ratcliff (1972) to find the required Lagrange multipliers, both for the Ising model (section 5.3) and the marginal distribution (section 5.4). This algorithm is appropriate for populations of up to $N = 10$ neurons, and thus sufficient for our needs. For even larger populations ($N \approx 40$ neurons), the exponential increase in the number of spike patterns (2^N) would necessitate an alternative approach, like Monte Carlo algorithms for example (Broderick et al. 2007).

6.2 Ising Models Fitted to Spontaneous Network Activity

We apply the theoretical framework from section 5.3 to the simultaneous spiking activity of retinal ganglion cells. Additional to the second-order (Ising) model, we also fit the extended third-order model to the experimental data.



(a) pattern probabilities ($N = 10$ cells) **(b)** histogram of J_{ij} ($N = 10$ cells)



(c) order 2 model **(d)** order 3 model

Figure 6.3: Maximum entropy models matching the observed correlations up to orders 1, 2 and 3, for $N = 10$ cells in a frog retina. The order 2 model is precisely the Ising model (5.11), and the order 3 model corresponds to the distribution (5.12). **(a)** Each dot represents one of the $2^N = 1024$ binary spiking patterns, where the x-axis shows the experimentally observed pattern probability and the y-axis the model prediction. Note the logarithmic scale of both axes. The black identity line corresponds to exact agreement between model and experiment. **(b)** Histogram plot of the pairwise interaction coefficients J_{ij} of the Ising model (order 2). **(c)**, **(d)** Mean values of the Lagrange multipliers h_i, J_{ij} (and K_{ijm}) of the maximum entropy models of order 2 and 3, respectively. The smaller networks are subsets of the largest set with 10 cells.

Figure 6.3 illustrates the maximum entropy models which match the observed correlations up to orders 1, 2 (Ising) and 3, respectively. Specifically, figure 6.3a compares the experimental pattern probabilities $P_{\text{exp}}(\sigma_1, \ldots, \sigma_N)$ to the model predictions. Each dot represents one of the 2^N binary spiking patterns $(\sigma_1, \ldots, \sigma_N)$, with the experimental probability as x-value and the model prediction as y-value. Order 1 (blue dots) means that the cells are assumed to be independent, which leads to significant deviations from experiment, similar to what Schneidman et al. (2006) find. Namely, the blue dots show large deviations between experimental pattern probabilities and the model prediction.

Recall from section 5.3 the physical interpretation of the Ising model: the h_i are the local "magnetic fields" and the J_{ij} the pairwise interactions. The histogram in figure 6.3b shows the distribution of these interactions J_{ij}. We observe some "strong interaction" outsiders around 1.5. These could result from ganglion cells which are coupled via gap junctions, for example. Interestingly, there are both positive and negative J_{ij}, which could lead to "frustrated" cell triplets: three neurons (denoted A, B and C) with strong positive interactions between A and B as well as between B and C, but negative interactions between C and A. On the one hand, A directly inhibits C, but on the other hand A indirectly drives C via B, such that no obvious activity pattern exists (e.g., all cells prefer spiking or all cells remain silent).

Figures 6.3c and 6.3d explicitly show the mean values of the Lagrange multipliers h_i, J_{ij} (and K_{ijm}) for the order 2 and 3 models, respectively, as well as for networks of various sizes. Starting from the largest network with $N = 10$ cells, we have successively removed one cell to arrive at the smaller networks. The average local fields $\langle h_i \rangle$ in figure 6.3c are negative, i.e., the cells tend to remain silent. Interestingly, the average interaction strength $\langle J_{ij} \rangle$ slightly decreases as the number of cells increases, but does not reach zero for $N = 10$ cells. These observations are plausible since the typical distance between cells grows with the network size, and in realistic physical systems the interactions decay algebraically with distance (e.g., Coulomb interactions) or even exponentially (due to screening effects). Unfortunately, we do not have the exact positions and mutual distances of the ganglion cells available and the network size is relatively small, such that we can hardly discriminate between exponential or algebraic decay. Different from our finding, $\langle J_{ij} \rangle$ retains a constant positive value in the study by Schneidman et al. (2006), and the authors conclude that the interactions play an even larger role for larger networks. The average of the third-order interaction coefficients K_{ijm} in figure 6.3d is negative and has a small amplitude; we will investigate these coefficients in more detail in the following section 6.3.

Concerning the question whether pairwise interactions suffice to describe the population statistics, we calculate the *connected information* (refer to equation (5.19) in section 5.4) between the independent model (order 1) and the Ising model (order 2), as well as between the Ising model and the order 3 model:

$$\begin{aligned} I_C^{(2)} &= H(P^{(1)}) - H(P_{\text{Ising}}) \\ I_C^{(3)} &= H(P_{\text{Ising}}) - H(P^{(3)}). \end{aligned}$$

Recall that the Ising model is equivalent to the maximum entropy model reproducing all second-order marginals (since we have *binary* variables), and analogously for $P^{(1)}$ and $P^{(3)}$. In our case, the independent model equals the product of the individual marginal distributions $P_i(\sigma_i)$,

$$P^{(1)}(\sigma_1, \ldots, \sigma_N) \equiv \frac{1}{Z} \exp\left[\sum_i h_i\, \sigma_i\right] = \prod_i P_i(\sigma_i),$$

since each binary probability distribution $P_i(\sigma_i)$ is fully specified by its expectation value $\langle \sigma_i \rangle$. In particular, $H(P^{(1)}) = \sum_i H(P_i)$.

According to equation (5.18) from section 5.4, the quantities $I_C^{(k)}$ range between zero (complete independence) and the multi-information I_{m} defined in (5.15). In the current notation, $I_{\mathrm{m}} = H(P^{(1)}) - H(P)$, where P denotes the full (experimental) probability distribution. Indeed, we obtain for the experimental data in figure 6.3 (without any bias correction) that

$$\begin{aligned} I_{\mathrm{m}} &= 0.870\,\mathrm{bits}, \\ I_C^{(2)} &= 0.837\,\mathrm{bits} \quad \text{and thus} \quad I_C^{(2)}/I_{\mathrm{m}} = 0.96, \\ I_C^{(3)} &= 0.011\,\mathrm{bits} \quad \text{and thus} \quad I_C^{(3)}/I_{\mathrm{m}} = 0.01. \end{aligned} \tag{6.1}$$

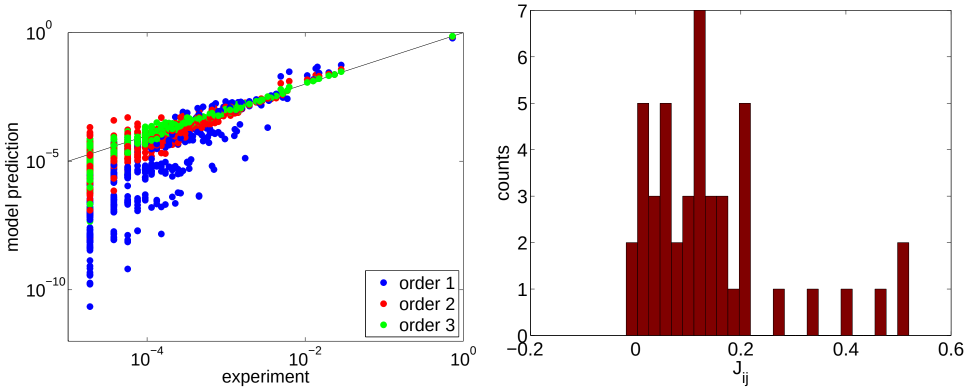
In other words, the Ising model captures 96% of the "connected information" in the network, according to the data. Thus, the pairwise interactions (instead of triple or even higher ones) seem to play the dominant role.

To estimate the uncertainty of I_{m} in (6.1) and correct for the bias due to finite sampling, we make use of the NSB entropy estimation from section 5.6. The Matlab/Octave implementation by the authors (Nemenman 2011) additionally calculates an error estimate of the entropy. Using this software, we obtain

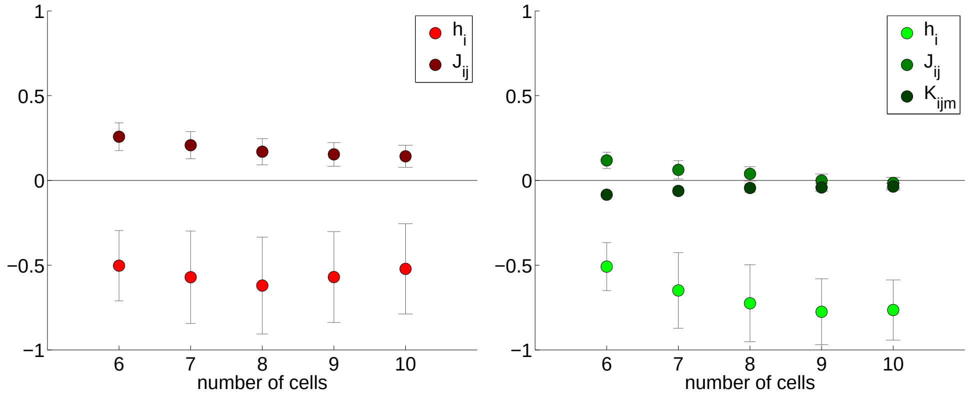
$$\begin{aligned} H(P) &= 1.322 \pm 0.006\,\mathrm{bits} \\ H(P^{(1)}) &= 2.190 \pm 0.02\,\mathrm{bits}, \quad \text{and thus} \\ I_{\mathrm{m}} = H(P^{(1)}) - H(P) &= 0.867 \pm 0.025\,\mathrm{bits}, \end{aligned}$$

which closely matches the (uncorrected) value from (6.1). Our own Mathematica implementation of the NSB algorithm precisely agrees with the above value for $H(P)$ in the first 8 digits.

So far our discussion of figure 6.3 was based on ganglion cell activity in a frog retina. For comparison, figure 6.4 shows the analogous analysis but for the spiking activity in an *axolotl* retina. Remember that we employ a spatially uniform, linearly increasing/decreasing brightness change as visual stimulus to elicit spiking activity. The averages of the interactions J_{ij} and K_{ijm} in subfigures 6.4c and 6.4d qualitatively resemble the interactions in figure 6.3 in the sense that their amplitude decreases with the



(a) pattern probabilities ($N = 10$ cells) **(b)** histogram of J_{ij} ($N = 10$ cells)



(c) order 2 model **(d)** order 3 model

Figure 6.4: Maximum entropy models matching the observed correlations up to orders 1, 2 and 3, for $N = 10$ cells in an *axolotl* retina. This is analogous to figure 6.3, except that we employ a spatially uniform, linearly increasing or decreasing brightness change to elicit spiking activity. **(a)** Binary spiking patterns, comparing the experimental probabilities with the model prediction. **(b)** Histogram plot of the pairwise interaction coefficients J_{ij} of the Ising model (order 2), showing that almost all J_{ij} are non-negative. **(c)**, **(d)** Mean values of the Lagrange multipliers h_i, J_{ij} (and K_{ijm}) of the maximum entropy models of order 2 and 3, respectively. The amplitude of the average interactions J_{ij} and K_{ijm} decreases with the network size N, similar to figure 6.3, and $\langle J_{ij} \rangle$ even reaches zero.

network size N, and $\langle J_{ij} \rangle$ is mostly positive whereas $\langle K_{ijm} \rangle$ is negative with smaller amplitude. The corresponding histogram in figure 6.4b shows several "strong interaction" outliers J_{ij} in the second-order model. However, $\langle J_{ij} \rangle$ changes notably in the third-order model and reaches zero, different from figure 6.4c. This could point to the conclusion that several positive interaction coefficients J_{ij} in the *second*-order model actually result from higher-order interactions, which the second-order model cannot explicitly account for. Note that individual third-order coefficients can be strong even though the average $\langle K_{ijm} \rangle$ is small. The connected information for the data in figure 6.4 (analogous to equation (6.1)) equals

$$\begin{aligned} I_{\mathrm{m}} &= 0.269\,\text{bits}, \\ I_C^{(2)} &= 0.223\,\text{bits} \quad \text{and thus} \quad I_C^{(2)}/I_{\mathrm{m}} = 0.83, \\ I_C^{(3)} &= 0.033\,\text{bits} \quad \text{and thus} \quad I_C^{(3)}/I_{\mathrm{m}} = 0.12. \end{aligned} \tag{6.2}$$

Interestingly, the connected information $I_C^{(3)}$ between the order 3 model and the Ising model is now much higher as compared to (6.1), which confirms that the triple interactions K_{ijm} become more important.

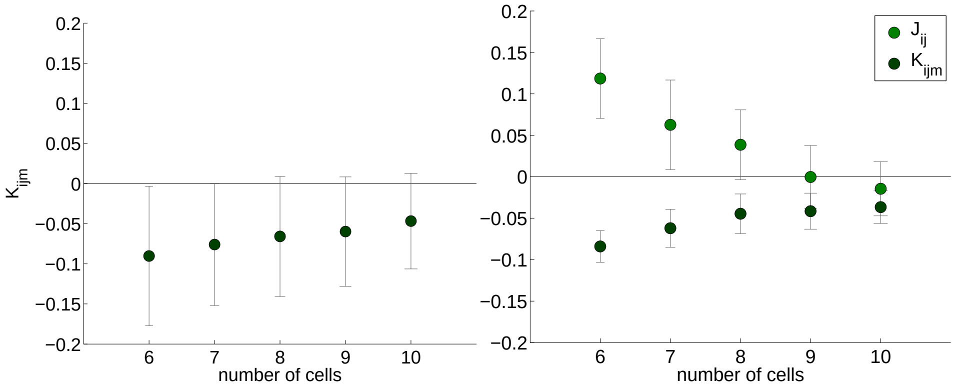
Discussion Our experimental data in figure 6.3 (frog) confirms the main finding by Schneidman et al. (2006), namely that pairwise interactions are largely sufficient for reproducing the network statistics. Importantly, we can exclude any stimulus-induced correlations in frog retinas since we measure spontaneous activity, whereas Schneidman et al. (2006) do not distinguish between stimulus-induced and intrinsic correlations.

Based on theoretical considerations, the authors report that the interactions play an increasing role with increasing network size N if the average $\langle J_{ij} \rangle$ retains a positive constant value, which they indeed observe experimentally. In contrast to that, our experimental data shows a decrease of $\langle J_{ij} \rangle$ in figure 6.3c. A plausible explanation could be locality, i.e., pairwise interactions are constrained to nearby cells, such that the majority of possible interactions is negligible (J_{ij} close to zero) since the cells are too far apart. As the percentage of these negligible interactions grows with N, the average value decreases.

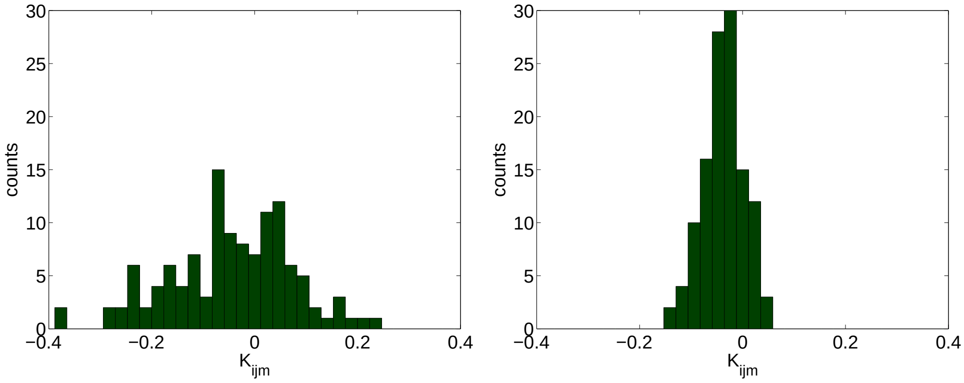
Last but not least, according to the experimental data in figure 6.4 (axolotl) and equation (6.2), third-order interactions might be more important than previously thought. This is substantiated by the speculative conjecture that strong pairwise interactions J_{ij} in the second-order model might actually be due to higher-order interactions (compare figures 6.4c and 6.4d).

6.3 Third-Order Interactions

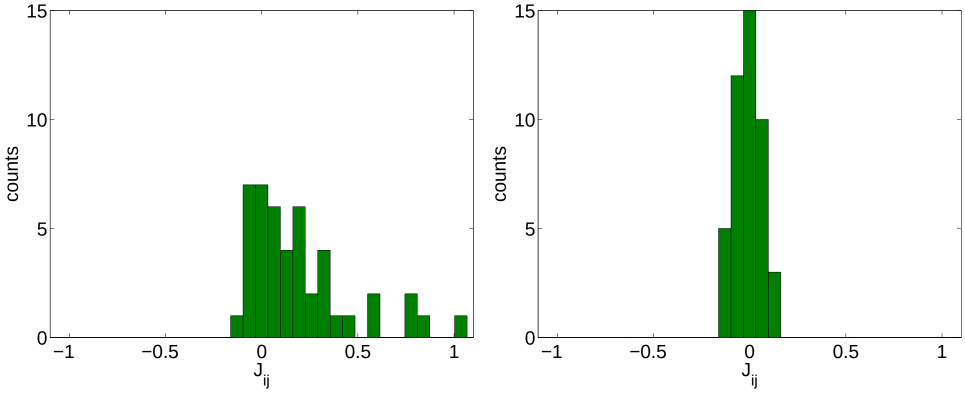
In what follows, we investigate the triple interaction coefficients K_{ijm} in more detail, which have been omitted in the study by Schneidman et al. (2006). We have seen in equations (6.1) and (6.2) that the connected information $I_C^{(3)}$ is much smaller than $I_C^{(2)}$ for the experimental data in figures 6.3 and 6.4, but might nevertheless be worth studying. A biological scenario for a third-order interaction between three cell types in the retina might be an amacrine cell which modulates the synaptic signaling between a bipolar cell and a ganglion cell (Manu and Baccus 2011).



(a) zoom-in of figure 6.3d (frog) **(b)** zoom-in of figure 6.4d (axolotl)



(c) histogram for 10 cells, figure 6.3 (frog) **(d)** histogram for 10 cells, figure 6.4 (axolotl)



(e) histogram for 10 cells, figure 6.3 (frog) **(f)** histogram for 10 cells, figure 6.4 (axolotl)

Figure 6.5: Interaction coefficients of the third-order model, for the data in figures 6.3 (frog, left side) and 6.4 (axolotl, right side). Top row: average and variance of K_{ijm} (and J_{ij}) for various network sizes N, showing an amplitude decrease with growing N. Middle row: histograms of K_{ijm}. Bottom row: histograms of J_{ij}. Note the different scale of the x-axis as compared to the middle row.

Figure 6.5 illustrates the pairwise and triple interaction coefficients of the *third*-order models for the experimental data in figures 6.3 (frog) and 6.4 (axolotl). We observe that the average of the coefficients K_{ijm} is negative and has a small amplitude, which decreases as the network size N increases. Interestingly, there are K_{ijm} of either sign in both retinas, according to the histograms 6.5c and 6.5d.

Discussion In the axolotl retina, the pairwise interactions J_{ij} of the third-order model (figure 6.5f) are centered around zero and much weaker as compared to the second order model (figure 6.4b). This is in line with our discussion in section 6.2, namely, that the strong interaction coefficients J_{ij} in the second-order model actually trace back to third-order interactions (which the second-order model cannot account for directly).

We have observed that the amplitude of $\langle K_{ijm} \rangle$ decreases when N grows (see figures 6.5a and 6.5b). As discussed in the previous section, a plausible explanation could be locality. In this case, the percentage of negligible K_{ijm} close to zero is higher as compared to the pairwise interactions J_{ij}, since for triple interactions, all three cells must be close together.

Our experimental data shows that the triple interaction coefficients K_{ijm} are biased towards negative values, which could indicate inhibitory modulations in the retina. However, given that third-order interactions actually exist, we can only speculate about the underlying biological mechanisms. Presumably, these should involve amacrine cells or bipolar cells since the axons of retinal ganglion cells only reach higher brain areas, but cannot transmit signals to any neurons in the retina.

Chapter 7

Conclusions and Outlook

In this thesis, we have aimed at understanding how an abundant feature of vision (saccadic eye movements) influences and translates to neuronal coding strategies in the retina. Specifically, we have investigated the spiking responses of individual retinal ganglion cells to a visual grating image, which shifts due to simulated eye movements (chapter 4). Most importantly, we have found that the latency and spike count response of some ganglion cells forms a sinusoidal tuning curve (as illustrated in figure 4.6, for example) dependent on the offset of the presented image (figure 4.3). This allows us to describe the response by a single number, the tuning phase φ. Analogously, the activation of the cell's receptive field by the grating stimulus also has a sine-shaped form, according to our mathematical derivation in section 4.3. Thus we can directly compare these phases as illustrated in figure 4.13. In other words, we have found a simple way to relate the stimulus to the response. The experimental data indeed confirms that the receptive field position can approximately predict the response tuning (figures 4.14 and 4.15 for the latency and spike count, respectively). In the following three sections 4.5, 4.6 and 4.7, we have then refined this model and tried to understand the responses in terms of the time-dependent activation of the ganglion cells during the saccadic image shift. Finally, we have employed pharmacology to study the effects of selectively blocking the signal transmission from photoreceptors to ON bipolar cells (section 4.8.1), or blocking inhibition in the retina (section 4.8.2).

In the second experimental part of this thesis, we have quantified the pairwise and third-order interaction strengths in a retinal ganglion cell population (chapter 6). The multi-electrode array setup is well suited for this tasks since we obtain the *simultaneous* spiking signals. Interestingly, we have found that third-order interactions might be more important than previously thought (section 6.3).

In the following two sections, we briefly review the experimental findings from chapters 4 and 6 in more detail and discuss several open questions.

Neuronal Coding in the Presence of Saccades

We have shown in section 4.2 that saccadic image shifts can indeed elicit neuronal spiking responses. We have compared these responses for different offsets of the visual image, in order to find features which contain information about the stimulus (i.e., the vertical grating position in our case). Two salient candidates are the latency and the spike count (see figure 4.4), which both show a sinusoidal tuning curve dependent on the image offset (figures 4.6 and 4.7).

According to our theoretical calculations in section 4.3, the activation of a Gaussian-shaped receptive field by a grating stimulus can be well approximated by a sine function (4.6) which only depends on the center position of the receptive field, as long as the receptive field size is at least as large as the grating width. This allows us to associate a phase value with the receptive field position on the grating, analogous to the response tuning phase (compare with figure 4.13). We have found that the receptive field position can indeed approximately predict the tuning phase of the latency or spike count response (figures 4.14 and 4.15 in section 4.4).

As next step, we have tried to narrow down the precise stimulus features which determine the response tuning. First, we have directly compared two stimulus scenarios which should lead to exactly the same results if the response was dependent on the image before or after the saccadic shift only. Thus, any experimental deviations must result from the time-dependent stimulus during the shift. We have indeed found small deviations in experiments (distance from dashed identity line in figure 4.18, section 4.5). A heuristic explanation could be the timing of the grating border crossing by the receptive field (see figure 4.19). Next, we have investigated the question whether retinal ganglion cells actually report the differential image change to higher brain regions, or the new image (independent of the previous one), or even the previous image. The average of the experimental data points in figures 4.23 and 4.24 of section 4.6 approximately agrees with a model (figure 4.22) maximizing the differential activation change. In any case, we can match the data points corresponding to individual ganglion cells by weighting the relative importance of the previous and new image (figure 4.21).

Next, we have specifically investigated the effect of the saccadic shift amplitude (section 4.7). For the particular shift amplitudes, the experimental latency data (figure 4.28) shows that the activation difference between previous and new image cannot solely account for the observed responses. Nevertheless, taking the transient activation during the beginning of the saccadic shifts into account (figure 4.30) can qualitatively explain the observed latency tuning responses.

Using pharmacology, we have studied the effects of selectively blocking the signal transmission from photoreceptors to ON bipolar cells via APB (section 4.8.1), or blocking inhibition in the retina (section 4.8.2). Surprisingly, the OFF pathway in the retina suffices to generate latency and spike count tuning curves (see for example figure 4.32). Moreover, the reliability of the receptive field position on the grating as predictor of the response latency tuning remains large unaffected by either blocking the ON pathway (figure 4.33) or blocking inhibition (figure 4.37), as compared to the control condition without pharmacology (figure 4.14).

Open questions We have seen in section 4.2 that only a subset of ganglion cells shows latency responses appropriate for defining a tuning curve. So the question remains whether the noise introduced by the MEA setup and the experimental in-vitro conditions change the apparent properties of the other ganglion cells, whether they preferentially respond to different stimuli, or whether they indeed employ a different code. We have investigated several coding schemes in section 4.2, but our search might not have been exhaustive. Thus one could look for a visual stimulus which particularly drives these ganglion cells as a future project, or search for alternative neuronal coding strategies. For example, we have investigated the intrinsic interactions of ganglion cells in chapter 6, but have not yet studied how these interactions contribute to population coding in the context of saccadic eye movements.

On the other hand, it would be desirable to further characterize the subset of ganglion cells which actually show latency tuning responses. Ideally, these ganglion cells could be distinguished by their morphology or by genetic tools, analogous to the study by Kay et al. (2011).

It is still open whether the latency tuning curves also exist in mammalian retina. For this purpose, we could repeat the experiments with mouse retina (recordings from mouse retina are indeed possible and have already been performed with our particular experimental setup). Mammalian systems typically operate faster then amphibian neuronal networks, so an interesting aspect are the (latency) timescales.

We have purposely used a rather simple spatial grating stimulus such that the overall brightness level and the spatial frequency distribution remain constant. In contrast to that, employing “natural stimuli” (i.e., movies of natural scenes) complicates the analysis which stimulus features are relevant for the retinal responses. Nevertheless, one eventually wants to understand how the retina actually functions under “real life” conditions.

Neuronal networks

We have performed multi-electrode recordings of several retinal ganglion cells to investigate the simultaneous network activity of a neuronal population (chapter 6). Our analysis procedure follows the approach by Schneidman et al. (2006), asking whether Ising models from theoretical physics can reproduce the experimental spike pattern probabilities. Since we are interested in the biological wiring of the network, we deliberately try to record the *spontaneous* spiking activity in the absence of any visual stimulus, to ensure that the observed correlations indeed stem from intrinsic interactions. This is different from the study by Schneidman et al. (2006), who present "natural movies" to the retina.

We have first investigated the activity in a frog retina (figure 6.3), where our results point to the same conclusion as in the study by Schneidman et al. (2006), namely, that the *pairwise* interactions captured by the Ising model are largely sufficient to reproduce the experimental spiking statistics. Quantitatively, the connected information (see section 5.4) between the Ising model and the independent model is 96% of the full "multi-information", according to equation (6.1).

Next, we have analyzed the spiking activity in an axolotl retina (figure 6.4), arriving at a smaller value for the connected information, namely 83% of the multi-information (see equation (6.2)). To understand where the remaining 17% come from, we start from the canonical extension of the Ising model to a third-order model. This model additionally respects all experimental triple correlations $\langle \sigma_i \, \sigma_j \, \sigma_m \rangle$, whereas the Ising model is equivalent to the maximum entropy model reproducing the averages $\langle \sigma_i \rangle$ and pairwise correlations $\langle \sigma_i \, \sigma_j \rangle$ (refer to the review in section 5.3). Indeed, the third-order model contributes 12% to the "connected information" according to our experimental data. Investigation of the interaction coefficients in this model (figure 6.5) reveals that the pairwise interactions J_{ij} become much weaker as compared to the Ising model. As speculative explanation of this effect, the strong interaction coefficients J_{ij} in the Ising model might actually stem from third-order interactions which the Ising model cannot represent explicitly. Thus, we conclude that third- or even higher-order interactions might actually be more important than previously thought.

Open questions We can only speculate why third-order interactions seem to be more important in the axolotl retina than compared to the spontaneous activity in the frog retina. This effect could either be due to the different species, or to the fact that we have used a linearly increasing or decreasing brightness change to elicit spiking in the axolotl retina. In any case, since we have only analyzed two experiments so far, a clarification requires repeating these experiments to see if the effect is indeed systematic.

Our analysis of neuronal population statistics in chapter 6 is based on the network state at the current time window, but does not explicitly take *temporal* interactions into account. Tang et al. (2008); Marre et al. (2009) try to address this issue by explicitly including succeeding time windows into the model, and can indeed show that the model reproduces the experimental data better than the Ising model. Thus, a goal for the future is such an extension of the Ising model. Similarly, one could investigate different time scales of interactions by employing various time bins Δt_{bin} (see figure 6.2).

Outlook

In chapter 4 we have chosen the visual stimulation parameters for imitating saccadic eye movements based on the requirement that they elicit ample spiking responses. However, we have not yet experimentally quantified saccadic eye movements *in vivo*, which is left as project for the future. Such a quantification could be performed by a search coil setup, which uses a miniature contact lens with a tiny electric coil inside. The contact lens is attached to the axolotl's eye, similar to a contact lens for humans. When we then position the animal into an external magnetic field, the eye movements induce a small voltage change in the coil and thus allow us to measure eye movement speed and amplitude.

A further refinement of our analysis in chapter 6 is a detailed investigation how the neuronal interactions depend on the physical distance between ganglion cells. This distance could be computed approximately from the receptive field centers, analogous to the study by Ganmor et al. (2011a). Such an analysis would simplify the interpretation in terms of physical models, which quantitatively depend on the decay of the interaction strength with distance. Additionally, we could use an experimental setup with more electrodes in order to quantify the interactions in larger populations. Due to the exponential grows of possible patterns with the number of neurons, this also necessitates different numerical approaches (like Monte-Carlo simulations, mentioned in section 6.2) to compute the maximum entropy models.

Neuronal coding strategies are a central ingredient of the inner workings of neuronal networks, but still remain far from comprehensively understood. The contribution of this thesis is a detailed investigation of latency and spike count coding schemes in the context of saccadic eye movements, which are an abundant feature of vision. In the long term, we hope that unveiling such schemes might contribute to the operation of electronic retinal implants (Zrenner et al. 2011; Ahuja et al. 2011), which could partially restore vision in patients with degenerative eye conditions.

Bibliography

E. D. Adrian. The impulses produced by sensory nerve endings. Part 1. *Journal of Physiology*, 61:49–72, 1926.

E. D. Adrian and Yngve Zotterman. The impulses produced by sensory nerve-endings. Part 2. The response of a Single End-Organ. *Journal of Physiology*, 61:151–171, 1926.

Ehud Ahissar and Amos Arieli. Figuring space by time. *Neuron*, 32:185–201, 2001.

A K Ahuja, J D Dorn, A Caspi, M J McMahon, G Dagnelie, L daCruz, P Stanga, M S Humayun, R J Greenberg, and Argus II Study Group. Blind subjects implanted with the Argus II retinal prosthesis are able to improve performance in a spatial-motor task. *British Journal of Ophthalmology*, 95:539–543, 2011.

Stephen A. Baccus, Bence P. Ölveczky, Mihai Manu, and Markus Meister. A Retinal Circuit That Computes Object Motion. *Journal of Neuroscience*, 28(27):6807–6817, 2008.

Horace B. Barlow and Richard M. Hill. Selective Sensitivity to Direction of Movement in Ganglion Cells of the Rabbit Retina. *Science*, 139 (3553): 412, 1963.

Horace B. Barlow and William R. Levick. The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology*, 178(3):477–504, 1965.

Horace B. Barlow, Richard M. Hill, and William R. Levick. Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *Journal of Physiology*, 173(3):377–407, 1964.

Michael J. Berry and Markus Meister. Refractoriness and Neural Precision. *Journal of Neuroscience*, 18(6):2200–2211, 1998.

Michael J. Berry, David K. Warland, and Markus Meister. The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences USA*, 94:5411–5416, 1997.

Michael J. Berry, Iman H. Brivanlou, Thomas A. Jordan, and Markus Meister. Anticipation of moving stimuli by the retina. *Nature*, 398:334–338, 1999.

Iman H. Brivanlou, David K. Warland, and Markus Meister. Mechanisms of Concerted Firing among Retinal Ganglion Cells. *Neuron*, 20:527–539, 1998.

Tamara Broderick, Miroslav Dudik, Gasper Tkacik, Robert E. Schapire, and William Bialek. Faster solutions of the inverse pairwise Ising problem. *arXiv*, 0712.2437:1–8, 2007.

Yoram Burak, Uri Rokni, Markus Meister, and Haim Sompolinsky. Bayesian model of dynamic image stabilization in the visual system. *Proceedings of the National Academy of Sciences*, 107:19525–19530, 2010.

Dwight A. Burkhardt, Patrick K. Fahey, and Michael Sikora. Responses of ganglion cells to contrast steps in the light-adapted retina of the tiger salamander. *Visual Neuroscience*, 15:219–229, 1998.

David C. Burr, M. Concetta Morrone, and John Ross. Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature*, 371:511–513, 1994.

Santiago Ramón y Cajal. *Die Retina der Wirbelthiere: Untersuchungen mit der Golgi-Cajal'schen Chromsilbermethode und der Ehrlich'schen Methylenblaufärbung*. Bergmann, 1894.

E. J. Chichilnisky. A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, 12:199–213, 2001.

P. Jeffrey Conn and Jean-Philippe Pin. Pharmacology and functions of metabotropic glutamate receptors. *Annual Review of Pharmacology and Toxicology*, 37:205–237, 1997.

J. N. Darroch and D. Ratcliff. Generalized Iterative Scaling for Log-Linear Models. *Annals of Mathematical Statistics*, 43(5):1470–1480, 1972. ISSN 00034851.

Peter Dayan and Laurence F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2005.

Jonathan B. Demb, Kareem Zaghloul, Loren Haarsma, and Peter Sterling. Bipolar Cells Contribute to Nonlinear Spatial Summation in the Brisk-Transient (Y) Ganglion Cell in Mammalian Retina. *Journal of Neuroscience*, 21(19):7447–7454, 2001.

R.W. Ditchburn. The function of small saccades. *Vision Research*, 20(3): 271–272, 1980. ISSN 0042-6989.

G.D. Field and E.J. Chichilnisky. Information Processing in the Primate Retina: Circuitry and Coding. *Annual Review of Neuroscience*, 30:1–30, 2007.

Nathan Fitzsimmons, Mikhail Lebedev, Ian Peikon, and Miguel A. L. Nicolelis. Extracting kinematic parameters for monkey bipedal walking from cortical neuronal ensemble activity. *Frontiers in Integrative Neuroscience*, 3:1–19, 2009.

Elad Ganmor, Ronen Segev, and Elad Schneidman. How fast can we learn maximum entropy models of neural populations? *Journal of Physics: Conference Series*, 197:012020, 2009.

Elad Ganmor, Ronen Segev, and Elad Schneidman. The Architecture of Functional Interaction Networks in the Retina. *Journal of Neuroscience*, 31(8):3044–3054, 2011a.

Elad Ganmor, Ronen Segev, and Elad Schneidman. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *Proceedings of the National Academy of Sciences USA*, 108: 9679–9684, 2011b.

Tim Gollisch and Markus Meister. Rapid neural coding in the retina with relative spike latencies. *Science*, 319(5866):1108–1111, 2008.

Tim Gollisch and Markus Meister. Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina. *Neuron*, 65:150–164, 2010.

Martin Greschner, Markus Bongard, Pal Rujan, and Josef Ammermüller. Retinal ganglion cell synchronization by fixational eye movements improves feature estimation. *Nature Neuroscience*, 5:341–347, 2002.

John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences USA*, 79:2554–2558, 1982.

Tomomi Ichinose and Peter D. Lukasiewicz. Inner and outer retinal pathways both contribute to surround inhibition of salamander ganglion cells. *Journal of Physiology*, 565:517–535, 2005.

Ernst Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik A*, 31:253–258, 1924.

Edwin Thompson Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620–630, May 1957a.

Edwin Thompson Jaynes. Information Theory and Statistical Mechanics. II. *Physical Review*, 108(2):171–190, Oct 1957b.

Jeremy N. Kay, Irina De la Huerta, In-Jung Kim, Yifeng Zhang, Masahito Yamagata, Monica W. Chu, Markus Meister, and Joshua R. Sanes. Retinal Ganglion Cells with Distinct Directional Preferences Differ in Molecular Identity, Structure, and Central Projections. *Journal of Neuroscience*, 31(21):7753–7762, 2011.

Justin Keat, Pamela Reinagel, R. Clay Reid, and Markus Meister. Predicting every spike: a model for the responses of visual neurons. *Neuron*, 30: 803–817, 2001.

In-Jung Kim, Yifeng Zhang, Masahito Yamagata, Markus Meister, and Joshua R. Sanes. Molecular identification of a retinal cell type that responds to upward motion. *Nature*, 452:478–482, 2008.

Hee-kyoung Ko, Martina Poletti, and Michele Rucci. Microsaccades precisely relocate gaze in a high visual acuity task. *Nature Neuroscience*, 13: 1549–1553, 2010.

M. F. Land. Motion and vision: why animals move their eyes. *Journal of Comparative Physiology A*, 185:341–352, 1999. ISSN 0340-7594.

Peter E. Latham and Sheila Nirenberg. Synergy, Redundancy, and Independence in Population Codes, Revisited. *Journal of Neuroscience*, 25(21): 5195–5206, 2005.

J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts. What the frog's eye tells the frog's brain. *Proceedings of the IRE*, 47:233–258, 1959.

Mihai Manu and Stephen A. Baccus. Disinhibitory gating of retinal output by transmission from an amacrine cell. *Proceedings of the National Academy of Sciences USA*, 108:18447–18452, 2011.

O. Marre, S. El Boustani, Y. Frégnac, and A. Destexhe. Prediction of Spatiotemporal Patterns of Neural Activity from Pairwise Correlations. *Physical Review Letters*, 102:138101, 2009.

Susana Martinez-Conde and Stephen L. Macknik. Windows on the mind. *Scientific American*, 8:56–63, 2007.

Susana Martinez-Conde, Stephen L. Macknik, and David H. Hubel. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5:229–240, 2004.

Susana Martinez-Conde, Stephen L. Macknik, Xoana G. Troncoso, and Thomas A. Dyar. Microsaccades counteract visual fading during fixation. *Neuron*, 49:297–305, 2006.

Susana Martinez-Conde, Stephen L. Macknik, Xoana G. Troncoso, and David H. Hubel. Microsaccades: a neurophysiological analysis. *Trends in Neurosciences*, 32:463–475, 2009.

Richard H. Masland. The fundamental plan of the retina. *Nature Neuroscience*, 4(9):877–886, 2001.

Markus Meister and Michael J. Berry. The Neural Code of the Retina. *Neuron*, 22:435–450, 1999.

Thomas A Münch, Rava Azeredo da Silveira, Sandra Siegert, Tim James Viney, Gautam B Awatramani, and Botond Roska. Approach sensitivity in the retina processed by a multifunctional neural circuit. *Nature Neuroscience*, 12(10):1308–1318, 2009.

Yoshiaki Nakajima, Hideki Iwakabe, Chihiro Akazawa, Hiroyuki Nawa, Ryuichi Shigemoto, Noboru Mizuno, and Shigetada Nakanishi. Molecular characterization of a novel retinal metabotropic glutamate receptor mGluR6 with a high agonist selectivity for L-2-amino-4-phosphonobutyrate. *Journal of Biological Chemistry*, 268:11868–11873, 1993.

Ilya Nemenman. Inference of entropies of discrete random variables with unknown cardinalities. *arXiv:physics*, 0207009:1–8, 2002.

Ilya Nemenman. http://nsb-entropy.sourceforge.net/, 2011.

Ilya Nemenman, Fariel Shafee, and William Bialek. Entropy and Inference, Revisited. *arXiv:physics*, 0108025:1–9, 2002.

Ilya Nemenman, William Bialek, and Rob de Ruyter van Steveninck. Entropy and information in neural spike trains: Progress on the sampling problem. *Physical Review E*, 69:056111, May 2004.

Miguel A. L. Nicolelis and Mikhail A. Lebedev. Principles of neural ensemble physiology underlying the operation of brain-machine interfaces. *Nature Reviews Neuroscience*, 10:530–540, 2009.

Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

Sheila Nirenberg, S. M. Carcieri, A. L. Jacobs, and Peter E. Latham. Retinal ganglion cells act largely as independent encoders. *Nature*, 441:698–701, 2001.

Bence P. Ölveczky, Stephen A. Baccus, and Markus Meister. Segregation of object and background motion in the retina. *Nature*, 423:401–408, 2003.

Bence P. Ölveczky, Stephen A. Baccus, and Markus Meister. Retinal Adaptation to Object Motion. *Neuron*, 56:689–700, 2007.

Lars Onsager. Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition. *Physical Review*, 65:117–149, Feb 1944.

Jonathan W. Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M. Litke, E. J. Chichilnisky, and Eero P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454:995–1000, 2008.

Xaq Pitkow, Haim Sompolinsky, and Markus Meister. A Neural Computation for Visual Acuity in the Presence of Eye Movements. *PLoS Biology*, 5(12):e331, 2007.

Tomaso Poggio, Manfred Fahle, and Shimon Edelman. Fast perceptual learning in visual hyperacuity. *Science*, 256:1018–1021, 1992.

Christophe Pouzat, Ofer Mazor, and Gilles Laurent. Using noise signature to optimize spike-sorting and to assess neuronal classification quality. *Journal of Neuroscience Methods*, 122(1):43–57, 2002. ISSN 0165-0270.

Jason L. Puchalla, Elad Schneidman, Robert A. Harris, and Michael J. Berry. Redundancy in the Population Code of the Retina. *Neuron*, 46: 493–504, 2005.

D.S. Reich, F. Mechler, and J.D. Victor. Independent and redundant information in nearby cortical neurons. *Science*, 294:2566–2568, 2001.

Fred Rieke. Temporal Contrast Adaptation in Salamander Bipolar Cells. *Journal of Neuroscience*, 21(23):9445–9454, 2001.

Botond Roska and Frank Werblin. Rapid global shifts in natural scenes block spiking in specific ganglion cell types. *Nature Neuroscience*, 6:600–608, 2003.

Yasser Roudi, Sheila Nirenberg, and Peter E. Latham. Pairwise Maximum Entropy Models for Studying Large Biological Systems: When They Can Work and When They Can't. *PLoS Computational Biology*, 5(5): e1000380, 2009.

Elad Schneidman, William Bialek, and Michael J. Berry. Synergy, redundancy, and independence in population codes. *Journal of Neuroscience*, 23(37):11539–11553, 2003a.

Elad Schneidman, Susanne Still, Michael J. Berry, and William Bialek. Network Information and Connected Correlations. *Physical Review Letters*, 91(23):238701, 2003b.

Elad Schneidman, Michael J. Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440:1007–1012, 2006.

Mark J. Schnitzer and Markus Meister. Multineuronal Firing Patterns in the Signal from Eye to Brain. *Neuron*, 37:499–511, 2003.

Ronen Segev, Elad Schneidman, Joe Goodhouse, and Michael J. Berry. Role of eye movements in the retinal code for a size discrimination task. *Journal of Neurophysiology*, 98:1380–1391, 2007.

Claude E Shannon and Warren Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

Lindsay T. Sharpe and Andrew Stockman. Rod pathways: the importance of seeing nothing. *Trends in Neurosciences*, 22:497–504, 1999.

Jonathon Shlens, Greg D. Field, Jeffrey L. Gauthier, Matthew I. Grivich, Dumitru Petrusca, Alexander Sher, Alan M. Litke, and E. J. Chichilnisky. The Structure of Multi-Neuron Firing Patterns in Primate Retina. *Journal of Neuroscience*, 26:8254–8266, 2006.

Jonathon Shlens, Greg D. Field, Jeffrey L. Gauthier, Martin Greschner, Alexander Sher, Alan M. Litke, and E. J. Chichilnisky. The Structure of Large-Scale Synchronized Firing in Primate Retina. *Journal of Neuroscience*, 29(15):5022–5031, 2009.

Malcolm M. Slaughter and Robert F. Miller. 2-Amino-4-Phosphonobutyric Acid: A New Pharmacological Tool for Retina Research. *Science*, 211: 182–185, 1981.

Aonan Tang, David Jackson, Jon Hobbs, Wei Chen, Jodi L. Smith, Hema Patel, Anita Prieto, Dumitru Petrusca, Matthew I. Grivich, Alexander Sher, Pawel Hottowy, Wladyslaw Dabrowski, Alan M. Litke, and John M. Beggs. A Maximum Entropy Model Applied to Spatial and Temporal Correlations from Cortical Networks In Vitro. *Journal of Neuroscience*, 28(2):505–518, 2008.

W. Rowland Taylor, Shigang He, William R. Levick, and David I. Vaney. Dendritic Computation of Direction Selectivity by Retinal Ganglion Cells. *Science*, 289(5488):2347–2350, 2000.

Gasper Tkacik, Elad Schneidman, Michael J. Berry II, and William Bialek. Ising models for networks of real neurons. *arXiv*, q-bio/0611072v1:1–4, 2006.

Gasper Tkacik, Elad Schneidman, Michael J. Berry II, and William Bialek. Spin glass models for a network of real neurons. *arXiv*, 0912.5409v1:1–15, 2009.

Gasper Tkacik, Jason S. Prentice, Vijay Balasubramanian, and Elad Schneidman. Optimal population coding by noisy spiking neurons. *Proceedings of the National Academy of Sciences USA*, 107(32):14419–14424, 2010.

David K. Warland, Pamela Reinagel, and Markus Meister. Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, 78:2336–2350, 1997.

Heinz Wässle. Parallel processing in the mammalian retina. *Nature Reviews*, 5:1–11, 2004.

Claudia Werner and Werner Himstedt. Mechanism of Head Orientation during Prey Capture in Salamander. *Zoologische Jahrbücher. Abteilung für allgemeine Zoologie und Physiologie der Tiere*, 89:359–368, 1985.

David H. Wolpert and David R. Wolf. Estimating functions of probability distributions from a finite set of samples. *Physical Review E*, 52:6841, 1995.

Xiong-Li Yang. Characterization of receptors for glutamate and GABA in retinal neurons. *Progress in Neurobiology*, 73:127–150, 2004.

Alfred L. Yarbus. *Eye Movements and Vision*. Plenum Press, 1967.

Eberhart Zrenner, Karl Ulrich Bartz-Schmidt, Heval Benav, Dorothea Besch, Anna Bruckmann, Veit-Peter Gabel, Florian Gekeler, Udo Greppmaier, Alex Harscher, Steffen Kibbel, Johannes Koch, Akos Kusnyerik, Tobias Peters, Katarina Stingl, Helmut Sachs, Alfred Stett, Peter Szurman, Barbara Wilhelm, and Robert Wilke. Subretinal electronic chips allow blind patients to read letters and combine them to words. *Proceedings of the Royal Society B*, 278:1489–1497, 2011.