Sensory coding in strongly correlated neural populations

Gabriel Mahuas

Sorbonne Université

École doctorale Cerveau, Cognition, Comportement, Nº 158

Laboratoire de Physique de l'École Normale Supérieure (UMR 8023)
&
Institut de la Vision (UMR S968)

Sensory coding in strongly correlated neural populations

Par Gabriel MAHUAS

Thèse de doctorat de NEUROSCIENCES

Dirigée par Ulisse FERRARI
Et par Thierry MORA

Présentée et soutenue le 16/09/2024

Devant un jury composé de :

Georges DEBRÉGEAS, DR-HDR	Laboratoire Jean Perrin	Président
Tim GOLLISCH, Prof.	Université de Göttingen	Rapporteur
Cristina SAVIN, Assist. Prof.	Université de New York	Rapportrice
Adrienne FAIRHALL, Prof.	Université de Washington	Examinatrice
Ulisse FERRARI, CR-HDR	Institut de la Vision	Directeur de thèse
Thierry MORA, DR-HDR	École Normale Supérieure	Directeur de thèse

Abstract

Title: Sensory coding in strongly correlated neural populations.

Keywords: Vision; Retina; Sensory coding; Noise correlations; Collective behavior

Abstract: Neurons in sensory systems encode information related to input stimuli in their collective activity. The neural population activity is shaped by two main sources of correlations. First, natural stimuli tend to contain strong spatio-temporal correlations, which transpire in neural responses through what are called stimulus or signal correlations. Second, interactions between neurons further promote correlations in the population activity. In particular, biological neurons are not deterministic but noisy, and neuronal interactions will correlate noise across the network in a phenomenon termed noise correlations. Positive noise correlations are widely observed across sensory systems, from the retina to the cortex, especially between neurons with similar sensitivity to the stimulus. This is at odds with a large body of theoretical work which suggests that simultaneously positive noise and stimulus correlations should reduce the sensory information encoded by the network compared to the case of uncorrelated noise. To investigate this discrepancy, we started by developing an inference approach for Generalized Linear Models that inherently distinguishes between the contributions of the stimulus and that of interactions. This approach allowed us to model accurately the collective behavior of OFF-alpha ganglion cells from the rat retina and showed that the network of interactions underlying noise correlations in retinal data is robust to strong changes in stimulus statistics. To investigate how the interplay between noise and stimulus correlations affect sensory coding, we then developed a small-correlation approximation of the mutual information between stimulus and response. This approximation is accurate and can be directly applied to quantify mutual information between response and stimulus in experimental data. Furthermore, our development also provides a clear and transparent view of how neural neural correlations impact information: positive noise correlations will benefit sensory coding not only when they are opposed to stimulus correlations, but also when they are sufficiently large. Analysis of rat retinal recordings revealed that both regimes could occur experimentally for populations of OFF-alpha ganglion cells depending on the statistics of the stimulus. We further investigated these effects in large population models and showed both theoretically and experimentally that positive noise correlations benefit coding by boosting the information encoded about fine-grained stimulus details.

À Eulalie.

Thanks

I would like to express my gratitude to all those who contributed in one way or another to this long journey. First and foremost to my PhD advisors, Ulisse Ferrari and Thierry Mora, for their mentorship and patient support throughout the years.

I would also like to thank the members of my PhD committee, Georges Debrégeas and Jonas Ranft, for the benevolent feedback they provided during the yearly thesis committees.

I want to thank Georges Debrégeas a second time, Adrienne Fairhall, Tim Gollisch and Cristina Savin for accepting to take part in my PhD jury, and especially the readers for their constructive comments on the manuscript.

To all past and current students and postdocs of the team at the IdV, Alina, Awen, Baptiste, Carlo, Chiara, Danica, Déborah, Filipo, Francesco, Geoffroy, Guilhem, Isaure, Jeanne, Kyle, Leon, Mathieu, Melide, Mouna, Pietro, Rémi, Rodrigo, Samuele, Sarah, Shawn, the Simones, Thomas, Timothé, Tobias, Tom, Tommaso and Victor, thank you for the lively atmosphere you brought to the office and for the good moments spent in and out of the lab. Our time sharing the stale air of the UCL open space will not be swiftly forgotten.

I also want to thank the researchers of the team: Matias, Matthew and particularly Olivier, for their help, benevolent advice and critical yet constructive insights.

I would like to thank the Sorbonne Center for Artificial Intelligence for financing this work, and in particular Xavier Fresquet and Nora Roger for their support and help to the organization of the Cosyne 2024 S.O.S. workshop.

I would like to thank the research support and administrative teams of the IdV, in particular Abir Djeddi, Céline Cadic, Stéphane Duhieu, and Olga Keh, as well as the teams of the LPENS, especially Olga Hodges, for their help in solving all things related to administration and logistics.

Last but not least, I want to thank my parents, family, and friends for their care, support, and understanding throughout the years. Above all, I am deeply grateful to Solenn, for I have never seen things more clearly than since you began sharing my life.

Contents

Abstract i

Thanks iii

Introduction 1

1 The retina 5

1.1 Physiology of the retina . . . 5

1.2 Elements of retinal computation . . . 8

1.2.1 The canonical receptive field of retinal ganglion cells . . . 8

1.2.2 Stimulus encoding by single ganglion cells . . . 11

1.3 Parallel processing in the retina . . . 14

1.3.1 The diversity of ganglion cell types . . . 15

1.3.2 The mosaic organization of the retina . . . 15

1.4 The collective behavior of retinal ganglion cells . . . 16

1.4.1 Stimulus induced correlations . . . 17

1.4.2 Intrinsic interactions further promote correlations . . . 17

1.5 Conclusion . . . 20

2 Population coding of stimulus information 21

2.1 Quantifying the collective behavior of sensory neurons . . . 22

2.1.1 Disentangling stimulus and noise correlations . . . 22

2.1.2 Experimental quantification . . . 24

2.2 The impact of correlated activity on stimulus encoding . . . 25

2.2.1 Geometrical picture . . . 25

2.2.2 Fisher information . . . 27

2.2.3 Mutual information . . . 32

2.3 Modelling the collective behavior of sensory neurons . . . 36

2.3.1 Maximum entropy approach . . . 37

2.3.2 Generalized Linear Model . . . 41

2.4 Conclusion 44

3 A new inference approach for training shallow and deep generalized linear models of noisy interacting neurons 47

3.1 Introduction 49
3.2 Recordings 51
3.3 Generalized linear model 51
3.4 Failure of GLM for complex stimuli 52
3.5 A two-step inference approach 54
3.6 Two-step inference allows for generalizing across stimuli 57
3.7 Deep GLM outperforms previous approaches 58
3.8 Discussion 59
S1 Empirical data and correlations 62
S2 Correction for the absolute refractory period 62
S3 Generalization results for moving bars stimulus 63
S4 Time Distributed Convolutional Neural Network 63

4 A small-correlation expansion to quantify information in noisy sensory systems 65

4.1 Introduction 67
4.2 Small correlation expansion of the mutual information 68
4.3 Noise synergy 70
4.4 Numerical test on synthetic data 71
4.5 Application to retinal data 73
4.6 Discussion 74
S1 Second order approximation 77
S2 Resummed expansion 80
S3 Generalized linear model simulations 84
S4 Link to the small time bin expansion (Panzeri *et al.* 1999) 88

5 Strong but not weak noise correlations are beneficial for population coding 95

5.1 Introduction 97
5.2 Results 98
5.3 Discussion 106
5.4 Methods 109
5.5 Supplementary information 117

6 Discussion and perspectives **119**

6.1 Stimulus dependency of noise correlations 120

6.2 Beyond pairwise instantaneous correlations 121

6.3 Stimulus specific impact of noise correlations 122

6.4 Population coding of specific stimulus features 123

6.5 Conclusion 125

Introduction

We often take for granted how effortlessly we perceive our environment. Yet, the problems our sensory organs have to solve to make sense of the flood of physical stimulation they collect continuously are far from being straightforward. The biological substrate in which sensory systems are implemented comes with limitations that make this task especially difficult. Sensory systems convert their physical inputs into spikes of electrical activity that are used by neurons throughout the central nervous system to convey information. Each spike costs energy and due to the limited resources available to them, sensory neurons have to make the most of the information they extract from the environment.

A consequence of the energetic efficiency of biological neurons is that they are not deterministic, but noisy. When presented many times with the exact same input stimulus, a sensory neuron may exhibit variability in its activity. This will limit the neuron's ability to convey information about the sensory signal it encodes. In the case of networks of neurons, neural noise is not independent from one neuron to the other, but is correlated across the population by neuronal interactions. This phenomenon, often called noise correlations in the literature, will modulate the impact of neural noise on stimulus encoding.

The effect of noise correlations on sensory coding has been widely debated in the literature during the last decades. Most theoretical investigations argued that positive noise correlations as those observed in many systems ranging from the retina to cortical areas would increase the detrimental effect of neural noise when stimuli also promote positive correlations in neural responses. By contrast, many experimental studies showed that noise correlations tend to be positive and significant when stimulus also positively correlate neural activity. Due to the lack of a direct and straightforward relationship between the structure of response correlations and the information carried by neural responses, this discrepancy between experimental observations and theoretical predictions remains poorly understood.

The retina is a thin piece of brain located at the back of the eye that takes as an input the visual stimulus, encodes it into the collective activity of its output neurons, and provides

all visual information available to the brain. Thanks to its layered organization that makes it easy to record from and to the virtual absence of feedback it receives from the rest of the nervous system, the retina is an ideal place to start investigating how stimuli are processed and encoded into the collective activity of sensory neurons.

The goal of this thesis is to investigate how noise correlations impact the encoding of stimulus information in the collective activity of sensory neurons, using the retina as a model system. The methods, results and insights we derive here are general and can be applied beyond the retina to other sensory systems.

Thesis outline

Introductory chapters

- **Chapter 1.** The first chapter gives a basic overview of retinal physiology and discuss how noise and signal correlations contribute to the collective behavior of retinal ganglion cells.
- **Chapter 2.** The second chapter focuses on quantitative analyses of noise correlations and how they impact sensory coding. It starts by a section giving precise definitions of noise and signal (or stimulus) correlations that will be used during the rest of the thesis. In a second section, we discuss the literature regarding the impact of noise correlations on sensory coding through the lens of two different information measures: the Fisher information and the mutual information.

Results

- **Chapter 3, previously published as reference [101].** Studying the collective behavior of sensory neurons often requires modeling neural population responses to incoming stimuli. In this chapter, we present a new inference approach for Generalized Linear Models that palliates to some well known inaccuracies of the classical maximum likelihood approach and disentangles signal and noise correlations during training, resulting in models that better describe the collective behavior of retinal ganglion cells.
- **Chapter 4, previously published as reference [102].** We develop a small correlation approximation for the mutual information that can be easily applied to data and gives accurate estimates. We validate the accuracy of this approach on synthetic data and showcase its application to retinal recordings. Further,

this approximation results in a simple and interpretable expression relating the structure of pairwise noise and signal correlations to their impact on information.

- **Chapter 5, submitted and available as a preprint [100].** In this final part, we develop the insights derived from the approximation introduced in the previous chapter. We start by using it to interpret the structure of noise and signal correlations observed in retinal recordings. Next, we extend analytically our conclusions to a large Gaussian population model and evaluate theoretically and experimentally the impact of noise correlations on different features of the encoded stimulus.

Conclusion

- **Chapter 6.** We conclude the thesis by discussing limitations of our findings as well as perspectives and potential future directions.

Chapter 1

The retina

The retina constitute the first step in the processing of visual information that ultimately gives rise to perception. It is a thin outgrowth of brain tissue that is located at the back of the eye. It takes as an input light from the external environment, which is focused on it by the eye's optics (cornea and lens), and outputs neural signals that are sent to higher visual centers for further processing. Every single bit of visual information that is available to the brain is extracted from visual inputs and packaged by this remarkably small and efficient neural network.

This first chapter aims to provide an overview of retinal physiology and functions, and set the stage for using the retina as a model to investigate stimulus encoding in strongly correlated neural populations.

We will start with a rapid examination of the retinal anatomy, describing the different cell classes and neuronal interactions that compose the retinal network. Second, we will briefly present the way the retina processes visual stimuli and how this processing relates to the retinal organization. Concluding the chapter, we will discuss the structure of collective neural activity in the retinal output as well as its different sources.

The considerations that will be made in the following sections will be valid for mammalians, and to a large extent for vertebrates in general.

1.1 Physiology of the retina

Despite the great diversity that exists in retinal organization across species, some general features of the retina's structure are remarkably conserved across vertebrates. The vertebrate

retina is organized in five alternating nuclear and synaptic layers (Fig. 1.1 A):

- The **outer nuclear layer (ONL)** contains the cell bodies of the photoreceptors, responsible for converting light into electrical signals.
- The first synaptic connections occur in the **outer plexiform layer (OPL)**, where photoreceptors connect with bipolar and horizontal cells.
- The **inner nuclear layer (INL)**, hosts the cell bodies of bipolar, horizontal, and amacrine cells that perform the bulk of retinal computations.
- The **inner plexiform layer (IPL)** is where bipolar cells synapses make contact with ganglion cells, and where amacrine cells connect primarily with bipolar cells.
- The **ganglion cell layer (GCL)** is composed of the ganglion cells' bodies.

Light from the external environment enters the eye through the pupil and is focused onto the retina by the cornea and the lens (Fig. 1.1 A). There, photons cross the entire width of the retina until they reach its outermost layer, where they will modulate the membrane potential of photoreceptors (cones and rods). During this phenomenon termed photo-transduction, a photon hits a photoreceptor cell where it is absorbed by a retinaldehyde molecule that's bounded to an opsin protein. The retinaldehyde undergoes photo-isomerization, which in turn changes the conformation of the opsin to which it is bounded. This in turns triggers a cascade of chemical reactions that eventually lead to the closing of sodium channels and the hyperpolarization of the cell. The resulting time varying electrical signal encodes the local level of light impinging the retina: rods output conveys the signal relative to local luminance while cones provide the signal necessary to color sensitivity.

In the following layer of the retina, bipolar cells can be both inhibited or excited by photoreceptors, leading to the emergence of the ON and OFF visual pathways. There, horizontal and amacrine cells provide (mostly) inhibition at different stages of the retinal processing (Fig. 1.1, B.). Amacrine cells, on the other hand, provide inhibition to the output of bipolar cells. They synapse at the inner plexiform layer, where they interact with both bipolar cells and ganglion cells. They play an important purpose in both the temporal and spatial aspects of visual processing, enhancing contrast and contributing to the regulation of signal timing.

The final and innermost layer of the retina is populated by retinal ganglion cells. These neurons integrate the signal processed by upstream cells and summarize the information extracted by the retina in the form of action potentials, short electrical impulses that

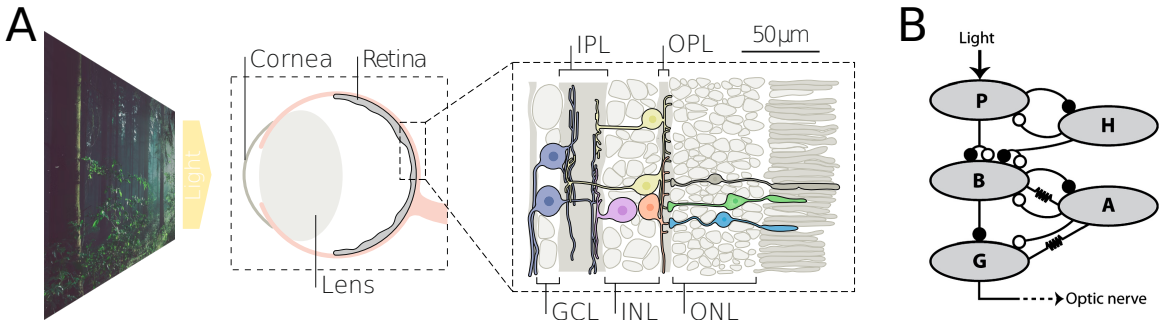

Figure 1.1: **A**. Light from the visual scene enters the eye through the iris and is focused on the retina by the cornea and the lens. Photons from incoming stimuli then penetrate the retina until they reach the photoreceptors, the rods (dark grey) and the cones (light blue and green), which bodies are located in the outer nuclear layer (ONL). Rods and cones convert incoming photons into electrical signal that is fed to bipolar cells (yellow) in the inner nuclear layer (INL), then further relayed to ganglion cells (dark blue) in the ganglion cells layer (GCL). Horizontal cells (orange) and amacrine cells (purple) are both located in the inner nuclear layer (INL) as well. Horizontal cells synapse in the outer plexiform layer (OPL) while amacrine cells synapse in the inner plexiform layer (IPL). Panel adapted from [12]. **B**. Neuronal interactions in the retina can be classified in two main categories. Synaptic connections on the one hand can be either excitatory (full circles) or inhibitory (empty circles). On the other hand, gap junctions (small resistor symbol) are excitatory electrical couplings. Photoreceptors (P) can excite or inhibit bipolar cells (P), which can only excite directly ganglion cells (G). Inhibitory feedback is exerted by horizontal cells (H) on photoreceptors and bipolar cells, while amacrine cells (A) inhibit ganglion cells through chemical synapses. Simultaneously, amacrine cells connect to bipolar and ganglion cells via gap junctions. Panel reprinted from [62].

support information transmission throughout the brain. The axons of all retinal ganglion cells bundle together to form the optic nerve, a cranial nerve that crosses the retina at the optic disk before connecting to higher visual areas.

Ganglion cells can be seen as the output of the retina, they provide the brain with a complex and sparse neural representation (or "neural code") of the visual input. Thanks to the layered structure of the retina and to the position of ganglion cells in its architecture, it is possible to record through *ex-vivo* electrophysiological experiments the electrical activity of large retinal ganglion cells populations. This makes the retina an ideal system to study the encoding of stimulus information in the collective activity of sensory neurons.

In this thesis, we will investigate how the structure of the neural representation formed by the collective activity of retinal ganglion cells affects the amount of information that is being provided to the brain about the visual input collected by the retina.

1.2 Elements of retinal computation

The purpose of early sensory systems is two-folds: first, they need to convert their physical input, the stimulus, into electrical signals that can be understood by the rest of the brain. Second, they need to extract relevant information from the stimulus and represent it efficiently in the neural code. Since the first recordings of single retinal ganglion cells by Hartline [69, 141], Kuffler [90] and Barlow [17], researchers have come a long way in their understanding of the retina. Far from being a simple camera that would convert "pixel wise" the visual input into a neural image, the retina already performs a variety of computations over the visual stimulus before sending it to the brain [62, 10, 12].

1.2.1 The canonical receptive field of retinal ganglion cells

All information extracted by the retina about the visual stimulus come from the activity of the few millions of photoreceptors that compose its input, and has to pass by the tens of thousands retinal ganglion cell axons that constitute its only link to the rest of the brain [76].

This tremendous decrease in dimension across the retinal hierarchy suggests that this remarkably small network compresses the visual input in an efficient neural representation before sending it to the brain. As Attneave [5] and later on Barlow [16] suggested, one way to reach such an efficient representation of visual information consists in discarding

redundant information from the sensory input.

1.2.1.1 Spatial receptive field

Natural scenes are full of such redundancies and exhibit blatant long range spatial correlations that originate from their distinctive scale invariant and hierarchical nature [150]. They are composed of overlapping objects that each possess a certain spatial extent and present significant redundancy in their visual properties.

Retinal neurons are sensitive to stimuli within a local area of the eye's field-of-view that is called the receptive of that neuron (Fig. 1.2). Ganglion cells receptive fields are generally characterized by a central area that, when flashed with a small bright spot, increases activity for some cells (the ON ganglion cells) and decreases that of others (the OFF cells). In the surrounding area of the receptive field, a small bright spot will have the opposite effect: the ON ganglion cells will be suppressed while OFF cells will be stimulated.

In the 50s, Kuffler observed that while ganglion cells responded strongly to a small, bright spot of light flashed in the center part of their receptive field, flashing two spots at the same time in the center *and* the surround of the cell led to a significantly suppressed response [90]. This experiment revealed how center and surround parts of the receptive field interact such that the stimulation of one and inhibition of the other leads to an overall diminution of activity (see Fig. 1.2, condition 3 versus conditions 1 and 4). This phenomenon, called center-surround antagonism, is considered a hallmark of retinal computation.

Through this antagonism, neurons encode preferentially features of the stimulus that induce strong local contrast such as edges while filtering out uniform patches of the visual input. This mechanism leads to a decorrelation of neural responses compared to their initial visual inputs, that, in agreement with Attneave and Barlow's theories, reduces redundancy and increases the informational content of ganglion cells activity.

1.2.1.2 Temporal receptive field

Beyond their sensitivity to spatial patterns, the receptive fields of retinal ganglion cells also have a temporal dimension which enables them to detect temporal variations in the stimulus. The natural world is composed of persistent objects that evolve in a smooth and continuous fashion over time. Akin to the spatial correlations that emerge from the spatial extent of visual objects, strong temporal correlations arise from the temporal extent of objects in natural movies. Holding similar considerations for temporal processing as were

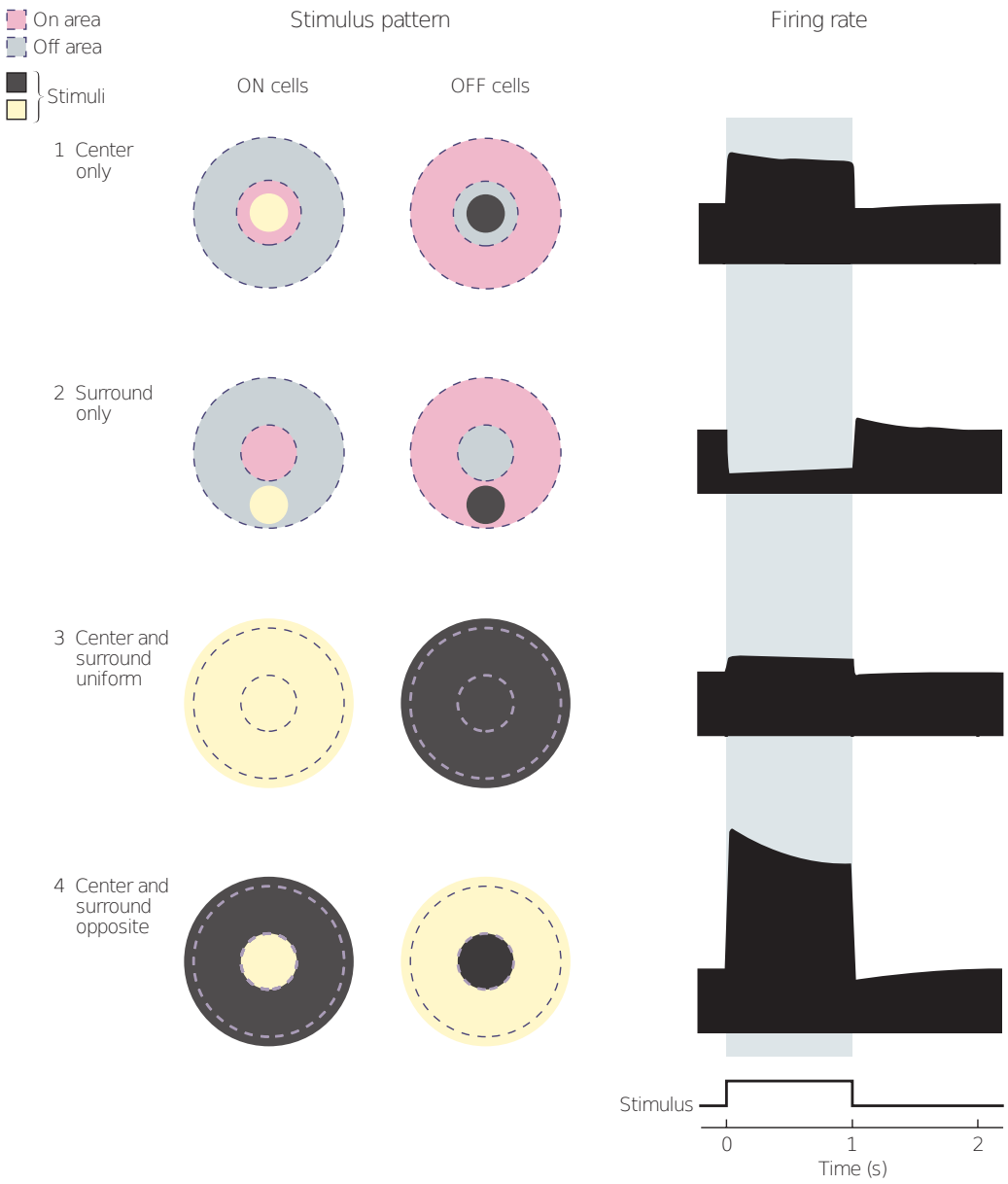

Figure 1.2: The center-surround organization of retinal ganglion cells' receptive fields (adapted from [79]). **A.** The two columns show idealized receptive fields of ON and OFF cells. Pink corresponds to ON areas and grey to OFF areas. Each row corresponds to a different stimulus configuration (1: center excitation, 2: surround inhibition, 3: antagonistic stimulation, 4: preferred contrast). The cells are stimulated with local increases (white) or decreases (black) of contrast. **B.** Each row illustrates the typical firing rate of ON and OFF cells to the corresponding stimulus configurations presented in panel A. The grey shaded area corresponds to the duration of stimulus presentation. Each cell fires the most when presented with its preferred contrast (rows 1 and 4), and firing is increased with contrast (cells fire more in condition 4 with respect to condition 1). Firing is suppressed by stimulation of the surround by a spot of the cell's preferred contrast (condition 2), and uniform stimulation of the receptive field drives only a weak response from the cell.

made for spatial processing in the previous section [4, 44, 111], we can explain the biphasic antagonism observed in the temporal receptive fields of retinal ganglion cells (see Fig. 1.3A for a theoretical prediction and Fig. 1.3B for experimental data).

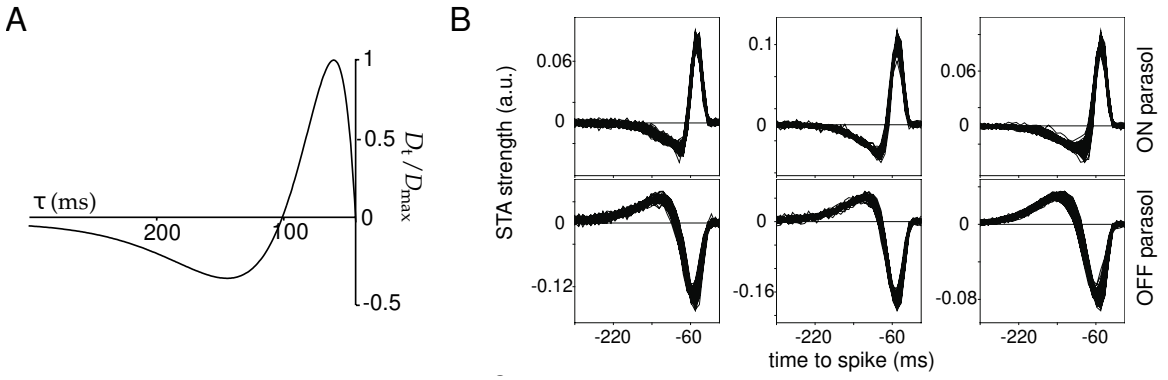

Figure 1.3: Temporal processing by retinal ganglion cells agrees with theoretical predictions of the redundancy reduction hypothesis. **A.** This theoretical temporal trace of an optimal linear filter was derived originally in the context of LGN cells [44], but extends straightforwardly to retinal processing [111, 37]. The filter exhibits clear antagonism that promotes redundancy reduction. Adapted from [37]. **B.** Time course of ON (top row) and OFF (bottom row) parasol cells from the macaque retina. The three columns correspond to different retinas and each plots shows the overlap of temporal traces from multiple neurons. Adapted from [55].

1.2.2 Stimulus encoding by single ganglion cells

1.2.2.1 Linear stimulus integration

So far we have seen that retinal ganglion cells can be canonically described as filters that preprocess the visual input through their receptive field.

Let's formalize this description of retinal processing and build a simple model of stimulus encoding by single ganglion cells. Let $\nu(t)$ be the rate at which ganglion cell emits action potentials at time t, and $S(x, y, t')$ the light level in point (x, y) of the visual space at time t' (a.k.a. the stimulus). It follows from the previous sections that the firing rate $\nu(t)$ is related to the filtering of $S(x, y, t')$ by a spatiotemporal kernel $K(x, y, \tau)$ [145, 49, 31]:

$$h_{\text{stim}}(t) = \iiint \mathrm{d}x\mathrm{d}y\mathrm{d}\tau K(x, y, \tau)S(x, y, t - \tau), \tag{1.1}$$

$$\nu(t) = \mathcal{F}\left(b + h_{\text{stim}}(t)\right). \tag{1.2}$$

This simple model is characterized by three main components:

- **The spatiotemporal receptive field** $K(x, y, \tau)$, which in a first approximation can be factorized in space and time components: $K(x, y, \tau) = K_{\mathrm{s}}(x, y)K_{\mathrm{t}}(\tau)$. Textbook and historical descriptions of the spatial component are often based on a difference of Gaussians:

$$K_{\mathrm{s}}(x, y) = \alpha_{\mathrm{c}}\mathcal{G}_{\sigma_{\mathrm{c}}}(x - x_{\mathrm{c}}, y - y_{\mathrm{c}}) - \alpha_{\mathrm{s}}\mathcal{G}_{\sigma_{\mathrm{s}}}(x - x_{\mathrm{c}}, y - y_{\mathrm{c}}). \tag{1.3}$$

  Here, $\mathcal{G}_\sigma$ is a bi-dimensional Gaussian function of standard deviation σ. α_{c} and α_{s} dictate the strength of the center and the surround while σ_{c} and σ_{s} dictate their width. The center of the receptive field is set by x_{c} and y_{c}. The temporal component on the other hand, needs to account for the temporal antagonism illustrated in Fig. 1.3.

- **The bias** b sets the baseline activity of the neuron.

- **The nonlinearity** $\mathcal{F}$ is a function that essentially prevents the firing rate from being negative. This function can be as simple as a rectifier function $\mathcal{F}(x) = \max(0, x)$ or an exponential $\mathcal{F}(x) = \exp(x)$. It can also account for the fact that the firing rate cannot be arbitrarily large by being a saturating function such as a sigmoid $\mathcal{F}(x) = (1 + \exp(-x))^{-1}$.

1.2.2.2 Stochastic spiking

Poisson spiking The response of single retinal ganglion cells is not solely characterized by the rate at which they fire action potentials. When presented with the same stimulus multiple times, the spiking response of retinal ganglion cells is stochastic and tend to fluctuate from one trial to the other. Several processes contribute to these random fluctuactions. First, noise originating in photoreceptors due to thermal noise affecting phototransduction is propagated through the retinal network [65, 2]. Second, noisy synaptic release across the retinal hierarchy is accumulated and propagated to retinal ganglion cells [2, 23], where it is combined with the intrinsic stochastic nature of ganglion cells [184].

This intrinsic variability is an important aspect of neural responses: it is ultimately what limits the neurons' ability to reliably encode stimulus information, but it is also the result of the neurons' energetic efficiency [169]. A straightforward way to incorporate the stochastic aspect of neural responses in our model consists of using the firing rate predicted by our previous model as the parameter of a random process. A common choice for the spiking process is the Poisson process, which assumes that the amount of spikes fired in

non-overlapping time windows are independent. In this setup, inter-spike-intervals follow an exponential distribution at constant firing rate. The resulting model is called a Linear-Nonlinear Poisson cascade model [31, 164], often referred to as LN (or LNP) model in the literature. The different parameters of the model (filter, bias, and eventually nonlinearity) can be fit to experimental recordings of single neurons [31, 126, 164, 134] through maximum-likelihood estimation. Once fit to data, the LN model has been shown to predict fairly well the response of macaque retinal ganglion cells to simple white noise stimuli [31, 135].

Spiking regularity and intrinsic variability Neural variability in single retinal ganglion cells tends to present more regularity than what would be expected from a Poisson process [15, 83, 52] (see Fig. 1.4). Due to refractory effects that follow the emission of a spike [20, 52], inter-spike-interval distributions of real neurons depart significantly from the exponential distribution observed in Poisson spiking.

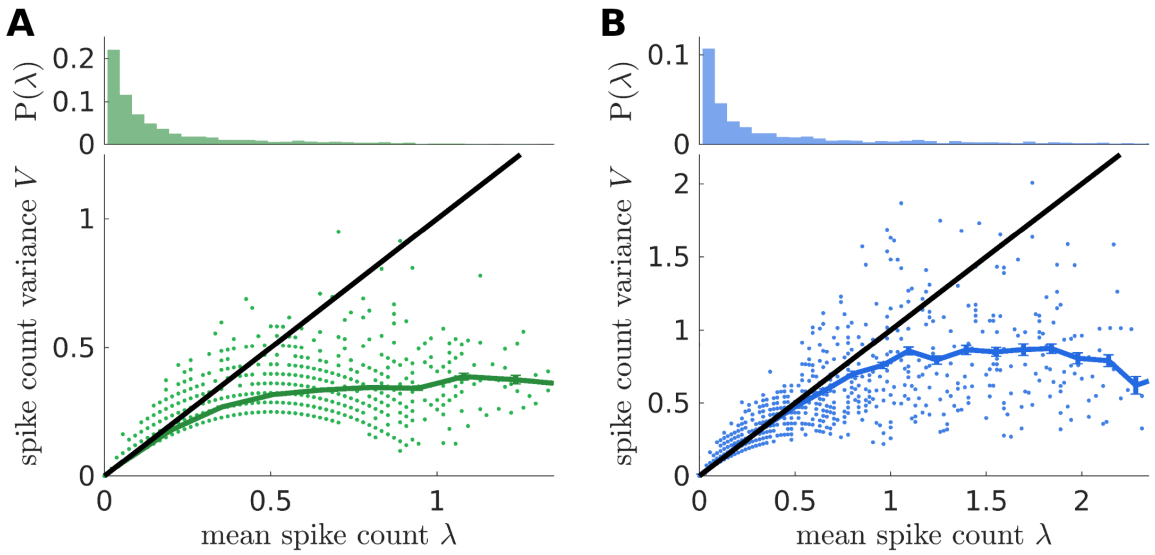

Figure 1.4: Sub-Poissonian variability of rat retinal ganglion cells, adapted from [52]. **A.** Mean-variance relationship of OFF-α cells. Each dot corresponds to the variability of one cell across many trials of the same stimulus condition, compared to the mean spike count (i.e. the firing rate) computed across repetitions. The black line corresponds to the mean-variance relationship of a Poisson process: OFF-α cells have a sub-Poissonian variability. **B.** Same as the previous panel, but for ON-α cells, in the same rat preparation.

A simple way to account for the specificity of neural variability consists of extending the LNP model presented above into a Generalized Linear Model (GLM) [182, 135] by feeding the nonlinearity with a second time varying field $h_{\text{past}}(t)$, added on top of the integrated stimulus. This field would result from the convolution of the neuron's past activity by a spike-history filter $J(\tau)$ that accounts for any refractory effects (seed Fig. 1.5 for a schematic). For simplicity, let's assume time has been discretized in bins of size Δt,

and that space has been discretized in pixels. Let $R(t)$ denote the number of spikes emitted by a ganglion cell in time bin t. The previous rate model now becomes:

$$h_{\text{stim}}(t) = \sum_{\tau} \sum_{x,y} K(x, y, \tau) S(x, y, t - \tau), \tag{1.4}$$

$$h_{\text{past}}(t) = \sum_{\tau} J(\tau) R(t - \tau), \tag{1.5}$$

$$\nu(t) = \mathcal{F}\left(b + h_{\text{stim}}(t) + h_{\text{past}}(t)\right). \tag{1.6}$$

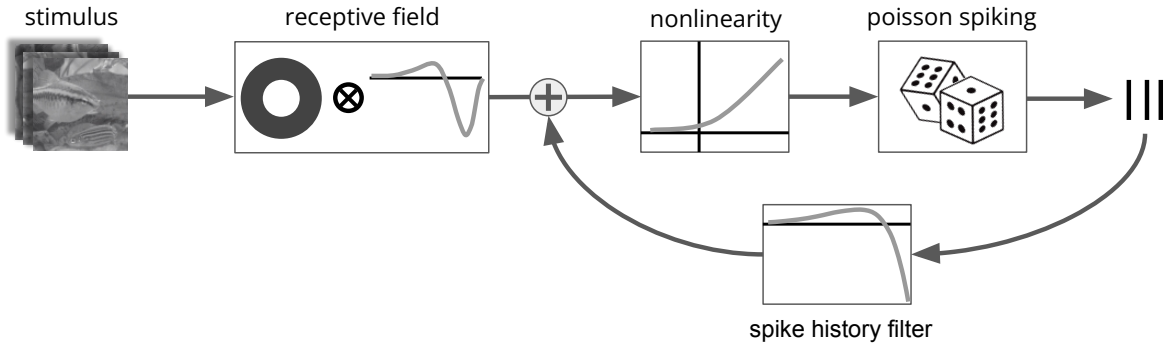

Figure 1.5: Schematic of the single cell Generalized Linear Model. The visual stimulus is convolved by a spatiotemporal receptive fields. The resulting processed stimulus signal gets added to the convolution of the neuron's past spiking activity by a spike history filter before being fed to a nonlinearity. The output of the nonlinearity is used as the parameter of a Poisson process from which spikes are sorted.

1.3 Parallel processing in the retina

A generic property of the different areas that compose the visual system is that they process the visual stimulus through parallel channels [118]. We have already seen how retinal ganglion cells can be coarsely categorized in ON and OFF cells, each of these groups signaling complementary contrast information to the brain. In reality, parallel processing in the retina goes far beyond this simple ON-OFF filtering of the visual input. Early investigations [95, 96] already suggested that there could exist a variety of retinal cell types that each respond to a specific feature of the stimulus. This view offers a better explanation for the intricate architecture and cellular diversity of the retinal network: the complex structure of the retina only mirrors the complex nonlinear computations that underlie parallel stimulus features extraction.

1.3.1 The diversity of ganglion cell types

As stated previously in this chapter, retinal ganglion cells summarize the visual information extracted by the retina into their spiking activity before sending it to the brain. These neurons have been historically classified in ON, OFF, and ON-OFF cells depending on their preferred contrast, and in transient or sustain cells depending on the temporal dynamics of their response [33, 10].

Despite the persistence of this rather simple view in the larger vision science community, nonlinear feature extraction has been known to occur in the retina since the 60s, when early work showed that some ganglion cells could be selective to local motion, direction of motion and orientation [95, 14, 33]. Recent analyses identified that there is more than 30 different ganglion cell types in the mouse retina [10, 185]. Nonlinearity and parallel processing go hand-in-hand in the retina: each cell type is characterized by a distinct nonlinearity that enables the extraction of specific stimulus features [179].

A certain class of retinal neurons, the α ganglion cells, hold a particular status in the mammalian retina: they are characterized by their large cell bodies, fast conducting axons and short response latency [89]. These features as well as the fact that they are remarkably conserved across species [132] suggest that they play a prominent role in visual processing. α cells come in four sub-types: OFF transient, ON transient, OFF sustain and ON sustain. Sustain cells seem to encode mostly local contrast information, while transient cells are also sensitive to motion [40, 188]. Interestingly, it has recently been established through single-cell transcriptomic analyses that parasol and midget ganglion cells from the primate retina are orthologs of α-cells [68]. In the case of simple stimuli, this broad category of cells seems to be fairly well modeled by linear-nonlinear models [31, 135, 149, 101], although the same models fail in the case of more naturalistic or complex stimuli [71, 110].

1.3.2 The mosaic organization of the retina

Most of the different ganglion cell types seem to cover an extensive range of the retina so as to extract feature maps across large portions of the visual field (Fig. 1.6 A.). Within a cell type, the receptive fields follow an approximate triangular lattice (Fig. 1.6 B.). This organization, akin to a sphere packing problem, allows ganglion cells to sample densely from the visual scene [43, 189, 106].

Many properties of the receptive fields that form these mosaics differ not only between cell types but also within a single type [21]. Local statistics in natural scenes and features

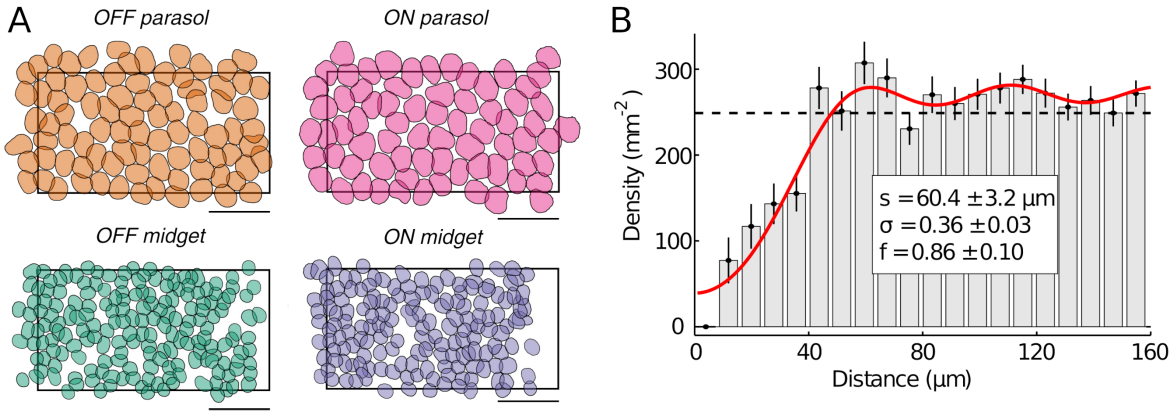

Figure 1.6: The mosaic organization of retinal ganglion cells. **A.** Receptive fields of midget and parasol cell from the primate retina tile tigthly the visual space. Largest gaps in the population are due to missing cells. Adapted from [142]. **B.** Radial density function of a single ganglion cell type (W3 cells, local motion detectors) in a small portion of the mouse retina. The red line corresponds to the fit of a noisy sphere packing model that models the approximate triangular lattice of W3 cells. Adapted from [198].

of behavioral interest vary across the animal's field of view. For instance, the sky tends to be uniform and salient flying objects are likely to be predators for a mouse, while the bottom part of the view is likely to be overcome with dark contrast and be where food is located. Mirroring these variation of local scene statistics, the properties of receptive fields often vary across regions of the retina. These variations are believed to be adaptations to the statistics of natural scenes [12, 66], in accordance with efficient coding theory.

Further considerations based on efficient coding theory have suggested that the relative arrangement of various cell types is likely optimized to enhance the encoding of stimulus information [77]. Recent research [148] showed that the arrangement of ON and OFF parasol (primate) and alpha (mouse) mosaics is not random but anti-aligned.

1.4 The collective behavior of retinal ganglion cells

Retinal ganglion cells are notorious for responding to visual stimuli in a substantially correlated manner. Within and across cell types, ganglion cells have a strong tendency to co-fire differently than what would be expected by chance [143, 153]. Correlated spiking has been characterized in the retina of many different mammals (rat [167], mouse [149, 124], primate [160, 159], cat [107, 108]). The striking conservation of this phenomenon across species suggests that it serves an important purpose in retinal information processing.

There are two main sources that shape the collective behavior of retinal neurons: on

the one hand ganglion cells respond to a common input, the stimulus, and the correlations it contains can drive correlated neural activity. Second, neuronal interactions within the ganglion cell layer, as well as interactions with upstream cells can further correlate the population response. These two sources of correlations are very different in nature, and will have substantially different impacts on the responses of retinal neurons.

1.4.1 Stimulus induced correlations

As was already discussed in this chapter, natural stimuli contain strong spatiotemporal correlations [150, 4], and computations in the retina are hypothesized to reduce redundancy by decorrelating the retinal output. At odds with this idea, research showed that retinal ganglion cells only partially decorrelate their visual inputs [163, 136]. This means that substantial residual stimulus-induced correlations persist between the activities of retinal ganglion cells (see Fig. 1.7). These correlations are often termed signal (or stimulus) correlations, and arise as the combination of two different effects. First, the visual stimuli itself contains strong spatio-temporal correlations that may not be filtered out by the retina [136]. Second, the processing performed by the retina itself can further promote signal correlations. For instance, receptive fields of neighboring ganglion cells have been shown to overlap, even within cell types [43]. Further, correlated responses can arise to signal specific features of the visual stimulus like motion reversal [154], or motion speed [40]. Recently, gaze shifts have been shown to strongly correlate the responses of retinal ganglion cells [84].

1.4.2 Intrinsic interactions further promote correlations

Intrinsic interactions within the retinal network also drive the correlated activity of retinal ganglion cells. The impact of interactions on population activity is of very different nature than the stimulus. Whereas the stimulus promotes correlations in the deterministic part of retinal responses (their firing rate), interactions induce correlations of the stochastic part of neural responses. The intrinsic variability (or noise) that characterizes neural spiking is reshaped through network interactions to give rise to the so-called noise correlations.

Noise correlations in the activity of retinal ganglion cells has been extensively characterized during the last decades [135, 167, 65]. Research has shown that noise correlated activity can arise from two main different sources [187, 181, 184]: shared noisy input and direct neuronal couplings.

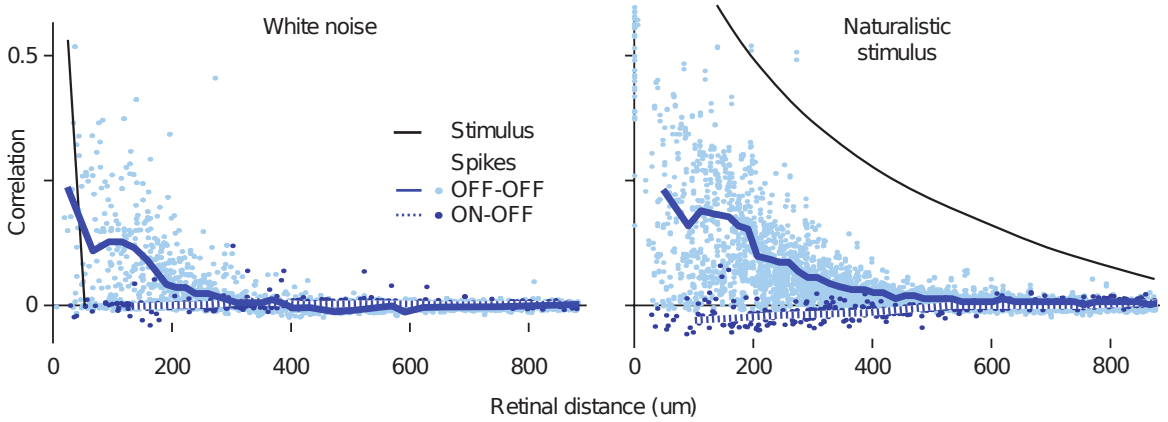

Figure 1.7: Ganglion cells from the salamander retina only partially decorrelate stimulus inputs (adapted from [163]). Black lines represent spatial correlations in the visual input, and blue lines correlations between retinal neurons of the same polarity (OFF cells). Left: Correlations between pairs of retinal ganglion cells are significant even in the absence of strong long range correlations in the stimulus (white noise). Right: For naturalistic stimuli, OFF ganglion cells significantly reduce correlations in their output compared to their input, down to a level that is close to the white noise case.

First, noise originating from photoreceptors is propagated across the retina. As this noise crosses the retinal layers, it becomes correlated due to network interactions while accumulating additional noise due to noisy synaptic transmission. In each layer of the retina, different neurons can receive common noisy inputs from upstream cells: bipolar cells receive synaptic inputs from multiple photoreceptors, likewise, retinal ganglion cells can receive inputs from several bipolar cells [106, 190]. In addition to this feed-forward noise propagation, horizontal and amacrine cells can further correlate the signal and its noise across the network. All in all, these common noisy inputs will result in retinal ganglion cells being noise correlated themselves, yielding what is sometimes termed "shared noise". Shared noise is significant both between pairs of cells of the same type and pairs of different types, and arises primarily in pairs with shared phtoreceptors input [181, 2].

Another prominent source of correlated variability arises at the ganglion cell layer. It is well-established that certain types of retinal ganglion cells are interconnected directly via gap junctions [186, 187]. These gap junctions function like small resistances, providing direct electrical coupling to neighboring ganglion cells of the same type [74, 180] as well as to some ganglion cells with amacrine cells [147]. Gap junctions induced noise correlations can be distinguished from shared noise by their short time-scale and typical double peaked cross-correlogram [25, 74, 187] (see Fig. 1.8 A). Interestingly, noise correlations induced by gap junctions are present in many different species (mouse [187], rabbit [74], rat [167], and primate [181]). This remarkable conservation across species in ortholog types of ganglion cells (α in mammals and parasol in primates), suggests fast noise correlations induced by

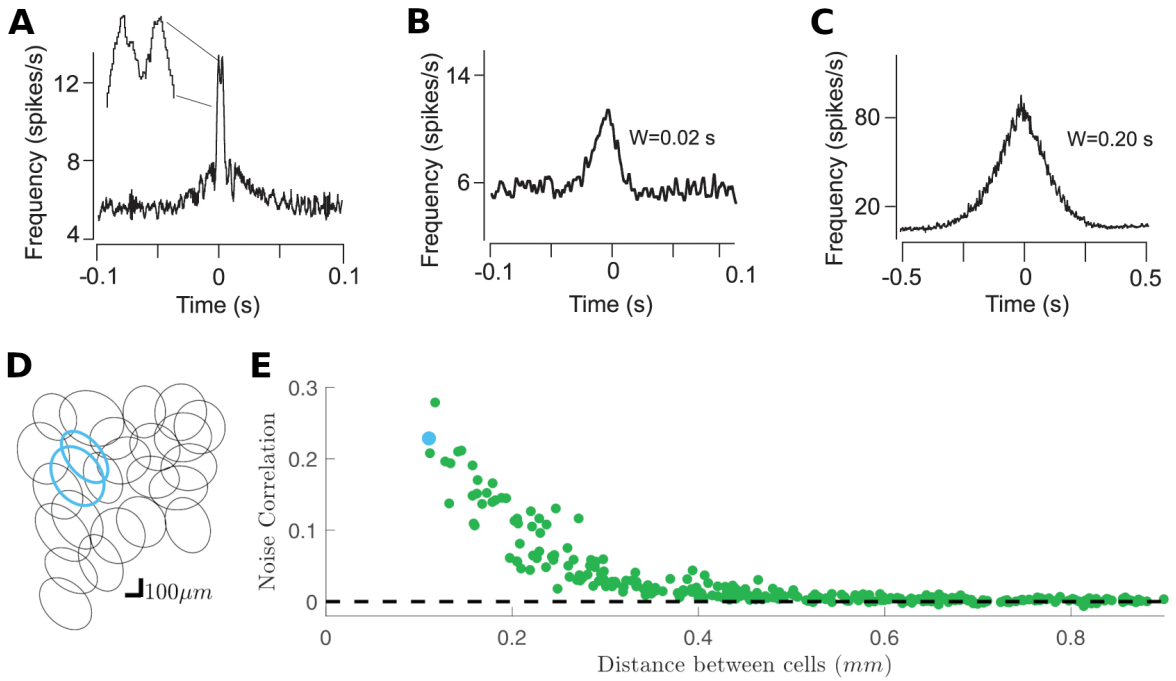

Figure 1.8: **A-C** Cross-correlograms between spontaneous spiking activity of pairs of mouse ganglion cells illustrate the diverse origins of noise correlations. (adapted from [187]) **A.** Cross-correlogram for an example pair of ganglion cells from the mouse retina. Direct couplings between the cells induce very fast correlations, with a typical double-peak centered on zero. **B.** In a second example pair, gap junctions between ganglion cells and amacrine cells mediate intermediate timing correlated spiking. **C.** For another pair, slow cross-correlations originate from both chemical and electrical synaptic inputs to the ganglion cells. **D,E** Illustration of the spatial dependency of noise correlations in a population ganglion cells from the rat retina (adapted from [167]). **D.** Partial mosaic of receptive fields from an OFF-α ganglion cells population. **E.** Noise correlations between pairs of OFF-α ganglion cells decrease exponentially with distance. The blue dot corresponds to the blue pair in the mosaic.

gap junctions may serve an important role in retinal processing.

1.5 Conclusion

We have seen in this chapter that the purpose of the retinal network is to process the visual stimulus to encode it into the collective behavior of its output layer, composed of retinal ganglion cells. In agreement with the efficient coding hypothesis, retinal computations seem to transform the stimulus to reduce redundancy in retinal outputs and increase the efficiency of this process.

The representation of visual stimuli in collective ganglion cell activity has a very specific organization: many different types of ganglion cells, each covering most of the visual field and encoding different features of the stimulus, relay parallel streams of information to the rest of the brain through the optic nerve. Ganglion cells activity is also very structured within each of these cell types, and neurons that encode information from neighboring locations in the visual field tend to be correlated by correlations in the stimulus and overlaps in their receptive fields.

We also saw that biological neurons are not deterministic: ganglion cell responses are noisy due to unreliable biological processes shaping the input they receive from upstream retinal layers, such as noisy synaptic release and photo-transduction in the photoreceptor layer. Like stimulus induced activity, neural noise is also correlated across the retinal network. In particular, within some ganglion cell types, shared synaptic inputs and direct electrical couplings via gap junctions result in noise correlations being strong for pairs of cells that also tend to be correlated by incoming stimuli. This phenomenon is broadly conserved across species, notably in mammalian α-cells and primate parasol cells, suggesting it may be an important feature of retinal computation.

Neural noise in ganglion cells activity plays a fundamental role in sensory processing: it limits reliable encoding of sensory signals by "blurring" them. To understand how retinal ganglion cells encode stimulus information in their collective activity, one therefore needs to understand how signal and noise are structured relatively to each other and how this structure impacts the encoded information. In the next chapter, we will delve on these two points. We will start by introducing measures of signal and noise correlations, then we will develop the main results of the literature on the impact of correlations on stimulus encoding, and finally describe the computational approaches that we will use in the rest of the thesis to model the collective behavior of retinal neurons.

Chapter 2

Population coding of stimulus information

A longstanding goal of neuroscience has been to crack the "neural code"—that is, to understand how stimulus related information is represented in the activity of sensory neurons. This problem is inherently complex because sensory coding cannot be fully understood by considering separately the activity of individual neurons: neural activity is characterized by its collective nature that emerges from stimulus input and neuronal interactions. This phenomenon has been widely observed in the retina, as discussed in the previous chapter [112, 161, 42], but also across a variety of other sensory systems such as the hippocampus [117], the somatosensory [139, 168], auditory [98, 48] and visual [199, 87, 165] cortices.

Characterizing the population activity of sensory neurons is challenging in itself, let alone understanding how this collective behavior relates to stimulus information encoding, due to the gigantic size of the "neural lexicon". Let's consider a set of neurons that can exist in one of two states: active or inactive. In this simplified case, the total number of possible states for a population of size N is 2^N. In reality, this problem is made even more complex by the fact that neural responses are dynamic and structured in time.

The retina is an ideal system to start investigating population coding, primarily thanks to its self-contained organization (it receives almost no feedback from the brain) and the accessibility of its output layer. These unique features allow recording *ex-vivo* from nearly complete local populations of neurons through electrophysiology, offering a sensible pip into collective ganglion cells activity. In addition, the retina is relatively well understood compared to other sensory centers such as the visual cortex. This understanding provides a solid ground to start relating the structure of population activity to retinal form and function.

The goal of this chapter is to give a general frame to the problem of population coding that is addressed in the rest of the thesis with a more specific emphasis on the retina. We will start by discussing ways to quantify the different sources that lead to the emergence of collective neural activity, i.e. stimulus and noise correlations. Second, we will review how the structure of this collective behavior is thought to impact stimulus information encoding. Finally, we will develop the computational tools that will be used throughout this thesis to investigate how neurons from the retina encode stimulus information.

2.1 Quantifying the collective behavior of sensory neurons

In order to study the impact of correlations on stimulus information encoding by sensory neurons, we first need to be able to characterize their collective behavior. The correlated activity of sensory neurons has been studied in the literature through the use of different quantities. Here we will define a set of correlation measures that will formalize our description of neural correlations. These quantities will be used throughout this thesis and will serve as the basis for our study of population coding.

2.1.1 Disentangling stimulus and noise correlations

Decomposition of marginal correlations To formalize our definition of signal correlations, let's consider the discretized activity of a population of N neurons that respond to a time varying stimulus. Let $R_i \mid S$ be the spike count random variable of neuron i in response to stimulus S. The Pearson correlation coefficient is a classical measure of linear correlations between two random variables. We call total correlations ρ_{ij}^{tot} the Pearson correlation between R_i and R_j, the responses of neurons i and j, described by the marginal distribution $P(R_i, R_j) = \langle P(R_i, R_j \mid S) \rangle_S$:

$$\rho_{ij}^{\text{tot}} = \frac{\text{Cov}\left(R_i, R_j\right)}{\sqrt{\text{Var}\left(R_i\right)\text{Var}\left(R_j\right)}}. \tag{2.1}$$

According to the law of total covariance, we can decompose the covariance $C_{ij}^{\text{tot}} = \text{Cov}\left(R_i, R_j\right)$ in stimulus and noise contributions C_{ij}^{s} and C_{ij}^{n}. If we note $\mu_i(S) = \overline{R_i(S)}$ the mean spike count of neuron i in response to stimulus s we get:

$$C_{ij}^{\text{tot}} = C_{ij}^{\text{s}} + C_{ij}^{n}, \tag{2.2}$$
$$C_{ij}^{\text{s}} = \text{Cov}\left(\mu_i(S), \mu_j(S)\right), \tag{2.3}$$
$$C_{ij}^{\text{n}} = \left\langle \text{Cov}\left(R_i, R_j \mid S\right)\right\rangle_S . \tag{2.4}$$

As a result, we can decompose total correlations in two contributions:

$$\rho_{ij}^{\text{tot}} = r_{ij}^{\text{s}} + r_{ij}^{\text{n}}, \tag{2.5}$$
$$r_{ij}^{\text{s}} = \frac{C_{ij}^{\text{s}}}{\sqrt{C_{ii}^{\text{tot}} C_{jj}^{\text{tot}}}}, \tag{2.6}$$
$$r_{ij}^{\text{n}} = \frac{C_{ij}^{\text{n}}}{\sqrt{C_{ii}^{\text{tot}} C_{jj}^{\text{tot}}}}, \tag{2.7}$$

where r_{ij}^{s} and r_{ij}^{n} correspond to the fractions of response correlations that are due respectively to stimulus and noise.

Signal (or stimulus) correlations The component of neural responses that corresponds to how the stimulus drives the neuron's activity is the mean spike count $\mu_i(S) = \overline{R_i(S)}$ (see Eq. 1.2). From this, a straightforward way to quantify signal correlations between cells i and j consists in taking the Pearson correlation between the two cells firing rates:

$$\rho_{ij}^{\text{s}} = \frac{C_{ij}^{\text{s}}}{\sqrt{C_{ii}^{\text{s}} C_{jj}^{\text{s}}}}. \tag{2.8}$$

Noise correlations To quantify noise correlations between cells i and j, we can therefore compute the Pearson correlation of their noise variables:

$$\rho_{ij}^{\text{n}} = \frac{C_{ij}^{\text{n}}}{\sqrt{C_{ii}^{\text{n}} C_{jj}^{\text{n}}}}. \tag{2.9}$$

This measure of noise correlations however, doesn't account for the fact that noise correlations can depend on stimulus, as it is based on the use of the average noise covariance C_{ij}^{n} defined in Eq. 2.4. The dependency of noise correlations on the stimulus is thought to impact the way information is encoded in population activity [57, 137]. To get a precise

picture of the structure of noise correlations, one can quantify them in a stimulus conditional manner:

$$\rho_{ij}^{\mathrm{n}}(S) = \frac{\mathrm{Cov}\left(R_i, R_j \mid S\right)}{\sqrt{\mathrm{Var}\left(R_i \mid S\right)\mathrm{Var}\left(R_j \mid S\right)}}. \tag{2.10}$$

2.1.2 Experimental quantification

Computing the aforementioned correlations from experimental data requires being able separate signal from noise in neuronal recordings. This can be done only by recording simultaneously the activity of a population of retinal ganglion cells to repeated presentations of a set of stimuli through electrophysiological recordings by multi-electrode arrays [104]. Repeating many times the presentation of each individual stimuli in the stimulus ensemble is what allows to disentangle signal from noise correlations: averaging neural activity over repeats amounts to averaging over the empirical neural noise distribution. This way, we can compute the mean spike counts of the neurons $\mu_i(S)$ as well as all the correlations described above.

In practice, estimating noise averaged quantities can be tricky as data are limited by experimental recording duration, and accurate estimation of firing rates and correlations may require from tens to hundreds of repetitions per stimuli. In particular, noise correlations will induce correlations in the residual noise of mean spike counts estimates, resulting in a bias in the empirical stimulus covariance estimates. This effect can be particularly striking in the case of neurons that are weakly modulated by the stimulus (small signal, large noise).

In the cortex, using repeated presentations of the stimulus isn't usually enough to separate intrinsic neural noise from neural signals. Any sensory cortical area receives inputs from other areas (sensory or even motor), and neural activity may be modulated by internal variables such as brain states and attention that are difficult to monitor accurately. These factors are not conditioned upon when repeating the presentation of the stimulus, which means that intrinsic neural noise that comes from the stochastic nature of neurons cannot be properly separated from varying inputs that originate from other part of the brain [46]. The nearly complete absence of feedback from the brain to the retina, combined with the ability to record from the ganglion cells layer *ex-vivo* in minimally disturbed retinal networks, establishes the retina as the ideal system for studying neural correlations and their impact on sensory coding.

2.2 The impact of correlated activity on stimulus encoding

Sensory neurons encode information about incoming stimuli in their collective activity. The accuracy of this encoding process will depend on the structure of the signal, that of the noise and how they relate to each other: the signal carries stimulus information while the noise corrupts it. In the current section, we will formalize this intuition by starting with simple geometrical considerations, following with an overview of the impact of noise correlations from the perspectives of estimation and information theory.

2.2.1 Geometrical picture

Noise in the response of sensory neurons will always hurt the representation of stimulus information. Due to noise, two stimuli may elicit similar responses from the population, while when averaged over noise, the mean spike counts may be different. Intuitively, noise is harmful because it blurs stimulus representations in the population activity and makes them overlap. In the simplest case of two neurons responding to two different stimuli, this can be pictured easily. The overlap between the responses to the stimuli is decreased compared to the uncorrelated noise case (Fig. 2.1 A) in the case where the direction of the noise is orthogonal to that of the signal (Fig. 2.1 B). Conversely, it is increased when the direction of the noise and signal are the same (Fig. 2.1 C). This phenomenon is sometimes called the sign-rule in the literature [75, 8]: when noise and signal correlations have the same sign, information decreases, but when they have opposite signs, information increases.

Real sensory neurons however, do not only encode pairs of stimuli, like illustrated in the simplistic example above: they respond to and code for large stimuli ensembles. To illustrate the impact of noise correlations on stimulus encoding in a more realistic setting, we can look at the response of a pair of neurons with mean responses specified by their overlapping receptive fields (see Fig. 2.1 D). In the example designed here, the pair's mean response lie along a manifold that encodes stimulus signal (see Fig. 2.1 E). In the case of positive noise correlations, omnipresent in sensory systems and in particular in the retina (see previous chapter), different effects will contribute to their overall impact on stimulus encoding. Locally, noise correlations may be detrimental or beneficial depending on whether they increase or decrease noise along the signal direction (see stimulus pairs (S_3,S_4) and (S_5,S_6) on the plot). But the effect of noise correlation on stimulus encoding isn't limited to local coordination with the signal direction. At a global scale, noise correlations also impact

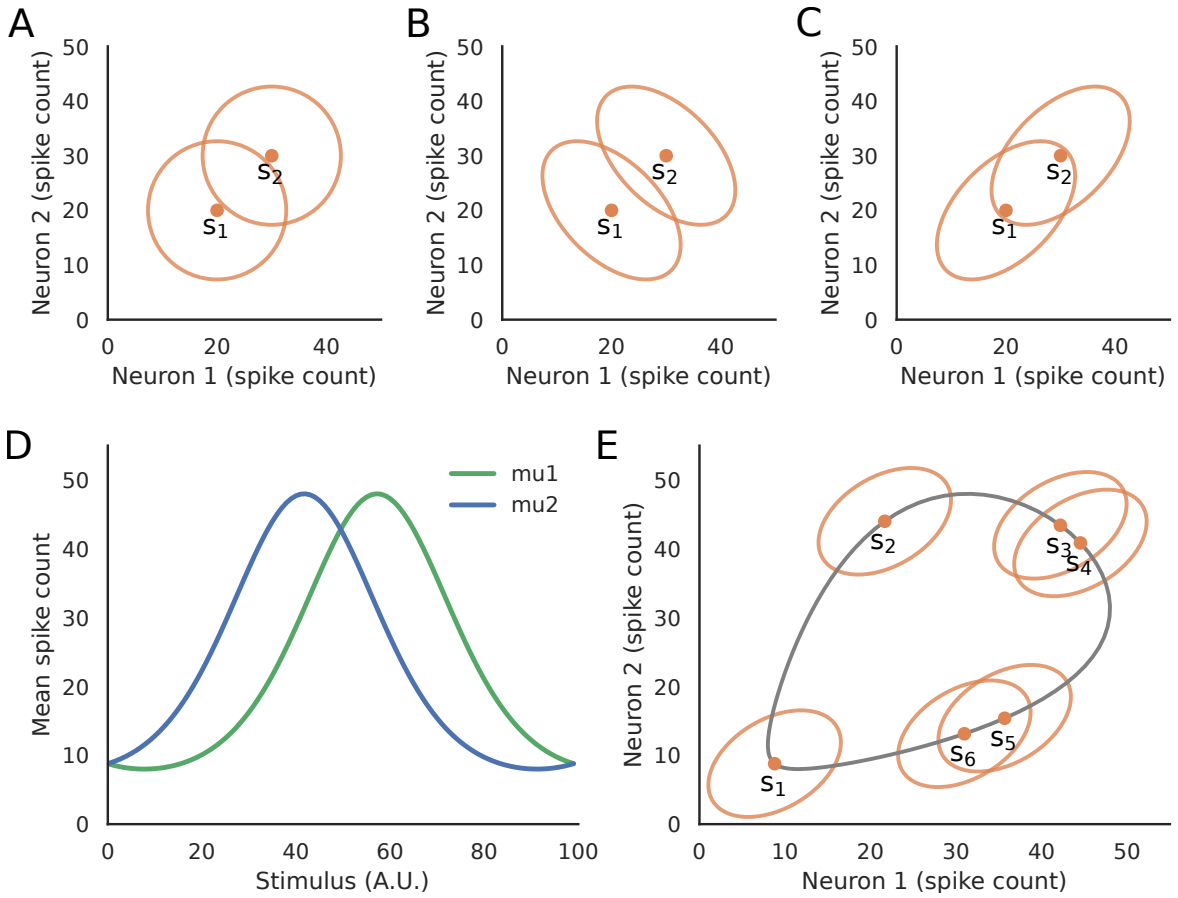

Figure 2.1: Geometrical illustration of the impact of noise correlations on stimulus coding. Responses are assumed to be Gaussian, and ellipses delimit the 2σ interval around the mean of pairwise responses conditioned on the stimulus. The simple case of two neurons responding to two different stimuli (S_1 and S_2) is sketched on panels **A-C**. In the absence of noise correlations (panel **A**), noise is isotropic at a given stimulus. Noise correlations are beneficial when they decrease the overlap between distributions compared to the uncorrelated case (panel **B**) and detrimental when they increase it (panel **C**). The case of two neurons responding to an ensemble of stimuli is sketched on panels **D,E**. The neurons spike according to their receptive fields defined by Von Mises functions $\mu_i(S) = A(\exp(\cos(S - c_i)) - e^{-1})/(e - e^{-1}) + B)$, where $A = 40$ and $B = 10$ set respectively the amplitude and the baseline of the mean spike count, and c_i the receptive field centers. In response space, the cells' mean activity (the signal) organizes along a manifold (grey line). Noise correlations now have different local and global effects on stimulus encoding. Locally, noise correlations can be beneficial (stimuli S_3, S_4) or detrimental (S_5, S_6), depending on how they align with the signal manifold. On a global scale, noise correlations can also be beneficial (i.e. stimuli S_2, S_5) or detrimental (i.e. S_1, S_3), depending on how they relate to the global structure of the manifold.

the discernability of stimuli lying on distant points on the manifold: while being detrimental for local stimulus pair (S_5,S_6), positive noise correlations allow better discrimination of pair (S_2,S_5) that lies on opposite sides of the manifold.

The picture outlined here gives an intuition about the impact of noise correlation on stimulus encoding. It remains however simplistic as it is based on simple geometrical considerations derived from a pair of sensory neurons encoding a simple stimulus into a very simple low dimensional representation. To fully understand how neural correlations, and in particular noise correlations, impact coding of complex stimuli in large populations of sensory neurons, one needs to first settle on a definition and a measure of what is information. Without such a measure, no definitive understanding of the phenomenon can arise.

2.2.2 Fisher information

2.2.2.1 Definition

Fisher information is a key concept of statistical estimation theory, a branch of statistics that deals with estimating the values of parameters underlying empirical data distributions. It measures the amount of information an observable random variable R carries about the parameter θ upon which the probability of R depends. In the context of sensory coding, the random variable R would represent the N dimensional population activity, while the unknown parameter θ (a scalar or a vector) would represent the stimulus that elicited neural responses.

The Fisher information of a parameter θ is defined as the variance of the score, which is the gradient (with respect to θ) of the log-likelihood function, evaluated in a particular point of parameter space θ^*. It is a local measure of information as it measures the information carried by R about θ in each point of parameter space. For a random variable R with probability density function $f(R;\theta)$, the Fisher information is given by:

$$\mathcal{F}(\theta) = \mathbb{E}_R\left[\left(\frac{\partial}{\partial\theta}\log f(R;\theta)\right)^2\right]. \tag{2.11}$$

Under few simplifying assumptions, its mathematical form becomes very tractable. Assuming that R is conditionally Gaussian such that $P(R \mid \theta) \sim \mathcal{N}(\mu(\theta), C^{\mathrm{n}}(\theta))$, with $\mu(\theta)$ the noise average of R and $C^{\mathrm{n}}(\theta)$ the covariance in θ, the Fisher information is often

approximated by:

$$F_{\text{lin}}(\theta) = \frac{\partial \mu(\theta)}{\partial \theta}^{\top} C^{\text{n}}(\theta)^{-1} \frac{\partial \mu(\theta)}{\partial \theta}. \tag{2.12}$$

This quantity is sometimes called the "linear Fisher" information. It coincides with the true Fisher information in the case of Gaussian responses with constant noise covariance. In the case where the covariance matrix depends on the stimulus θ, it is a lower bound to the true Gaussian Fisher information, which contains another positive term that accounts for the information provided by stimulus-dependent changes of the covariance matrix.

An important inequality that provides an intuitive interpretation of the Fisher information is the Cramér-Rao bound. Noting $\hat{\theta}(R)$ an unbiased estimator of θ such that $\mathbb{E}_R(\hat{\theta}(R)) = \theta$, we have:

$$\text{Var}_R(\hat{\theta}) \geq \frac{1}{\mathcal{F}(\theta)}. \tag{2.13}$$

This inequality states that no unbiased estimator of the parameter θ can have a variance smaller than the inverse of the Fisher information. In the context of sensory coding, it means that the inverse Fisher bounds the precision with which stimuli can be read out from population activity. Interestingly, the linear Fisher information defined previously corresponds to the inverse variance of an unbiased locally optimal *linear* decoder [19, 80].

2.2.2.2 Noise correlations and Fisher information

Due to its tractable form, Fisher information and its linear approximation have been the basis of numerous investigations of the impact of noise correlations on stimulus encoding [1, 166, 45, 80, 193, 194, 165, 116, 75, 57, 86]. Most theoretical and computational studies based on this quantity were inspired from the primary visual cortex, with large populations of neurons coding for a scalar stimulus angle. These studies mainly focused on the large population density limit, when a high number of tuning curves cover any single point of the stimulus ensemble (see Fig. 2.2A for an illustration) and considered constant noise correlations independent of the stimulus. The main conclusions of these studies are threefold:

- Positive uniform noise correlations (i.e. with infinite range) contribute positively to the Fisher information [1, 166, 193, 195].
- Positive short range noise correlations decrease the average Fisher information $\langle \mathcal{F}(\theta) \rangle_\theta$ compared to the independent case. In this case, the linear Fisher information even saturates with system size instead of scaling linearly as in the independent case (see Fig. 2.2B, left plot) [1, 166, 193, 45, 195].

- Introducing heterogeneity in tuning curves properties (see Fig. 2.2A right versus left) like amplitude and width may allow positive short range noise correlations to become beneficial if sufficiently strong (Fig. 2.2B right) [193, 45].

One way to get intuition about the linear Fisher information is to note that it corresponds to the squared norm of the vector of mean response derivatives $\mu'(\theta)$, computed using the covariance matrix as a metric (i.e. expressed in units of noise variance). This is akin to a local definition of the signal-to-noise ratio, where the local signal variation corresponds here to the derivative of the mean response with respect to the stimulus, and the noise to the amount of noise along the direction of the derivative vector. For the simpler case of constant positive noise covariance [1, 166], noise correlations will be beneficial for each stimulus value that corresponds to response derivatives $\mu'(\theta)$ located within specific fixed areas of the space of derivatives (see Fig. 2.2C for an illustration with two neurons). The distance between any points in these areas and the origin ($\mu' = 0$) will be larger when computed using the noise covariance matrix as a metric, compared to when using the diagonal noise variance matrix corresponding to the uncorrelated case as a metric (red areas in Fig. 2.2C). Let's consider a pair of cells with overlapping and similar tuning curves like shown in Fig. 2.1D. Mean response derivatives $\mu'_1(\theta)$ and $\mu'_2(\theta)$ will often have the same sign for certain ranges of stimulus values, leading to $\mu'(\theta)$ spending a significant amount of time within the detrimental areas (blue zones in Fig. 2.2C). In the case of non-overlapping cells with very distant tuning curves, the mean response derivatives will never be simultaneously large for both cells for any given stimulus. Either one or both cells will have a very small response derivative, which will result in the derivative $\mu'(\theta)$ being close to the axes of the space and thus in the beneficial zones (red areas in Fig. 2.2C encompass the vertical and horizontal axes). This intuition extends to large homogeneous neural populations such as the one depicted in Fig. 2.2A. In the case of positive short range noise correlations, the Fisher information will decrease compared to the uncorrelated case, as only the neighboring neurons with very similar tuning curves will be noise correlated. By contrast, in the case of uniform correlations, distant pairs of cells will also be noise correlated, leading to beneficial effects that may overall increase the Fisher information.

Another prominent case studied in the literature concerns a kind of stimulus dependent correlations termed "information-limiting" or "differential" correlations [116, 81]. Recent research has shown the presence of such information-limiting correlations in the cortex [18, 78]. These correlations correspond to a non-constant noise covariance matrix that always aligns the direction of the noise with that of the mean response gradient. An example of such correlations is illustrated in Fig. 2.2D. As a result, the Fisher information will always be decreased by the presence of differential correlations, compared to the independent

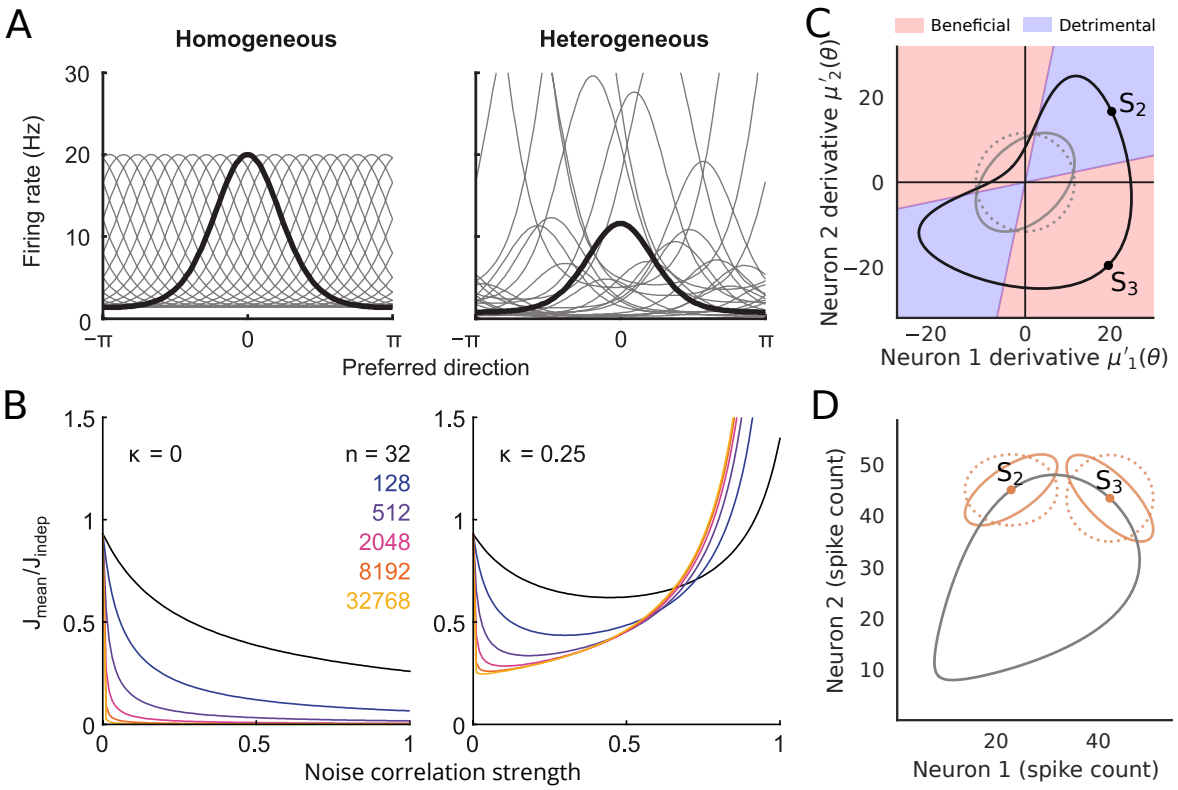

Figure 2.2: Overview of the impact of noise correlation on Fisher information. **A-B**, adapted from [45]. The Fisher information has been mostly studied in the case of systems inspired from the visual cortex. The neurons' receptive fields cover densely the stimulus ensemble (here an angular variable), such that any point in the stimulus ensemble is covered by many neurons. The case of homogeneous (left in **A**) and heterogeneous (right in **A**) tuning curves is shown here. Panel **B** shows the ratio of linear Fisher J_{mean} over uncorrelated Fisher J_{indep} for homogeneous (left) and heterogeneous populations (right), versus the noise correlation strength, and for different population sizes (n). In the homogeneous case, noise correlations always decrease information compared to the independent case, while in the heterogeneous case, strong noise correlations are beneficial. **C** Here we show (black line) the response derivative $\mu'(\theta)$ of the pair from Fig. 2.1D,E. The grey dotted circle corresponds to the two standard deviations ellipse of the variance matrix, and the grey ellipse to that of the covariance matrix (correlation strength $\rho^{\text{n}} = 0.4$ and noise variance of each cell $C^{\text{n}}_{ii} = 15$). Noise correlations increase the Fisher in any point of the red area of this space and decrease it in the blue area. These areas are delimited by lines passing by the intersections of the covariance and variance ellipses. **D** Information-limiting correlations are such that they align the noise with the local direction of the signal (the derivative of the response) for any given stimulus value. Therefore, they are always detrimental to the Fisher information.

case with same noise variance. Interestingly, it has been shown that some types of retinal ganglion cells exhibit noise correlations that vary with the stimulus in a way orthogonal to that of information-limiting correlations. Direction selective ganglion cells from rat and mouse retinas show noise correlations that make the noise orthogonal to the direction of the local signal, resulting in significant increase of the Fisher information [57, 200].

2.2.2.3 Limitations and perspectives

The Fisher information has been extensively used in neuroscience to evaluate the impact of noise correlations on sensory coding. Despite the insight it provides, this measure of information has specificities that limits its usability and interpretability.

First, quantifying the Fisher information requires having a parametric model of stimulus dependent responses. Estimating such a model from neural data tends to be challenging. This issue can be tackled in the case of low dimensional stimuli like the angular stimuli presented in the previous sections, but these stimuli are mostly relevant for the visual cortex and direction selective cells in the retina. For more relevant high-dimensional stimuli like naturalistic images in the case of retinal ganglion cells, solving this task starts to be challenging.

Another limitation of the Fisher information is that it measures the local sensitivity of neural responses to small changes in the stimulus. The Fisher information doesn't account for the existence of the stimulus ensemble, and is thus oblivious to the global effects that were highlighted in the first part of this section where we discussed a geometrical view on noise correlations. In the regime where neural noise is low compared to the strength of the signal, these global effects may be negligible and coding may be dominated by local statistics [28]. In the case where the neural noise is significant however, global statistics of the stimulus ensemble may have to be taken into account, as different stimuli can yield overlapping noisy responses.

To summarize, while the Fisher information can provide valuable insights on the effect of noise correlations on sensory coding, its intrinsic limitations make it likely to give a limited account of the full phenomenon. In the next section, we introduce the mutual information that will be used throughout the thesis to assess sensory coding and which palliates to some of the limitations of the Fisher.

2.2.3 Mutual information

2.2.3.1 Definition

In the late 40s, Claude Shannon laid the basis of information theory [156] and developed a mathematical framework providing grounded definitions of quantities such as the surprisal of a random process outcome and information. The entropy of a discrete random variable R with ensemble of possible outcomes $\mathcal{R}$ and probability distribution P is defined as:

$$H(R) = \left\langle \log \frac{1}{P(R)} \right\rangle_R \tag{2.14}$$

$$= -\sum_{R \in \mathcal{R}} P(R) \log P(R). \tag{2.15}$$

This quantity measures the uncertainty inherent to the outcome of process R, as it is the average of the surprisal $\mathcal{S}(R) = \log 1/P(R)$, a quantity that is high for unlikely values of R and low for its common values. The entropy of a random variable is expressed either in bits or nats depending on whether the logarithm used for its calculation is in base 2 or in base 10.

Mutual information is to information theory what Fisher information is to estimation theory. It measures how much information the observation of one random variable provides about another variable. Let our two random variables be R and S. From the intuition that a gain of information corresponds to a decrease of uncertainty and using the notion of entropy defined above, one recovers the definition of the mutual information:

$$I(R;S) = H(R) - \langle H(R \mid S) \rangle_S \,. \tag{2.16}$$

The mutual information is symmetric, and is expressed in bits or nats, like the entropy. Contrary to the Fisher information, the mutual information is a global measure of information. It accounts for the entire ensembles of both stimuli and responses, and integrates the structure of these ensembles as well as their statistical dependencies at all orders.

Connections between the mutual information and the Fisher information have been made in the literature, with a focus on sensory neuroscience [26, 166, 192]. When S is a one dimensional random stimulus variable and R a sensory measurement or neural response, it has been shown that the mutual information can be approximated by the Fisher information in the small Gaussian noise limit. This is in accordance with the considerations outlined in the previous section: in the small noise limit, coding is dominated by local effects and the Fisher information therefore gives a good account of sensory coding. When the noise is not

small this connection breaks down as both local and global effects accounted for by the mutual information but not by the Fisher impact sensory coding.

2.2.3.2 The effects of noise correlations on mutual information

Compared to the Fisher information, the mutual information has a less tractable form that hinders its use to study sensory coding. While the Fisher information is derived solely from the conditional response distribution, understanding the mutual information requires both conditional response and marginal response distributions to be tractable. Nonetheless, several theoretical studies investigated the impact of noise correlations on stimulus information encoding [128, 166, 137, 93, 178, 75, 117].

The impact of noise correlations on the mutual information can be evaluated through one quantity that we will call noise synergy and note ΔI in the rest of the thesis. Noting $I(R;S)$ the mutual information (defined in Eq. 2.16) between the stimulus S and the response R in the presence of noise correlations, and $I_{\text{ind}}(R;S)$ the mutual information when noise correlations are removed, we have:

$$\Delta I = I(R;S) - I_{\text{ind}}(R;S). \tag{2.17}$$

Concretely, removing noise correlations consists in assuming that the conditional response distribution $P(R \mid S)$ gets replaced by $P_{\text{ind}}(R \mid S) = \prod_i P(R_i \mid S)$, where $P(R_i \mid S)$ is the conditional response of the i-th neuron from the population. We thus assume that neurons are conditionally independent, all other things being equal.

The current common wisdom in sensory neuroscience about this can be summarized through a decomposition of the mutual information in correlation contributions [137]. This decomposition of the mutual information constitutes and exact and model free version of a previous small-firing rate approximation [128] that overall gives the same insights. This general decomposition has the advantage of making no assumptions on the underlying noise and response models, and decomposes the noise synergy in two components:

$$\Delta I = I_{\text{corr-ind}} + I_{\text{corr-dep}}. \tag{2.18}$$

Noting $P(R_i)$ the marginal and $P(R_i \mid S)$ the conditional responses of neuron i, $P(R) = \langle P(R \mid S) \rangle_{\text{S}}$ the marginal population response distribution and $P_{\text{ind}}(R) = \langle \prod_i P(R_i \mid S) \rangle_{\text{S}}$ the marginal population response distribution in the absence of noise correlations, we can

express the terms of the above decomposition as [137]:

$$I_{\text{corr-ind}} = \sum_R \left[P\left(R\right) - P_{\text{ind}}\left(R\right)\right] \log\left(\frac{\prod_i P\left(R_i\right)}{P_{\text{ind}}(R)}\right), \tag{2.19}$$

$$I_{\text{corr-dep}} = \left\langle \sum_R P\left(R \mid S\right) \log\left(\frac{P\left(R \mid S\right)/P_{\text{ind}}\left(R \mid S\right)}{\left\langle P\left(R \mid S'\right)\right\rangle_{S'} / \left\langle P_{\text{ind}}\left(R \mid S'\right)\right\rangle_{S'}}\right)\right\rangle_S. \tag{2.20}$$

In this decomposition, the effect of correlations can be gauged by recalling that the difference between conditional distributions $P\left(R \mid S\right)$ and $P_{\text{ind}}\left(R \mid S\right)$ on the one hand and marginal distributions $P\left(R\right)$ and $P_{\text{ind}}\left(R\right)$ on the other hand are solely due to noise correlations. From this, the two aforementioned components can be shown to describe two different effects of noise correlations on mutual information:

- $I_{\text{corr-ind}}$ specifically captures the effect of noise correlations on the marginal response distribution, that is the effect of noise correlations, irrespective of how they vary with the stimulus. If noise correlations increase the probability of a certain response patterns R, then $P\left(R\right) - P_{\text{ind}}\left(R\right) \geq 0$. On the other hand, if signal correlations increase the probability of a certain response patterns R, then $\log\left(\prod_i P\left(R_i\right)/P_{\text{ind}}(R)\right) \leq 0$. As a result, if noise and signal correlations promote the same patterns, then overall $I_{\text{corr-ind}}$ will be negative, while it will be positive otherwise. This phenomenon is sometimes called the sign-rule in the literature [75, 8] and captures the intuition outlined in the previous sections that when correlations align noise and signal, they are harmful to coding (see Figures 2.1 and 2.2).

- $I_{\text{corr-dep}}$ is the term that accounts for the stimulus-dependence of noise correlations. This term is always positive and is usually interpreted as the information carried by the fluctuations of noise correlations themselves (here the fluctuations of $P\left(R \mid S\right)/P_{\text{ind}}\left(R \mid S\right)$).

Interestingly, we can rephrase $I_{\text{corr-dep}}$ and $I_{\text{corr-indep}}$ in a way that leads to the emergence of another quantity:

$$I_{\text{corr-ind}} = I_{\text{cross}}(R;S) - I_{\text{ind}}(R;S), \tag{2.21}$$

$$I_{\text{corr-dep}} = I(R;S) - I_{\text{cross}}(R;S). \tag{2.22}$$

If we define the cross-entropy between two probability distributions P and Q as

$$H\left[P,Q\right] = -\sum_R P(R) \log Q(R), \tag{2.23}$$

the quantity $I_{\text{cross}}(R;S)$ is given by:

$$I_{\text{cross}}(R;S) = \mathrm{H}\left[P(R), P_{\text{ind}}(R)\right] - \left\langle \mathrm{H}\left[P(R \mid S), P_{\text{ind}}(R \mid S)\right]\right\rangle_S , \tag{2.24}$$

$$= \left\langle \sum_R P(R \mid S)) \log\left(\frac{P_0(R \mid S))}{P_0(R)}\right)\right\rangle_{\text{s}} . \tag{2.25}$$

I_{cross} resembles a mutual information, with the important difference that it is built from the cross-entropy between the dependent and independent distributions instead of the entropy of one or the other. Therefore, this quantity can be seen as the information available to an ideal decoder that ignores noise correlations, and it has been shown in the literature that the quantity $I_{\text{corr-dep}} = I(R;S) - I_{\text{cross}}(R;S)$ is an upper bound on the information lost by such a decoder [124, 93].

2.2.3.3 Perspectives

The aforementioned breakdown of mutual information summarizes and formalizes various observations and considerations that constitute the common wisdom found in the literature. Namely, noise correlations are thought to be beneficial for coding whenever they oppose stimulus correlations and follow the sign-rule, as well as when they vary with the stimulus. To some extent, this picture seem to generalize the conclusions drawn from the Fisher information, for which noise correlations are detrimental when they constrain noise in the direction of the signal. This phenomenon is akin to a violation of a local version of the "sign-rule" evoked previously and was exemplified in the case of population coding by densely overlapping tuning curves (inducing positive stimulus correlations) and positive short range noise correlations.

The apparent cohesion that emerges from the theoretical literature on the informational effects of noise correlations and in particular on the sign-rule seem to be at odds with the overwhelming experimental evidence showing both in retina [112, 161, 187, 167, 107] and cortex [199, 94, 13, 87, 73, 7] that noise correlations are particularly strong and positive for neurons that also have positive stimulus correlations.

Pinning down this contradiction or understanding where this discrepancy originates from would require being able to straightforwardly relate the theoretical frameworks outlined above to experimental measures of the structure of stimulus and noise correlations. Such attempts in the case of the Fisher information have been limited to low dimensional stimuli [57, 200, 70, 86, 165] for reasons that were discussed in a previous section.

Several studies have approached this problem using the mutual information [137, 93, 178,

75, 124, 117]. However, the reliance of the mutual information on estimation of probability distributions greatly hinders its applicability but even more its interpretability. Beyond the difficulties that stem from estimating response distributions from data, the formulation of the mutual information prevents relating structures of correlations to their impact on coding. For instance, the aforementioned mutual information breakdown has been applied to experimental data [114, 123], but the measures of correlations as comparisons of correlated and uncorrelated probability distributions on which it is based render its interpretation difficult. This is due to the fact that the behavior of such measures do not align straightforwardly with pairwise correlations or covariances. As an example, constant pairwise noise covariances do not imply that the ratio $P\left(R \mid S\right) / P_{\mathrm{ind}}\left(R \mid S\right)$ (from Eq. 2.20) is constant, even for pairwise response distributions.

The goal of this thesis is to address these points by filling the gap that exist in the literature on relations between mutual information and structures of noise and signal correlations. We will try to develop a framework that provides theoretical understanding of the impact of noise and signal correlations on the noise synergy ΔI, which can also be related straightforwardly to experimental data.

2.3 Modelling the collective behavior of sensory neurons

We have seen in the previous section that a grounded way to understand sensory coding and how noise correlations impact it consists in quantifying information theoretic quantities such as the mutual information. To do so one has to estimate the distributions of population responses R to the stimulus S. This problem is rendered difficult by the fact that the population activity R can be high-dimensional. For a population of N neurons that can each exist in q different states of activity, the probability distribution of the response is described by $q^N - 1$ parameters. Estimating directly and reliably the probability distributions of responses for each given stimuli requires having amounts of data that are difficult to gather in reasonable experimental time. A solution to circumvent this problem consists of building response models that are characterized by less parameters, but which still provide a good description of the data after fitting.

The goal of this section is to describe the computational tools that will be used and built upon in the rest of the thesis. We will start by describing a maximum entropy approach to the modeling of neural responses, following by an introduction of the Generalized Linear Model and how it can be used to model populations of sensory neurons.

2.3.1 Maximum entropy approach

2.3.1.1 Model description

Estimating accurately the probability distributions of neural responses from limited experimental recordings in the case of large neural populations is a difficult problem, especially in a model free manner. One way to curb this problem is to make careful assumptions (impose inductive biases) on the structure of the model in order to reduce the number of parameters to infer, while still capturing the gist of the data statistics. Maximum entropy (maxent) models are a wide category of models that are built by maximizing the entropy under constraints that will enforce reproduction of a few selected empirical statistics (the observables) from the data.

Lets consider a network of N binary neurons with response vector R that follows some empirical distribution P_{data}. Let $\mu_i^{\text{data}} = \langle R_i \rangle_{\text{data}}$ and $C_{ij}^{\text{data}} = \langle R_i R_j \rangle_{\text{data}} - \mu_i^{\text{data}} \mu_j^{\text{data}}$ denote respectively the mean and covariances computed from empirical data that we wish to model. Noting $P(R)$ the model probability of response R, the pairwise maxent model can be derived by maximizing the Lagrangian:

$$\begin{aligned}\mathcal{L} = H(R) &+ \sum_i h_i \left(\sum_R P(R) R_i - \mu_i^{data} \right) \\ &+ \sum_{i<j} J_{ij} \left(\sum_R P(R) R_i R_j - (C_{ij}^{\text{data}} + \mu_i{}^{\text{data}} \mu_j{}^{\text{data}}) \right) \\ &+ \gamma \left(\sum_R P(R) - 1 \right),\end{aligned} \tag{2.26}$$

where $H(R) = -\sum_R P(R) \log(P(R))$ is the entropy of the model, with Lagrange multipliers h_i and J_{ij} enforcing reproduction of the empirical means and covariance, while γ enforces normalization of the probability distribution. This results in the probability distribution:

$$P(R) = \frac{1}{Z(\{h_i\}, \{J_{ij}\})} \exp\left(\sum_i h_i R_i + \sum_{i<j} J_{ij} R_i R_j \right), \tag{2.27}$$

where $Z(\{h_i\}, \{J_{ij}\}) = \sum_R \exp\left(\sum_i h_i R_i + \sum_{i<j} J_{ij} R_i R_j \right)$ normalizes the probability distribution. Parameters h_i and J_{ij} are often called respectively magnetic fields and couplings in reference to the Ising model of statistical physics.

2.3.1.2 Inverse problem

The process of optimizing the parameters of the maxent model from empirical data to reproduce the set of chosen observables is called the inverse problem. This inference can be tackled in different ways, and significant effort has devoted to this problem in the literature.

Maximum likelihood estimation The most straightforward way to optimize the parameters of the pairwise maxent model consists in maximizing the log-likelihood of the empirical data given parameters of the model. Let us try to model the probability distribution of R the binary $(0, 1)$ multivariate neural response. Assuming we have access to T data samples of the distribution, the observed data are $\{R^{(t)}\}_{t=1,T}$ where t denotes the sample index. Noting $\mu_i^{\text{data}} = \langle R_i \rangle_t$ the empirical mean response and $C_{ij}^{\text{data}} = \langle R_i R_j \rangle_t - \mu_i^{\text{data}} \mu_j^{\text{data}}$ empirical covariances, the log-likelihood of the observed dataset is given by:

$$\begin{aligned}\log \ell(\{h_i\}, \{J_{ij}\}; \{R^{(t)}\}) &= \sum_{t=1}^{T} \left(\sum_i h_i R_i^{(t)} + \sum_{i<j} J_{ij} R_i^{(t)} R_j^{(t)} - \log Z(\{h_i\}, \{J_{ij}\}) \right), \\ &= T \left(\sum_i h_i \mu_i^{\text{data}} + \sum_{i<j} J_{ij} (C_{ij}^{\text{data}} + \mu_i^{\text{data}} \mu_j^{\text{data}}) - \log Z(\{h_i\}, \{J_{ij}\}) \right).\end{aligned} \tag{2.28}$$

The log-likelihood can be related directly to the cross-entropy, another quantity of interest that also quantifies the match between empirical and model distributions. Recalling its definition Eq. 2.23 and noting P_{data} the empirical data distribution we have:

$$\log \ell(\{h_i\}, \{J_{ij}\}; \{R^{(t)}\}) = -T \cdot H(P_{\text{data}}, P). \tag{2.29}$$

Inference of model parameters that match predicted and empirical observables can thus be done equivalently by maximizing the log-likelihood or minimizing the cross-entropy between data and model. In the case were the model predicts accurately empirial means and covariances, the cross-entropy $H(P_{\text{data}}, P)$ is minimized and becomes equal to the entropy of the model $H(R)$.

To maximize the log-likelihood, we start by computing its gradient with respect to the parameters h_i and J_{ij}. Calling $\mu_i = \sum_R P(R) R_i$ and $C_{ij} = \sum_R P(R) R_i R_i - \mu_i \mu_j$ the model predicted means and covariances, gradients are given by:

$$\frac{\partial \log \ell(\{h_i\}, \{J_{ij}\}; \{R^{(t)}\})}{\partial h_i} = T \left(\mu_i^{\text{data}} - \mu_i \right), \tag{2.30}$$

and

$$\frac{\partial \log \ell(\{h_i\},\{J_{ij}\};\{R^{(t)}\})}{\partial J_{ij}} = T\left(C_{ij}^{\text{data}} - C_{ij} + \mu_i^{\text{data}}\mu_j^{\text{data}} - \mu_i\mu_j\right). \tag{2.31}$$

Optimizing the log-likelihood function with respect to the fields and couplings will therefore ensure reproduction of the empirical means and covariances. This optimization can be simply performed by performing gradient ascent of the log-likelihood using the gradients derived above.

Approximate inference and small-correlation expansion Solving the inverse problem by maximizing the log-likelihood above Eq. 2.28 is doable analytically for very small groups of neurons, but quickly becomes intractable as the population size increases. For larger populations of up to ≈ 20 neurons, exact maximization of the likelihood is feasible numerically by the exact gradient ascent approach described in the previous section. Beyond ≈ 20 neurons however ($\approx 10^6$ configurations), it becomes difficult to do things exactly, and one has to resort to monte-carlo simulations to estimate numerically the model's observables in the right hand side of Eq. 2.30 and 2.31 and solve the optimization. Another strategy that has been pursued in the literature has been to search for approximate solutions to the inverse problem. Many of these approximate solutions were derived in the context of statistical mechanics, and such approaches include mean-field [82, 172], Thouless-Anderson-Palmer [175] solutions and small-correlation expansions [155].

The Sessak-Monasson expansion [155] is a perturbative approach that allows to estimate the parameters of the maxent model under the assumptions that the covariances C_{ij} are small. The starting point of this approximation is the minimization of the cross-entropy between the empirical data and the model distributions $H(P_{\text{data}}, P)$, which is equivalent to the maximization of the log-likelihood (see Eq. 2.29). Noting again μ_i^{data} and C_{ij}^{data} the empirical observables, this cross-entropy can be written:

$$H(P_{\text{data}}, P) = -\sum_i h_i\mu_i^{\text{data}} - \sum_{i<j} J_{ij}(C_{ij}^{\text{data}} + \mu_i^{\text{data}}\mu_j^{\text{data}}) + \log Z(\{h_i\},\{J_{ij}\}). \tag{2.32}$$

As already seen in the previous section, when the fields h_i and couplings J_{ij} minimize effectively this cross-entropy, the models' means and covariances coincide with that of the data, and the minimum of the cross-entropy equals the entropy of the fitted model.

The main trick of the expansion consists of replacing all covariances C_{ij} by βC_{ij}, where β is a positive scaling parameter. As a result, the optimal fields and couplings of the above equation will now depend on this parameter and when $\beta = 0$, the solution will correspond to independent binary variables with null couplings. The gist of the derivation consists

of performing a Taylor expansion of the cross-entropy with scaled covariances at small β. Evaluating this expansion in $\beta = 1$ results in approximate solutions for the optimal fields h_i and couplings J_{ij}. Keeping in mind that the entropy $H(R)$ of the maxent distribution $P(R)$ is *in fine* given by the optimized cross-entropy, the small-correlation expansion evaluated in $\beta = 1$ also gives an approximation of the maxent entropy. In the case of binary variables $R_i = 0, 1$, this approximation can be written as:

$$\begin{aligned} H(R) = &- \sum_i (1-\mu_i)\log(1-\mu_i) + \mu_i \log(\mu_i) \\ &- \frac{1}{2}\sum_{i<j} \frac{C_{i,j}^2}{\Gamma_i^{(1)}\Gamma_j^{(1)}} \\ &+ \frac{1}{6}\sum_{i<j} \frac{C_{i,j}^3}{(\Gamma_i^{(1)}\Gamma_j^{(1)})^3}\Gamma_i^{(2)}\Gamma_j^{(2)} \\ &+ 1\sum_{i<j<k} \frac{C_{i,j}C_{j,k}C_{k,i}}{\Gamma_i^{(1)}\Gamma_j^{(1)}\Gamma_k^{(1)}}, \end{aligned} \tag{2.33}$$

where:

$$\Gamma_i^{(1)} = \left\langle (R_i - \mu_i)^2 \right\rangle = \mu_i(1-\mu_i), \tag{2.34}$$

and

$$\Gamma_i^{(2)} = \left\langle (R_i - \mu_i)^3 \right\rangle = \mu_i(1-\mu_i)(1-2\mu_i). \tag{2.35}$$

2.3.1.3 Application to data

The pairwise maxent model has proven to be a data-efficient tool to model the collective behavior of biological neurons [152]. It has been applied successfully to model populations from the retina [153, 160, 176, 51, 39], as well as from the cortex [105, 174, 113, 122]. In previous sections, we showed how this approach could be used to describe an arbitrary distribution of the population activity of N binary neurons, without worrying about the nature of this distribution. The maxent approach can indeed be used to characterize the collective behavior of populations of neurons through the inferred coupling network, or to estimate neural activity entropies as well as other information theoretic quantities. In practice, maxent models can be applied to describe both stimulus conditioned response distributions $R \mid S$ and marginal response distributions R alike. Interpretation of the inferred models however, differs significantly in these two settings. Applying the maxent model to conditional neural responses $R \mid S$ yields couplings J_{ij} that solely account for noise correlations and describe the effective connectivity of the population. By contrast, when modeling the marginal distribution R, couplings will account for both noise and stimulus

induced correlations, giving no information about how these two sources of correlations differently impact the collective behavior. Further, maxent models can be seen as "static" models and their interpretation depends on the temporal preprocessing of the data on which they were fitted. If empirical data are binned at size $\Delta t = 20ms$, the model will solely account for fast neural correlations that exist across the network within that bin size while correlations within the network across time bins will be lost. As a result, network couplings or information theoretic quantities computed from these models always need to be interpreted in light of the temporal properties of the underlying experimental data.

2.3.2 Generalized Linear Model

A different approach to modeling the response of sensory neurons, and in particular how they respond to time varying input stimuli, is to use a "dynamic" model such as the Generalized Linear Model (GLM) already mentioned in section 1.2. Unlike the pairwise maxent model which describes statically a given conditional or marginal response distribution, the GLM models neural responses based on past stimulus values as well as past population responses, and is used to simulate dynamically, time bin after time bin, neural responses to input stimuli.

2.3.2.1 Model description

We introduced in section 1.2 a single neuron stimulus encoding model for retinal ganglion cells, to illustrate the main characteristics of retinal responses. The independent model we developed accounted for stimulus induced activity and past activity effects. To model noise correlations, one as to account for how the past activity of all neurons in the population impact the activity of each neuron. This can be done by extending the model detailed in Fig. 1.5 and defined in Eq. 1.6 by adding a set of couplings filters to the model [182, 135]. Let $J_{ij}(\tau)$ denote the coupling filter that accounts for the impact of cell j on cell i, and $J_{ii}(\tau)$ the self-coupling (or spike-history) filter that accounts for the cell's own past effects (see Fig. 2.3A for a schematic). We note $R_i^{(t)}$ the spike count variable of neuron i in time bin t and $\nu_i(t)$ the parameter of the random (Poisson or Bernoulli) process that yields $R_i^{(t)}$. As seen in section 1.2, the neuron responds to a stimulus movie S, and its response function is characterized by a spatio-temporal filter K_i, a bias b_i and a nonlinearity $\mathcal{F}$. The extended

population model is described by:

$$h^i_{\text{stim}}(t) = \sum_{\tau} \sum_{x,y} K_i(x, y, \tau) S(x, y, t - \tau), \tag{2.36}$$

$$h^i_{\text{past}}(t) = \sum_{i,j} \sum_{\tau} J_{ij}(\tau) R_j^{(t-\tau)} \Delta t, \tag{2.37}$$

$$\nu_i(t) = \mathcal{F}\left(b_i + h^i_{\text{stim}}(t) + h^i_{\text{past}}(t)\right). \tag{2.38}$$

Once the parameters of the model fitted to the data, simulations of the GLM should reproduce empirical cross-correlations. This approach has been shown to work well, in particular for ON and OFF parasol ganglion cells from the primate retina [135] (see reprint in Fig. 2.3B).

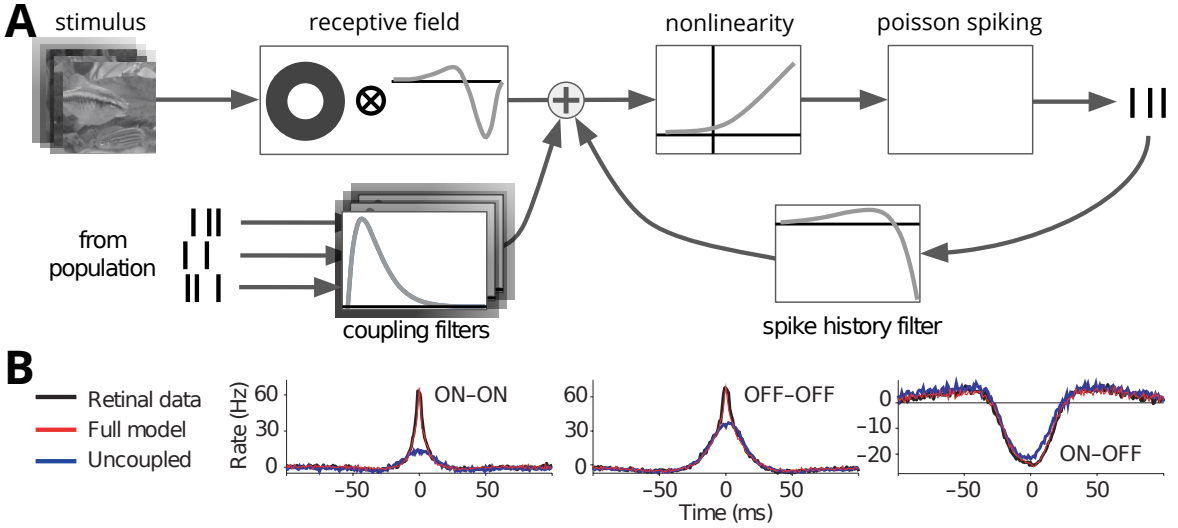

Figure 2.3: Generalized Linear Models (GLM) to describe the collective behavior of sensory neurons. **A** Schematic of the GLM, extended from Fig. 1.5. The activity of other cells in the population is now linearly processed by a set of couplings filters that will correlate neural variability across the system. **B** These couplings allow us to model the spikes cross-correlograms of neural responses (adapted from [135]). This plot shows three cross-correlograms for three example pairs of retinal ganglion cells from the rat retina. The first plot corresponds to a pair of ON cells, the second to OFF cells, and the third to a pair of ON and OFF cells. The uncoupled model (blue lines) captures only the part of cross-correlations that correspond to stimulus induced correlations. The coupled model (red) however, captures most of the remaining correlations in experimental data (black). These remaining correlations correspond to noise correlations, which are thus well accounted for by the coupled model.

We have seen in the last section that maximum entropy models have to be applied to spiking data with bin size Δt larger than the typical timescale of cross-correlations in order to integrate them effectively. By contrast, GLMs have to be used along with very small time bins (a few milliseconds maximum) to accurately describe empirical cross-correlations. This

is due to the fact that GLM couplings only account for the effect of past time bins on the current one, resulting in the activity of two different cells to be conditionally independent within the same time bin. To capture fast timed cross-correlations as the ones described in section 1.4.2, the GLM must therefore be applied to data binned at a resolution much higher than the latency of the fastest correlations present in the data.

2.3.2.2 Maximum likelihood estimation

The most common approach to fit the parameters of a GLM model is log-likelihood maximization. In the case of the GLM, we model the probability $P(R_i \mid t)$ of observing response R_i for neuron i at time t, given the past stimulus and the population response from the empirical data. We note the empirical response of neuron i as $\{R_i^{(t)}\}_{t=1,T}$, where t designates the time bin index of the sample. Given the aforementioned model definition and further assuming that the nonlinearity $\mathcal{F}$ is exponential and the noise process Poissonian, the log-likelihood of the GLM for neuron i becomes:

$$\log \ell_i \left(\{J_{ij}(\tau)\}, \{K_i(x,y,\tau)\}, \{b_i\}; \{R_i^{(t)}\} \right) = \sum_t R_i^{(t)} \log \left(\nu_i(t)\right) - \nu_i(t) - \log \left(R_i^{(t)}! \right). \tag{2.39}$$

Maximization of the GLM log-likelihood is straightforward if the chosen nonlinearity $\mathcal{F}$ is convex and log-concave, as in this case the log-likelihood function is concave [126] and optimization can be done by gradient ascent. Gradients of the log-likelihood with respect to the parameters of the model are given by:

$$\frac{\partial \log \ell_i}{\partial b_i} = \sum_t R_i(t) - \nu_i(t), \tag{2.40}$$

$$\frac{\partial \log \ell_i}{\partial K_i(x,y,\tau)} = \sum_t \left(R_i(t) - \nu_i(t)\right) S(x,y,t-\tau), \tag{2.41}$$

$$\frac{\partial \log \ell_i}{\partial J_{ij}(\tau)} = \sum_t \left(R_i(t) - \nu_i(t)\right) R_j(t-\tau). \tag{2.42}$$

These gradients illustrate how the bias b_i is here to enforce the mean spike count, while the stimulus filter K_i promotes reproduction of stimulus driven activity and couplings J_{ij} enforce that of cross-correlations.

We need to note here that this inference approach has some important caveats, as maximization of the above likelihood does not fully ensure reproduction of firing rates and cross-correlations during simulation of the inferred model. This can happen because of limitations of the model itself, for instance when stimulus-response relationships in the

data are too complex or nonlinear to be captured by the simple point-wise nonlinearity of the GLM. On top of this, as is shown by gradients Eq. 2.42, couplings enforce reproduction of total cross-correlations, but not separately of noise and stimulus correlations. A potential consequence of this limitation is that mismatch between predicted and empirical firing rates, which leads in mismatch between predicted and empirical stimulus correlations, can impact estimations of noise correlations. The model could use couplings to compensate for firing rates mismatch and eventually recover total correlations.

Another intrinsic limitation that may interact with the two previous points has to do with the fact that likelihood maximization is performed on the predicted Poisson parameter $\nu_i(t)$ given in Eq. 2.38, where $h^i_{\text{past}}(t)$ is computed with response history from the data, and not from response history predicted by simulation of the model. This can amplify the mismatch between predicted firing rates and empirical ones, affect inferred coupling values and eventually lead to ill-behaved simulations due to runaway population activity [61, 72, 130].

2.4 Conclusion

In this chapter, we laid the bases that will be used throughout the thesis to investigate the structure of correlated activity in the retina, as well as its impact on stimulus information encoding. We started by introducing precise definitions for pairwise signal and noise correlations that will serve in the rest of this thesis to characterize the collective behavior of retinal neurons. An overview of the literature showed that the effect of noise correlation on stimulus encoding could be summarized through the "sign-rule". A local version of the sign-rule, related to the Fisher information, suggested that noise correlations are detrimental when they align noise to the signal. Considerations based on the Fisher information further resulted in the conclusion that positive short range noise correlations are overall detrimental. Similarly, we saw that theoretical work on the mutual information suggested that noise correlations are detrimental when of the same sign than stimulus correlations, although fluctuations in their value could encode information about the stimulus.

These results are at odds with observations made in many sensory systems such as the retina or V1, where noise correlations tend to be strong and positive for cells that also have similar stimulus tuning. To pin down and understand this contradiction, one would need to relate directly the structure of signal and noise as measured by the quantities introduced in the first part of this chapter to a global measure of the stimulus information carried by neural responses. Doing so in the case of the mutual information is rendered difficult by

the fact that this quantity relies on estimations of probability distributions, which hinders greatly interpretability and straightforward application to data.

One of the main results of this thesis is the derivation of a mutual information approximation that is not only straightforward to apply to data, but most importantly provides an interpretable picture of the effects of noise and stimulus correlations on sensory coding. To develop, test and apply this approach, we relied and built on computational tools that we described in the last section of this chapter. Generalized Linear Models were used to model the population activity of retinal ganglion cells, and to perform synthetic experiments by simulating their responses. Maximum entropy models were used to describe (marginal and stimulus conditional) response distributions from experimental recordings of ganglion cells populations and compute efficiently from data information theoretic quantities such as the mutual information. In particular, the small-correlation approximate inference approach for maximum entropy models described previously served as the basis for the mutual information approximation we developed.

The next three chapters of this manuscript consist of articles that were written in the context of this thesis project. In the first article, we develop a new inference approach for Generalized Linear Models that palliates to some limitations of their classical inference approach and allows us to describe accurately the collective behavior of ganglion cells populations. In the second article, we detail the derivation of a small correlation approximation for the mutual information that yields both accurate estimates of information and an interpretable picture of the impact of noise correlations on stimulus encoding by sensory neurons. In the last paper (in preparation for submission), we delve on the implications of the picture derived from the mutual information approximation and showcase the different regimes in which noise correlations benefit stimulus encoding both in data and synthetic experiments.

Chapter 3

A new inference approach for training shallow and deep generalized linear models of noisy interacting neurons

A preliminary step to study how neural populations encode stimulus information in their collective activity often consists in building models of the population response to the encoded stimulus. Ideally, such models should allow us to understand how collective neural activity emerges from the different sources that shape its structure. As we have seen in two previous chapters, there are two main sources to the collective behavior of sensory neurons: on the one hand, incoming stimuli and sensory computations correlate neural responses across the population but also across time, inducing what are often called stimulus or signal correlations. On the other hand, the neural noise that arise from unreliable biological processes in the system such as noisy synaptic release is also correlated across network and time due to direct neural interactions between cells and shared inputs from upstream neurons.

A successful approach to model the collective behavior of sensory neurons, and in particular of retinal ganglion cells has been to use Generalized Linear Models (GLM) [135, 149]. However, the GLM seem to suffer from several limitations: a first well known problem of the GLM is that it is subject to very unnatural runaway activity transients [130, 72, 61] that limit its applicability. A second limitation of the GLM, is the fact that its inference via likelihood maximization does not disentangle stimulus and noise correlations. This can result in models that accurately predict total correlations, but not separately stimulus and noise correlations, as well as to models that do not generalize well to new stimuli not used in the training set.

In this chapter, we introduce a new inference approach for the GLM that yields models that capture separately accurately stimulus and noise correlations, and that do not exhibit runaway activity transients. The inference approach is compared to the results given by the classical log-likelihood maximization approach and applied to electrophysiological data from the rat retina. The integrality of the chapter that is reproduced below is a strict version of record of an article published previously as:

Gabriel Mahuas, Giulio Isacchini, Olivier Marre, Ulisse Ferrari, and Thierry Mora. “A new inference approach for training shallow and deep generalized linear models of noisy interacting neurons”. In: *Advances in neural information processing systems* 33 (2020), pp. 5070–5080

Abstract. Generalized linear models are one of the most efficient paradigms for predicting the correlated stochastic activity of neuronal networks in response to external stimuli, with applications in many brain areas. However, when dealing with complex stimuli, the inferred coupling parameters often do not generalize across different stimulus statistics, leading to degraded performance and blowup instabilities. Here, we develop a two-step inference strategy that allows us to train robust generalized linear models of interacting neurons, by explicitly separating the effects of correlations in the stimulus from network interactions in each training step. Applying this approach to the responses of retinal ganglion cells to complex visual stimuli, we show that, compared to classical methods, the models trained in this way exhibit improved performance, are more stable, yield robust interaction networks, and generalize well across complex visual statistics. The method can be extended to deep convolutional neural networks, leading to models with high predictive accuracy for both the neuron firing rates and their correlations.

3.1 Introduction

The pioneering work of J.W. Pillow and colleagues [135] showed how the Generalized Linear Model (GLM) can be used for predicting the stochastic response of neurons to external stimuli. Thanks to its versatility [191], high performance, and easy inference, the GLM has become one of the reference models in computational neuroscience. Nowadays, its applications range from retinal ganglion cells [135], to neurons in the LGN [9], visual [88], motor [182], parietal [131] cortices, as well as other brain regions [151, 144]. However, the GLM has also shown some significant limitations that has prevented its application to an even wider spectrum of contexts. In particular, the GLM shows unsatisfying performance when applied to the response to complex stimuli with spatio-temporal correlations much stronger than white noise, as for example naturalistic images [110] or videos [71].

A first limitation is that the inferred parameters depend on the stimulus used for training. This happens not only for the part of the model that deals with the external stimulus, which typically suffers a change in the stimulus statistics, but also for the couplings parameters quantifying interactions between the neurons of the network. However, if these couplings are to reflect an underlying network of biological interactions, they should be stimulus independent. In addition, and as we show in this paper, this lack of generalizability comes with errors in the prediction of correlated noise between neurons. This issue can strongly limit the application of GLM for unveiling direct synaptic connections between the recorded neurons [85] and for estimating the impact of noise correlations in information transmission [135].

A second issue is that the GLM can be subject to uncontrollable and unnatural self-excitation transients [130, 72, 61]. During these strong and positive feedback loops, the network's past activity may drive its current state to excitations above naturalistic levels, in turn activating neurons in subsequent time steps and resulting in a transient of very high, unrealistic activity. This problem limits the use of the GLM as a generative model—it is often necessary to remove those self-excitation runs by hand. Ref. [130] proposed an extension of the GLM that also includes quadratic terms limiting self-excitations of the network, but this comes at the price of more fitting parameters and harder inference. Ref. [72] showed that a GLM that predicts the responses several time-steps ahead in time [140] limits self-excitation, but this implies higher computational complexity and the risk of missing fine temporal structures. Alternatively, Ref. [61] proposed an approximation to estimate the stability of the inferred GLM model, and then used a stability criterion to constrain the parameter space over stable models. However the resulting models are sub-optimal, with degraded performance.

Thirdly, because neuronal responses are highly non-linear and hard to model for complex stimuli, the GLM fails to predict those responses correctly, even for early visual areas such as the retina [71]. Recently deep convolutional neural networks (CNNs) have been shown to outperform the GLM at predicting individual neuron mean responses [110, 29, 171, 30].

Compared to the GLM, these deep CNNs benefit from a more flexible and richer network architecture allowing for strong performance improvements [110]. However, the GLM retains an advantage over CNNs: thanks to the couplings between neurons in the same layer, it can account for both shared noise across the population and self-inhibition due to refractoriness. This feature, which is missing from deep CNNs [110], can be used to study how noise correlated in space and time impacts the population response [135]. A joint model combining the benefits of the deep architecture of CNNs and the neuronal couplings of the GLM is still lacking. It would allow for a more detailed description of the neuronal response to stimulus.

In this paper we develop a two-step inference strategy for the GLM that solves these three issues. We apply it to recordings in the rat retina subject to different visual stimulations. The main idea is to use the responses to a repeated stimulus to infer the GLM couplings without including the stimulus processing component. Then, in a second, independent step, we infer the parameters of the model pertaining to stimulus processing. Our approach allows for a wide variety of architectures, including deep CNNs. Finally, we introduce an approximation scheme to put together the two inference results into a single model that can predict the joint network response from the stimulus. All codes and data for the algorithms presented in this paper are available at `https://github.com/gmahuas/2stepGLM`.

3.2 Recordings

Retinal ganglion cells of a long-evans rat were recorded through a multi-electrode array experiment [103, 40] and spike-sorted with *SpyKING CIRCUS* [196]. Cell activity was stimulated with one unrepeated and two repeated videos of checkerboard (white-noise) and moving bars. For the checkerboard, we used the unrepeated ($1350s$) and one of the two repeated videos ($996s$ in total for 120 repetitions) for training, and the second repeated video for testing ($756s$ in total for 120 repetitions). Similarly, for the moving bars video we used the unrepeated ($1750s$) and one of the two repeated videos ($165s$ in total for 50 repetitions) for training, and the second repeated video for testing ($330s$ in total for 50 repetitions). In addition, we also recorded responses from a full-field movie with naturalistic statistics [40].

After sorting, we applied a spike-triggered average analysis to locate the receptive fields of each cell. Then, we used the response to full-field stimulation to cluster cells into different cell-types. In this work we focus on a population of $N = 25$ OFF Alpha retinal ganglion cells, which tile the visual field through a regular mosaic. The responses to both checkerboard and moving bars stimulations showed strong correlations, which we decompose into the sum of stimulus and noise correlations. Stimulus correlations are correlations between the cell mean responses (Peristimulus time histogram or PSTH). They are large only for the bars video, mostly because the video itself has strong and long-ranged correlations. Noise correlations, on the other hand, are due to shared noise from upstream neurons and gap junctions between cells in the same layer [25], and mostly reflect the architecture of the underlying biological network. Consistently, noise correlations were similar in the response of the two stimulations. In Suppl. sect. S1 we present additional statistics of the data.

3.3 Generalized linear model

In our Poisson GLM framework, $n^i(t)$, the number of spikes emitted by cell i in the time-bin t of duration $dt = 1.67ms$, follows a Poisson distribution with mean $\lambda^i(t)$: $n^i(t) \sim \text{Pois}(\lambda^i(t))$. The vector of the cells' firing rate $\{\lambda^i(t)\}_{i=1}^N$, with $N = 25$ is then estimated as

$$\lambda^i(t) = \exp\left\{h^i_{\text{offset}} + h^i_{\text{stim}}(t) + h^i_{\text{int}}(t)\right\} \ , \tag{3.1}$$

where h^i_{offset} accounts for the cell's baseline firing rate and where

$$h^i_{\text{int}}(t) = \sum_j \sum_{\tau>0} J_{ij}(\tau) n^j(t-\tau) \tag{3.2}$$

accounts for both past firing history of cell i itself and the contribution coming from all other cells in the network: J_{ii} are the spike-history filters, whereas $J_{i\neq j}$ are coupling filters. Both integrate the past up to 40ms. $h^i_{\text{stim}}(t)$ is a contribution accounting for stimulus drives, which takes the form of a linear spatio-temporal convolution in the classical GLM:

$$h^i_{\text{stim}}(t) = \sum_{\tau>0} \sum_{xy} K_{x,y}(\tau) S_{x,y}(t-\tau) \ , \tag{3.3}$$

where $S_{x,y}(t)$ is the stimulus movie at time t, $\{x, y\}$ being the pixel coordinates and $K_{x,y}(\tau)$ is a linear filter that integrates the past up to 500 ms. Later in the paper, we will go beyond this classical architecture and will allow for deep, non-linear architectures.

In order to regularize couplings and spike-history filters during the inferences, we projected their temporal part over a raised cosine basis [135] of 4 and 7 elements respectively, and added an $L1$-regularization $= 0.1$, which we kept the same for all the inferences. In addition, we imposed an absolute refractory period of τ^i_{refr} time-bins (calculated from the training set) during simulations and consequently the $J_{ii}(\tau)$ were set to zero for $\tau \leq \tau^i_{\text{refr}}$. In order to lower its dimension, the temporal behavior of stimulus filter $K_{x,y}(\tau)$ was projected on a raised cosine basis with 10 element. In addition we included an L1 regularization over the basis weights and a L2 regularization over the spatial Laplacian to induce smoothness.

All the inferences were done by log-likelihood (log-ℓ) maximization with Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, using the empirical past spike activity during training [135]. For easy comparison, all the performances discussed below are summarized in Table 3.1.

3.4 Failure of GLM for complex stimuli

We inferred the GLM by whole log-ℓ maximization from both the response to the checkerboard and moving bars non-repeated stimulations, and then simulated its response to the repeated videos (Fig. 3.1). Consistent with [135], in the case of the checkerboard stimulus, the model can predict with high accuracy the PSTH of all cells (Fig. 3.1A, mean Pearson's $\rho = 0.82 \pm 0.05$ std). It also reproduces the values of the zero-lag (17 ms window) noise correlations for all cell pairs (Fig. 3.1B, coefficient of determination CoD$= 0.94$, computed as

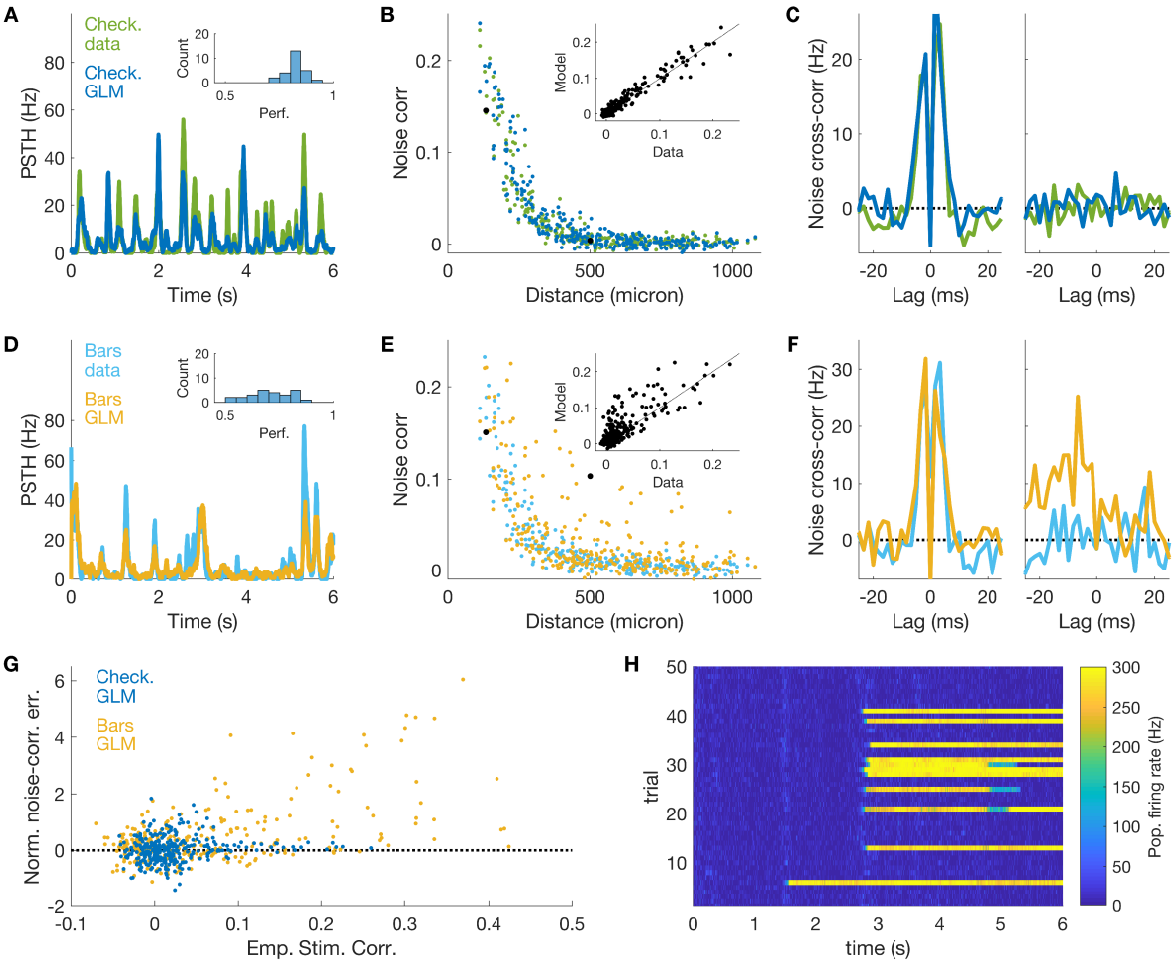

Figure 3.1: **GLM fails to predict noise correlations in the presence of strong stimulus correlations.** A) PSTH prediction for the response of an example cell to checkerboard stimulation. Inset: histogram of the model performance (Pearson correlation between empirical and model PSTH) for all cells in the population. B) Empirical and model predicted noise correlations versus distance between the cells. Inset: scatterplot. C) Empirical and model predicted noise cross-correlation between a nearby and a distant example cells. D,E,F) same as A,B,C, but for the response to moving bars stimulation. Note that the model overestimates noise correlations between certain pairs of distant cells. G) Error in the prediction of noise correlations normalized over their empirical value versus the empirical value of stimulus correlations. H) Population firing rate in time during model simulations of the responses to the moving bars stimulus. Note the transient of unnatural high activity due to self-excitation within the model.

$1 -$ var(error)/var(data)), and the temporal structure of noise cross-correlations (Fig. 3.1C).

The model performance is very degraded for the moving bars video—a stimulus characterised by long-range correlations. The model reproduces the empirical PSTH with rather good accuracy (Fig. 3.1D, $\rho = 0.71 \pm 0.10$ std) and shows fair overall accuracy on the noise correlations (Fig. 3.1E, CoD$= 0.55$). However it overestimates the value of the noise correlation for certain distant cell pairs (Fig. 3.1E&F). A closer look reveals that the model overestimates noise correlations for pairs of cells that are strongly stimulus-correlated (Fig. 3.1G). Here the error in the estimates is normalized over the empirical value of the noise correlations with a cut-off at three standard deviations. Interestingly, the effect is strong only for the moving bars video, as stimulus correlations are small for checkerboard stimulation. These results show that the inferred couplings of the GLM do not depend only on the correlated noise among the neurons, but can also be influenced by stimulus correlations. This prevents the inferred couplings from generalizing across stimuli. In addition, we observed several self-excitation transients when simulating the GLM inferred from the moving-bars stimulus (10% of the time, in 36% of the repetitions, Fig.3.1H, versus 0% for the model inferred from the checkerboard stimulus). This effect is probably the consequence of the over-estimation of those cell-to-cell couplings in the moving-bars stimulus, which drive the over-excitation of the network.

All these issues can be ascribed to the fact that by maximising the whole log-ℓ over all the parameters simultaneously, the GLM mixes the impact of stimulus correlations with neuronal past activity. In the next section we develop an inference strategy that disentangles stimulus from noise correlations and infer their parameters independently.

3.5 A two-step inference approach

In order to disentangle the inference of the couplings between neurons from that of the stimulus filters, we split the model training into two independent steps. We name this approach "two-step" inference (Fig. 3.2).

Coupling inference. We run a log-ℓ maximization inference over the response to a repeated video stimulation. Instead of inferring the parameters of a stimulus filter ($K_{x,y}(\tau)$ in Eq. 3.3), we treat the terms $h^i_{\text{stim}}(t)$ of Eq. 3.1 as auxiliary parameters that we infer directly from data (Fig. 3.2B). The log-ℓ derivative over these parameters is proportional to the difference between empirical and model-predicted PSTH. As a consequence, thanks to the repeated data, the addition of these parameters allows for enforcing the value of the PSTH exactly when the corresponding log-ℓ gradient vanishes. In this way, stimulus correlations

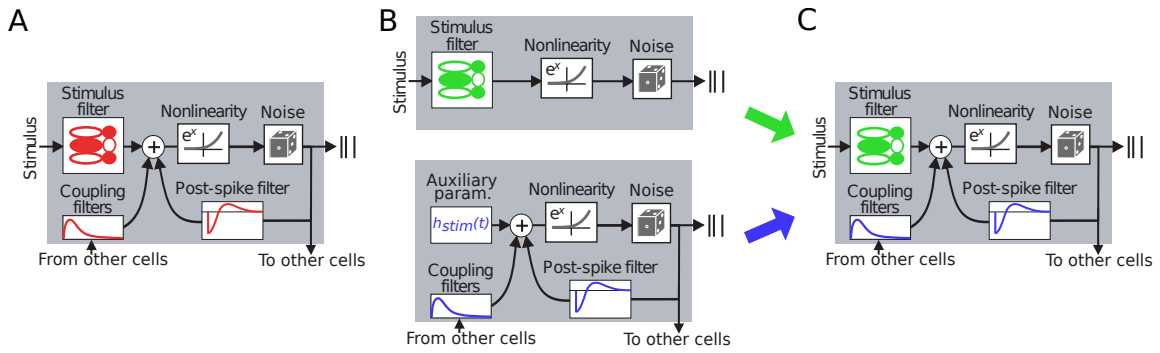

Figure 3.2: **Two-step inference of couplings and spike-history filters.** A) Whole log-ℓ maximization [135] trains couplings and spike-history filters together with the stimulus filter. B) Two-step inference trains couplings filters and stimulus filters by running two independent log-ℓ maximizations. Top: we remove coupling filters and infer the equivalent of an LNP model for each cell. Bottom: we run an inference over repeated data where we add auxiliary variables (instead of the stimulus filter) to exactly enforce the PSTH prediction. C) We build together the model by using the previously inferred parameters. A correction needs to be added (not shown, see text).

are perfectly accounted for, and the couplings only reflect correlated noise between neurons. As for the GLM inferred with whole log-ℓ maximization, we imposed an absolute refractory period of τ^i_{refr} time-bins and thus set $J_{ii}(\tau)$ to zero for $\tau \leq \tau^i_{\text{refr}}$.

Filter inference. We run the inference of a GLM model without the couplings between neurons or with themselves (spike-history filter) using the responses to the unrepeated stimulus (single-neuron linear-nonlinear Poisson (LNP) models [31], Fig. 3.2B). This inference allows us to predict the mean firing rate λ^i of each neuron.

Full model. Once couplings and stimulus filters are inferred, we can combine them to build up the full model (Fig. 3.2C). This cannot be done straightforwardly because the addition of the couplings will change the firing rate prediction of LNP model. As the average contribution of interactions on the activity of the cells were not taken into account during the inference of the stimulus model, we need to correct for this effect.

To do so, we subtract the mean contribution of the coupling term: $h^i_{\text{int}}(t) \to h^i_{\text{int}}(t) - \langle h^i_{\text{int}}(t) \rangle_{\text{noise}\sim\text{Pois}}$. This correction is equivalent to modify Eq. 3.2 into

$$\sum_j \sum_{\tau>0} J_{ij}(\tau) n^j(t-\tau) \to \sum_j \sum_{\tau>0} J_{ij}(\tau) \Big(n^j(t-\tau) - \lambda^i(t-\tau) \Big) \,. \tag{3.4}$$

Lastly, in order to account for the addition of absolute refractory periods, we added a term $\sum_{\tau=1}^{\tau^i_{\text{refr}}} \lambda^i(t-\tau)$ for each neurons (Suppl. Sect. S2). To compute all the corrections, we therefore only need the past firing rates $\lambda^i(t)$ of all neurons in the absence of the couplings,

which are given by the LNP model predictions. This allows the full model to predict the neuronal response to unseen (testing) data.

Note that this last correction is only an approximation. An exact alternative would be to perform the inference of the GLM stimulus filters as before, but in the presence of coupling filters fixed to the values inferred in the first step. Applying this approach to our data brought no improvement in terms of model accuracy, at the cost of more complex and time-consuming inferences.

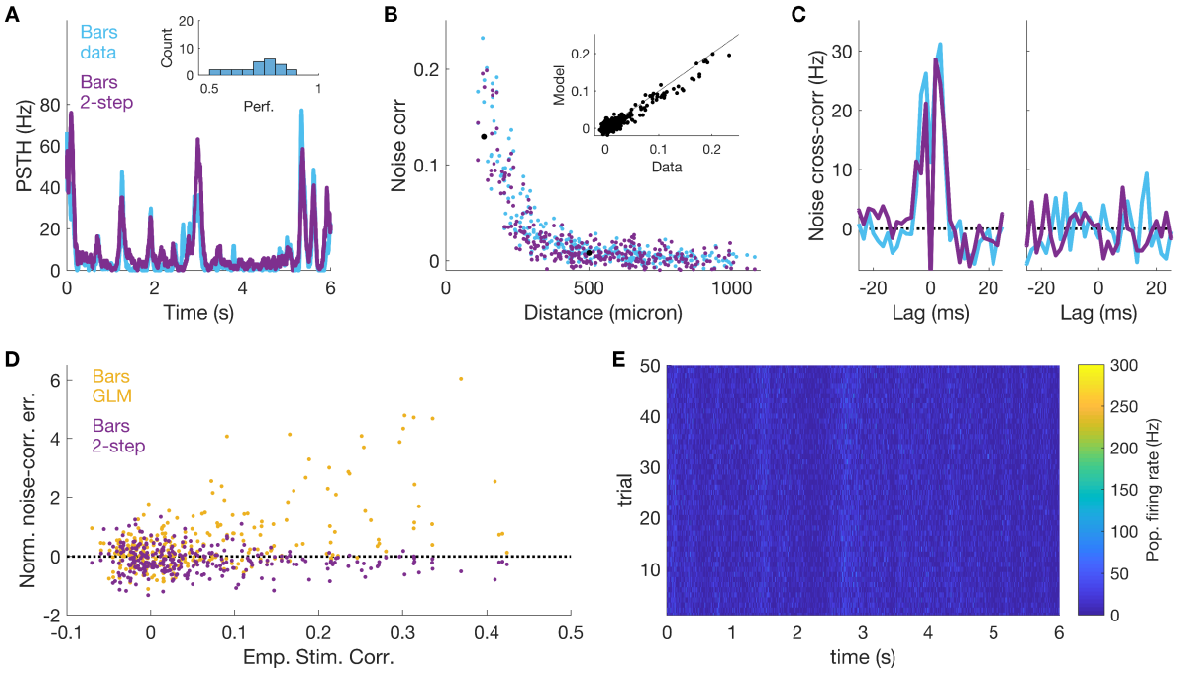

Figure 3.3: **Two-step inference retrieves noise correlations independently of the strong stimulus correlations in the moving bars video.** A) PSTH prediction for an example cell. Inset: histogram of the model performance for all cells. B) Empirical and model-predicted noise correlations versus distance between the cells. Inset: scatterplot. C) Empirical and model predicted noise cross-correlation between example pairs of nearby and distant cells. D) Normalized error in the prediction of noise correlations plotted versus the empirical value of the stimulus correlations. E) Population activity during model simulation shows no self-excitation transients.

We first applied our two-step inference to the response to checkerboard stimulation and obtained very similar results to whole log-ℓ maximization (Table 3.1). By constrast, performance was improved in the case of the moving bars stimulus (Fig. 3.3). The two inference approaches yielded similar performances for the PSTH (Fig. 3.3A, $\rho = 0.72 \pm 0.10$ std, versus $\rho = 0.71 \pm 0.10$ std), but for noise correlations we obtained much better results (Fig. 3.3B, CoD$= 0.91$, versus CoD$= 0.55$). In particular, the model avoids the overestimation the noise correlations for distant pairs (Fig. 3.3B&C) that we obtained with whole log-ℓ maximization (Fig. 3.1E&F). With the two-step inference, the strong stimulus

correlations of the moving bars video do not affect the model inference as was the case for whole log-ℓ maximization (Fig. 3.3D). In addition the model is much more stable, and we never observed self-excitation for either stimulus when simulating the model (Fig.3.3E, versus 10% of the time, Fig. 3.1H). In Table 3.1 we report all the performance for the different cases.

3.6 Two-step inference allows for generalizing across stimuli

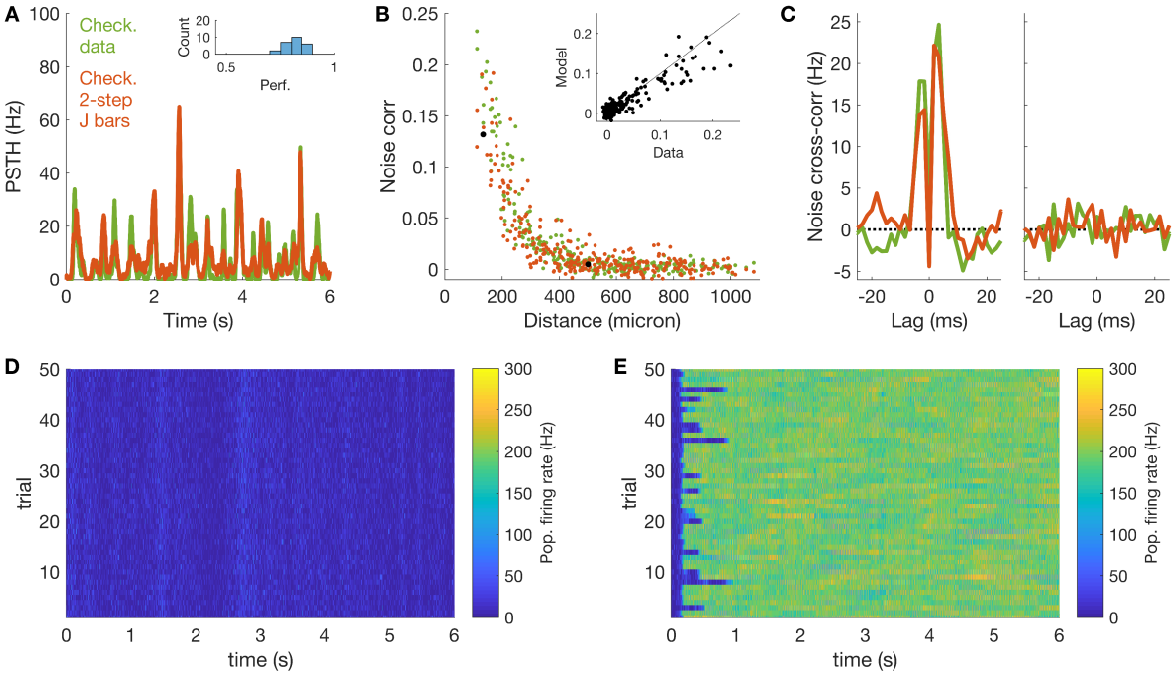

Figure 3.4: **Two-step inference allows for generalizing across stimulus ensembles** A,B,C,D) Simulation of the checkerboard responses for a model where stimulus filters were inferred from the response to checkerboard, and couplings filter were inferred from the moving bars data with our two-step inference. A) PSTH predictions. B) Noise correlations. C) Noise cross-correlation. D) Population activity showed no self-excitation transients E) Simulation of checkerboard responses when couplings filters are those inferred from moving bars data with whole log-ℓ maximization. The model shows self-excitation during all runs.

So far we have shown how our two-step approach can disentangle the inference of neuronal couplings from stimulus correlations. If these couplings are only due to network effects, one should expect them to generalize across stimulus conditions. To test for this, we run model simulations of one stimulus using its stimulus filter and the coupling filters inferred from the other. For the checkerboard movie (Fig. 3.4), and compared to the case where couplings are inferred on the same stimulus, with our two-step inference we obtained

performances that are almost equal for the PSTH ($\rho = 0.81 \pm 0.05$ std, versus $\rho = 0.81 \pm 0.05$ std) and rather good for noise correlations (CoD$= 0.84$, versus CoD$= 0.95$). In addition, we never observed self-excitation (Fig. 3.4D). By contrast, when we used the couplings inferred by whole log-ℓ maximization, self-excitation happens so often (93% of the time in 100% of the repetitions) that we were not able to estimate the model performance (Fig. 3.4E).

For the moving bars video (Fig. S2), our two-step inference yielded performances similar to the case where couplings were inferred on the same stimulus (Table 3.1). Using the couplings inferred by whole log-ℓ maximization instead, the model performance decreased for the PSTH ($\rho = 0.65 \pm 0.12$ std, versus $\rho = 0.71 \pm 0.10$ std), and improved for noise correlations (CoD$= 0.80$, versus CoD$= 0.55$). In conclusion, two-step outperforms whole log-ℓ maximization on both stimuli (Table 3.1).

3.7 Deep GLM outperforms previous approaches

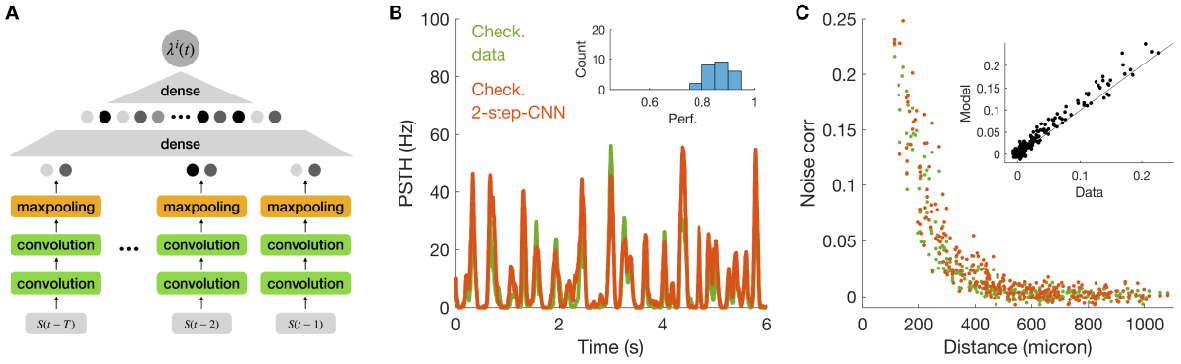

Figure 3.5: **Deep CNN can be included in our two-step approach to improve model performance** A) Architecture of our deep, time-distributed CNN. B) PSTH prediction for the response of an example cell to checkerboard stimulation. Inset: histogram of model performance for all cells. C) Empirical and model predicted noise correlations versus distance between cells. Inset: scatterplot.

Our two-step inference decomposes the model training into two independent components, one for the stimulus processing and one for network effects. In the previous experiments we still used a linear convolution to process the stimulus, but thanks to this decomposition, we can also consider any machine capable of predicting the neurons firing rates $\{\lambda^i(t)\}_{i=1}^N$. In order to predict the response to checkerboard stimulation with higher accuracy, we inferred a deep, time-distributed CNN, a special case of CNNs [110] with the additional constraint that the weights of the convolutional layers are shared in time [32]. In our architecture, two time-distributed convolutional layers are followed by a max-pooling and eventually by two dense layers that output the firing rate $\lambda^i(t)$ (Fig. 3.5A, see supple-

	Checkerboard stimulus			Moving bars stimulus		
	PSTH	noise-corr.	self-exc.	PSTH	noise-corr.	self-exc.
whole log-ℓ maximization	0.82 ± 0.05	0.94	0%	0.71 ± 0.10	0.55	10%
two-step approach	0.81 ± 0.05	0.95	0%	0.72 ± 0.10	0.91	0%
coupl. exchange max logℓ	unstable	unstable	93%	0.65 ± 0.12	0.80	0%
coupl. exchange two-step	0.81 ± 0.05	0.84	0%	0.73 ± 0.09	0.91	0%
CNN	0.87 ± 0.04	0.93	0%	—	—	—

Table 3.1: **Model performance for different inference approaches.** We computed Pearson's correlation coefficients between empirical and model predicted firing rate (PSTH). For noise correlations, we estimated the CoD between model predictions and data. The third and forth rows refer to simulations that use the coupling filters inferred from the other stimulus.

mentary section 4 for more details). After training, we included the model in our two-step inference to build a model with both a deep architecture for the stimulus component, and a network of coupling filters.

The model shows higher performance in predicting the PSTH: $\rho = 0.87 \pm 0.04$ std, versus $\rho = 0.82 \pm 0.05$ std and $\rho = 0.81 \pm 0.05$ std, when compared to our previous models (Fig. .3.5B). In addition, the model was capable of predicting noise correlations with high accuracy (Fig. .3.5C, CoD$= 0.93$, versus CoD$= 0.94$ and CoD $= 0.95$). We also did not observe any self-excitation transient. In summary, the model combines the benefits of deep networks with those of the GLM with its neuronal couplings.

We summarise all the different model performances in Table 3.1.

3.8 Discussion

In this work we have studied the application of the GLM to the case of retinal ganglion cells subject to complex visual stimulation with strong correlations. We have shown how whole log-ℓ maximization over all model parameters leads to inferring erroneous coupling

filters that reflect stimulus correlations (Fig. 3.1G). This effect introduces spurious noise correlations when the model is simulated (Fig. 3.1E&F), prevents its generalization from one stimulus ensemble to another (Fig. 3.4E), and increases the chance of having self-excitation in the network dynamics (Fig. 3.1G). This last issue poses a major problem when the GLM is used as a generative model for simulating spiking activity.

To solve these issues we have proposed a two-step algorithm for inferring the GLM that takes advantage of repeated data to disentangle the stimulus processing component from the coupling network. A similar approach has been proposed in the context of maximum entropy models [64, 51], and here we have fully developed it for the GLM. Our method prevents the rise of large couplings reflecting strong stimulus correlations (Fig. 3.3D). The absence of these couplings lowers the probability of observing self-excitation (Fig. 3.3E) and the inferred GLM does not predict spurious noise correlations (Fig. 3.3B&C). In addition, with our two-step inference the couplings are robust to a change of stimulus, and allows for generalizations (Fig. 3.4). In particular we showed that a model with the stimulus filter inferred from checkerboard data but couplings inferred from moving bars stimulation predicts with high accuracy the response to checkerboard.

The strongest drawback of using our method is the requirement of repeated data, which are not necessary for whole log-ℓ maximization of GLM. This may limit the application of our inference approach. However we emphasize that only $165s$ of repeated data were needed for inferring the couplings. In addition, another possibility that deserves to be tested is the use of spontaneous activity instead of repeated stimuli. For the retina, this activity can be recorded while the tissue is exposed to a static full-field image (blank stimulus). However, as spontaneous activity is usually very low, these recordings need to be long enough to measure correlations with high precision.

Another important contribution of our work is the possibility to easily include deep CNNs into the GLM to increase its predicting power. Deep CNNs represent today one of the best options for modelling and predicting the mean response of sensory neurons to complex stimuli such as naturalistic ones [110, 29, 171, 30], and architectures based on deep CNNs expanded with recurrence are therefore of great interest for studying the neural dynamic of sensory systems [119]. However, building a deep network that take as input both stimulus and the past activity of the neural population can be very challenging, as it implies dealing with very heterogeneous inputs. Our framework solves this problem by separating the CNN inference from that of coupling and spike-history filters, and can thus be easily added on an already inferred CNN.

The GLM has been used to estimate the impact of correlated noise on information

transmission, but mostly for stimuli with low complexity [135, 57]. Future works can apply our method to model the responses to complex stimulations and study its impact on stimulus encoding.

Broader Impact

In this work we present a computational advance to improve the inference and performance of the GLM. As the GLM is one of the most used models in computational neuroscience, we believe that many researchers can benefit from this work to advance in their investigations. The fight against blindness, which affects about 45 millions people worldwide, is one of such possible applications. Retinal prostheses, where an array of stimulating electrodes is used to evoke activity in neurons, are a promising solution currently under clinical investigation. A central challenge for such implants is to improve the information that is sent to the brain. A central challenge for retinal implants is thus to mimic the computations carried out by a healthy retina to optimize information sent to the brain. Modeling retinal processing could thus help optimize vision restoration strategies in the long term [53].

We believe that no one will be put at disadvantage from this research, that there are no consequences of failure of the system. Biases in the data do not apply to the present context.

Acknowledgement We thank M. Chalk and G. Tkacik for useful discussion. This work was supported by ANR TRAJECTORY and DECORE, by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 785907 (Human Brain Project SGA2), by a grant from AVIESAN-UNADEV to OM, by the French State program Investissements d'Avenir managed by the Agence Nationale de la Recherche [LIFE-SENSES: ANR-10-LABX-65], by the Programme Investissements d'Avenir IHU FOReSIGHT (ANR-18-IAHU-01) and by Agence Nationale de la Recherche grant ANR-17-ERC2-0025-01 "IRREVERSIBLE".

The authors declare no competing interests for this paper.

S1 Empirical data and correlations

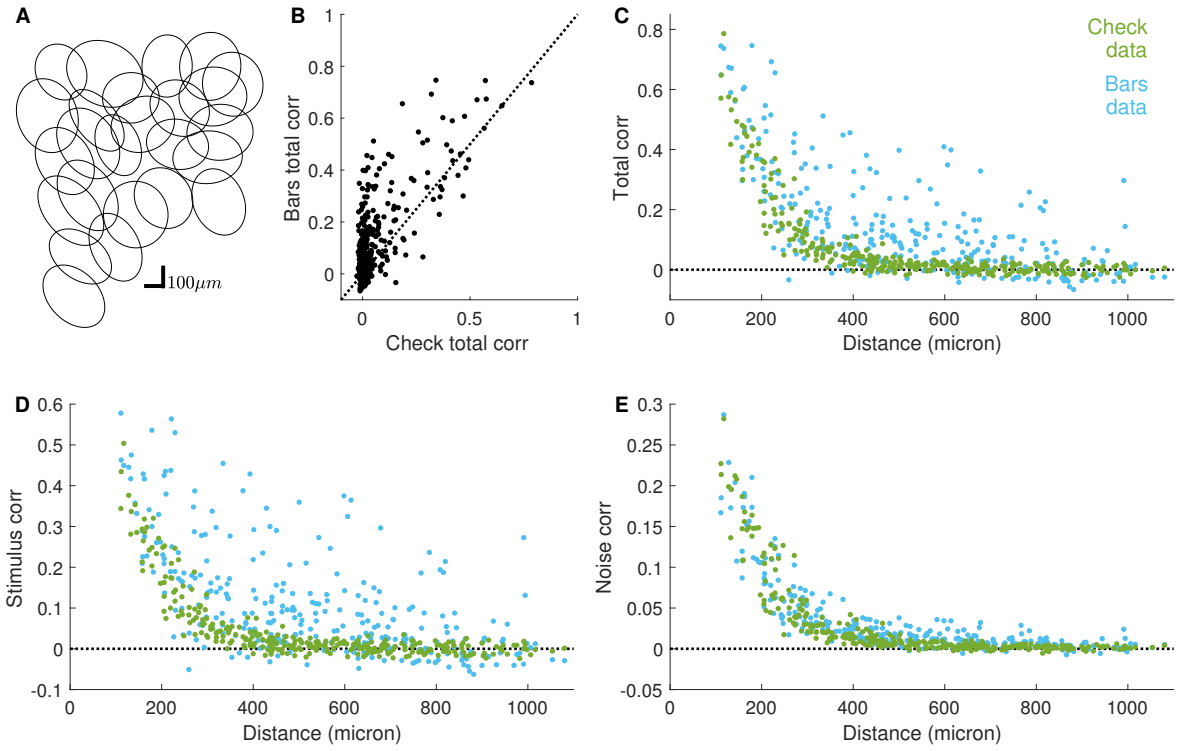

Figure S1: **Stimulus and noise correlation in the retinal response** A) Mosaic for $N = 25$ OFF alpha cells. B) Scatterplot of total pairwise correlation between the spiking activity in response to checkerboard and moving bars video. C) Total pairwise correlation versus cell distance D) Stimulus correlation versus cell distance E) Noise correlation versus cell distance

Responses to checkerboard and moving bars stimuli show different correlation patterns (Fig. S1). The moving bars video induces much stronger and long-ranged stimulus correlations, especially for certain pairs of distant cells. On the contrary, noise correlations decrease smoothly with distance and are of similar magnitude in the two datasets.

S2 Correction for the absolute refractory period

As explained in the main text, when we add the two-step coupling filters to the LNP model, we need to correct the h^i_{int} by its mean, Eq.3.4. However this correction does not take into account the addition of an absolute refractory period. In fact, if we start with an LNP model with rate $\lambda(t)$, and we prevent the cell to spike if it has spiked in the previous τ^i_{refr} time-bins during simulations, then the model rate will become a random variable itself with an average lower than $\lambda(t)$. In order to correct for this effect, we need first to quantify the

mean of $n(t)$, the spike-count at time t:

$$\begin{aligned}
\mathbb{E}\left(\, n(t) \,\right) &= \mathbb{E}\left(n(t) \sim Pois(\lambda(t)) \,\middle|\, \sum_{\tau} n(t-\tau) = 0 \right) \\
&= \mathbb{E}\Big(n(t) \sim Pois(\lambda(t)) \Big)\ \text{Prob}\left(\sum_{\tau} n(t-\tau) = 0 \right) \\
&\approx \mathbb{E}\Big(n(t) \sim Pois(\lambda(t)) \Big)\ \prod_{\tau} \text{Prob}\left(n(t-\tau) = 0 \right) \\
&= \lambda(t)\ \prod_{\tau} \exp\{-\lambda(t-\tau)\}
\end{aligned} \tag{3.5}$$

where the approximation is valid under the hypothesis of small λ. By taking the log of Eq. 3.5, we obtain the correction term $\sum_{\tau} \lambda(t-\tau)$ that needs to be added to $h_{int}(t)$ in order to correct for the addition of the absolute refractory period.

S3 Generalization results for moving bars stimulus

S4 Time Distributed Convolutional Neural Network

In section 3.7 we introduce the constrained architecture of Time Distributed Convolutional Neural Networks. In order to exploit the information in the 2D spatial structure of the data we use two convolutional layers, as it is successfully done in [110], with kernels of 8x8 and 5x5 size and two feature channels each. A MaxPooling layer of pool size 2x2 is then subsequently applied to complete the spatial computation of the network. We additionally impose a Time Distributed architecture [32], i.e. the independent application of the same spatial computation to each time slice of the input, as can be seen Fig. 3.5. Each temporal slice of the input is compressed though the convolutional part of the network to two real numbers. Subsequently the temporal information is combined through a dense layer of 100 units with softplus activation function. A Dropout layer is additionally implemented before the last layer in order to enforce regularisation.
This architecture reduces the number of parameters to ≈ 3000. Each model is trained for 30 epochs using the Adam optimiser on batches of 200 samples. A validation set was used to monitor the inference.

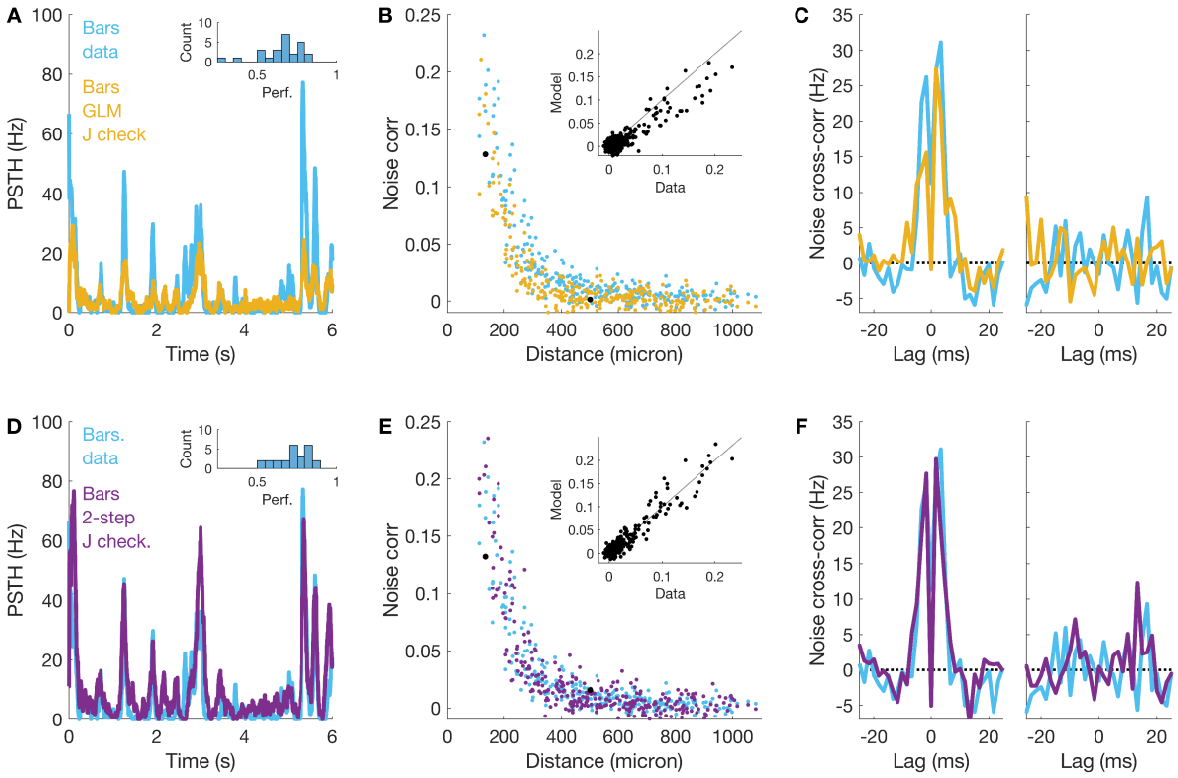

Figure S2: **Generalization results for moving bars stimulus** Simulation of the moving bars responses for a model where stimulus filters were inferred from the response to moving bars and couplings filter were inferred from the checkerboard data (opposite of Fig. 3.4) with whole log-ℓ maximization (A,B,C) and with our two-step inference (D,E,F). A,D) PSTH predictions. B,E) Noise correlations. C,F) Noise cross-correlation.

Chapter 4

A small-correlation expansion to quantify information in noisy sensory systems

As we have seen in the introductory chapters to this thesis, investigating how sensory neurons encode stimuli into their population response requires being able to relate the structure of collective neural activity to a measure of the information carried by neural responses about the visual stimulus. Further, we have discussed how the collective behavior of neural responses originates from both the structures of neural noise and stimulus induced activity (i.e. neural signal). Has we have further seen in chapter 1, the contributions of noise and signal to the collective activity of sensory neurons can be described to some extent by pairwise noise and signal correlations that can be estimated directly from data.

A theoretically grounded way to estimate the information neural responses carry about the stimulus consists in calculating the mutual information between stimulus and response. However, the mutual information is notoriously difficult to evaluate accurately from limited experimental data as it requires estimating the high-dimensional probability distributions that describe neural responses. A way to circumvent this issue consists in building models of response distributions, fit them to experimental data and use them to estimate information theoretic quantities such as the mutual information. A successful approach that has been applied to model the collective activity of many sensory systems is the maximum entropy approach. Pairwise maximum entropy models as introduced in chapter 2 are the least structured (and thus maximally agnostic) models that reproduce means and covariances from the data they are fit to.

In this chapter, we propose an analytical approach that bridges the gap between empirical measures of pairwise neural correlations and mutual information between stimulus

and response. The approach relies on the use of a small-correlation expansion [155] that was originally developed as an approximate inference approach for pairwise maximum entropy models. Here, we use it to express the entropies of neural response distributions in terms of their correlations, and find an approximation of the mutual information between stimulus and responses expressed in terms of these correlations. We validate the approach on synthetic data generated by a biologically inspired Generalized Linear Model of 12 RGC-like neurons and show that it provides accurate measures of mutual information even under significant stimulus and noise correlation strengths. We then apply the approach on retinal data to assess the effect on sensory coding of noise correlations between neurons as well as that of refractory effects (noise correlations across time) in single neurons. The integrality of the chapter below is a strict version of record of an article published previously as:

Gabriel Mahuas, Olivier Marre, Thierry Mora, and Ulisse Ferrari. “Small-correlation expansion to quantify information in noisy sensory systems”. In: *Physical Review E* 108.2 (2023), p. 024406

Abstract. Neural networks encode information through their collective spiking activity in response to external stimuli. This population response is noisy and strongly correlated, with complex interplay between correlations induced by the stimulus, and correlations caused by shared noise. Understanding how these correlations affect information transmission has so far been limited to pairs or small groups of neurons, because the curse of dimensionality impedes the evaluation of mutual information in larger populations. Here we develop a small-correlation expansion to compute the stimulus information carried by a large population of neurons, yielding interpretable analytical expressions in terms of the neurons' firing rates and pairwise correlations. We validate the approximation on synthetic data and demonstrate its applicability to electrophysiological recordings in the vertebrate retina, allowing us to quantify the effects of noise correlations between neurons and of memory in single neurons.

4.1 Introduction

Networks of neurons from sensory systems are characterized by strong correlations that shape their collective response to stimuli [25, 153, 97, 87, 138, 57]. These correlations have two sources [25]: *stimulus correlations*, which originate from shared or correlated stimuli that affect the mean activities of different neurons in a concerted way; and *noise correlations*, which stem from network interactions that couple noise across cells. These two sources of correlations impact how well the population encodes stimulus information, and detailed investigations have explored this effect both experimentally [138, 57, 135, 149, 70] and theoretically [199, 1, 129, 137, 166, 45, 75], showing a wide variety of scenarios in which noise correlations could either hurt or improve information transmission (see [8] for a recent review).

While geometric arguments about the structure of stimulus and noise correlations can help interpret and evaluate the impact of their interplay on information transmission for pairs or small groups of cells [75, 8], specific challenges arise when dealing with large populations of cells. A common way to quantify these effects is by computing the mutual information between the stimulus and the activity of the whole population. However, attempts at quantifying this information are inherently limited by the curse of dimensionality, whereby the size of the state space to be sampled grows exponentially with the system's size. Models based on the principle of maximum entropy have been proposed to build explicit probabilistic models of the collective activity of many neurons, based on mean spike rates and correlation functions [153, 173, 177, 158, 174, 54]. These distributions map onto known models of statistical mechanics, and can be used to evaluate entropies as well as

mutual informations.

In this paper, we leverage these techniques from statistical physics to compute the information of experimental spike trains using a small correlation expansion [155]. We show on synthetic data that this approach outperforms previous approximations of the mutual information and is computationally efficient. The resulting formulas are expressed as simple functions of the experimental observables, yielding an intuitive picture of how correlations affect information encoding in sensory systems beyond the previously discussed "sign rule" [75], which states that noise correlations are beneficial when of opposite sign to stimulus correlations. We apply our formulas to real electrophysiological recordings from the retina, to illustrate how it can be used to quantify the effect of noise correlations between neurons and across time.

4.2 Small correlation expansion of the mutual information

The collective response of a neural network of size N can be described by the neuronal activities $\mathbf{n} = (n_1, \ldots, n_N)$, taking value 0 or 1 depending on whether the neuron spikes or not within a short time window Δt (typically 10 – 20 ms). In general, because of processing delays and adaptation, the response is a stochastic function $P(\mathbf{n}|\mathbf{s})$ of the history of the stimulus $\mathbf{s}$ up to the response. The mutual information $I(\mathbf{n}, \mathbf{s})$ quantifies the amount of information conveyed by the neural response about the stimulus [156, 197]. Since it is expressed as a difference of entropies $I = H[\mathbf{n}] - \langle H[\mathbf{n}|\mathbf{s}]\rangle_{\mathbf{s}}$, where $H[x] = -\sum_x P(x) \ln P(x)$, its quantification requires good entropy estimators. Direct estimation methods from data exist, and can be applied for relatively small groups of neurons [170]. However, the estimation problem quickly becomes intractable as the number of neurons increases and the size of the response space grows exponentially. To deal with large networks, we thus developed a method based on a small correlation expansion of entropies [155], which allows us to express them as analytical functions of the empirical correlations.

We start by assuming that both $P(\mathbf{n})$ and $P(\mathbf{n}|\mathbf{s})$ follow the form of maximum entropy models consistent with empirical pairwise covariances and spike rates. Later we will discuss the limitations of this assumption. The total covariance between two cells i and j across stimuli, $C_{ij} \equiv \mathrm{Cov}(n_i, n_j)$, can be decomposed into two contributions corresponding to the effects of stimulus and noise: $C_{ij} = C^{\mathrm{s}}_{ij} + \langle C^{\mathrm{n}}_{ij}(\mathbf{s})\rangle_{\mathbf{s}}$, with $C^{\mathrm{s}}_{ij} \equiv \mathrm{Cov}_{\mathbf{s}}(\langle n_i\rangle_{n_i|\mathbf{s}}, \langle n_j\rangle_{n_j|\mathbf{s}})$, $C^{\mathrm{n}}_{ij}(\mathbf{s}) \equiv \mathrm{Cov}(n_i, n_j|\mathbf{s})$, which can be computed from the response to repeated presentations of the same stimulus. Likewise, the Pearson correlation coefficient $\rho^{\mathrm{tot}}_{ij} \equiv C_{ij}/\sqrt{C_{ii}C_{jj}}$

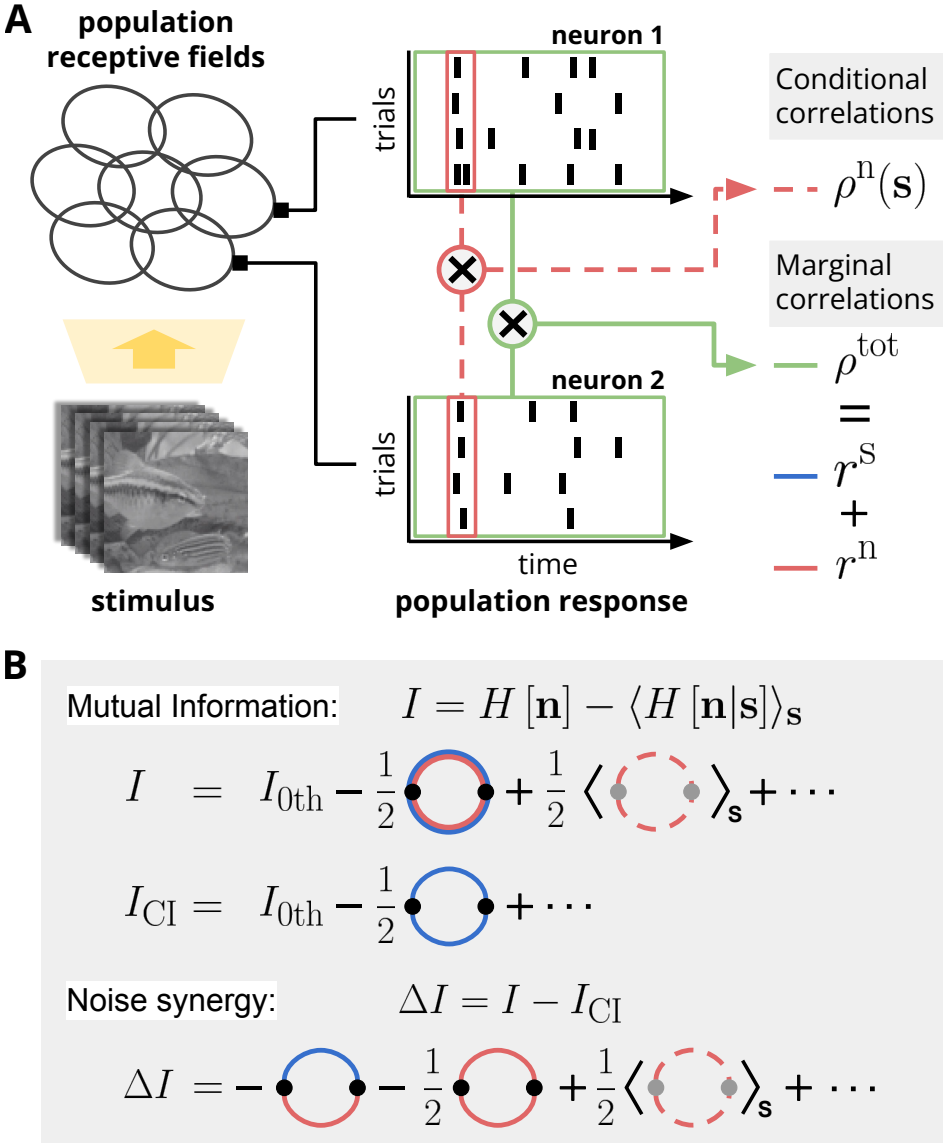

Figure 4.1: **How correlations affect information.** A) Visual stimuli drive the noisy response of sensory neurons (spikes, represented by vertical ticks). $\rho^{\mathrm{n}}_{ij}(\mathbf{s})$ is the pairwise Pearson correlation between the activities of cells i and j in a short window Δt, conditioned on past stimuli $\mathbf{s}$. The total Pearson correlation, ρ^{tot}, can be decomposed into stimulus and noise contributions, r^{s} and r^{n}. B) Small-correlation expansion of the mutual information I and synergy $\Delta I = I - I_{\mathrm{CI}}$ between stimulus $\mathbf{s}$ and response $\mathbf{n}$, based on [155]. I_{CI} is the information in absence of noise correlations. In the diagrams each line corresponds to a correlation term; double lines are sums of two correlations; multiple lines connecting the same two points are multiplied.

can also be decomposed into stimulus- and noise-induced contributions: $\rho_{ij}^{\text{tot}} = r_{ij}^{\text{s}} + r_{ij}^{\text{n}}$ (Fig. 4.1A), with $r_{ij}^{\text{s}} \equiv C_{ij}^{\text{s}}/\sqrt{C_{ii}C_{jj}}$ and $r_{ij}^{\text{n}} \equiv C_{ij}^{\text{n}}/\sqrt{C_{ii}C_{jj}}$. Note however that these two terms are not proper correlation coefficients because of the normalization. Stimulus correlations may instead be quantified by $\rho_{ij}^{\text{s}} \equiv C_{ij}^{\text{s}}/\sqrt{C_{ii}^{\text{s}}C_{jj}^{\text{s}}}$, and noise correlations in a stimulus-dependent manner through: $\rho_{ij}^{\text{n}}(\mathbf{s}) \equiv C_{ij}^{\text{n}}(\mathbf{s})/\sqrt{C_{ii}^{\text{n}}(\mathbf{s})C_{jj}^{\text{n}}(\mathbf{s})}$.

Following [155], we expand the entropy of the maximum entropy models—and thus the mutual information—at small values of the covariance parameter (C_{ij} or $C_{ij}^{n}(\mathbf{s})$), $I = I_{0\text{th}} + I_{1\text{st}} + I_{2\text{nd}} + \dots$ (App. S1). The leading order of this expansion is the sum of the information carried by each neuron: $I_{0\text{th}} = \sum_i [H[n_i] - \langle H[n_i|\mathbf{s}]\rangle_{\mathbf{s}}]$. The first order term vanishes, while the second one reads (Fig. 4.1B, App. S1):

$$I_{2\text{nd}} = -\frac{1}{2}\sum_{i<j}\left(\rho_{ij}^{\text{tot}\,2} - \langle\rho_{ij}^{\text{n}}(\mathbf{s})^2\rangle_{\mathbf{s}}\right). \tag{4.1}$$

We can compute higher-order terms using Feynman diagrams rules [92], but they quickly become unwieldy. However, some of these terms can be re-summed to yield a better approximation of the mutual information than (4.1) in terms of first and second order moments [155] (App. S2):

$$I \approx I_{0\text{th}} + I_{\text{pairs}} + I_{\mathcal{G}} - I_{\text{dbl}}. \tag{4.2}$$

I_{pairs} is the sum of the mutual information gains (with respect to single cells) of each pair (i,j) calculated one by one, ignoring the rest of the network. $I_{\mathcal{G}}$ is the mutual information gain computed through a mean-field (or *loop*) approximation [34, 155], which is equivalent to assuming that all fluctuations (stimulus and noise) are Gaussian:

$$I_{\mathcal{G}} = \frac{1}{2}\log\left(|\rho^{\text{tot}}|\right) - \frac{1}{2}\langle\ln\left(|\rho^{\text{n}}(\mathbf{s})|\right)\rangle_{\mathbf{s}}, \tag{4.3}$$

where $|\rho|$ denotes the determinant of the correlation matrix. Finally I_{dbl} corrects for terms that are double-counted in I_{pairs} and $I_{\mathcal{G}}$.

4.3 Noise synergy

These expansions can be used to investigate the impact of noise correlations on information transmission. We define the *noise synergy*, $\Delta I \equiv I - I_{\text{CI}}$, as the gain in information relative to the conditionnally independent case (Fig. 4.1B, bottom line). I_{CI} can be computed in practice by shuffling the response of individual neurons across repetitions of the same stimulus, which preserves stimulus correlations but destroys noise correlations. At second

order we obtain (App. S1):

$$\Delta I \approx \sum_{i<j} \left[-r^{\mathrm{n}}_{i,j} r^{\mathrm{s}}_{i,j} + \frac{1}{2} \left(\left\langle \rho^{\mathrm{n}}_{i,j}(\mathbf{s})^2 \right\rangle_{\mathbf{s}} - {r^{\mathrm{n}}_{i,j}}^2 \right) \right]. \tag{4.4}$$

This expression shows how noise synergy depends on noise correlations through r^{n} and ρ^{n}. The first term is positive when noise and stimulus correlations have opposite signs. This effect is known in the literature as the *sign rule* [75] and can be interpreted in terms of the whitening of the output power spectrum: it is beneficial for the network to "cancel out" input correlations by adding noise correlations of opposite sign, in order to approach a uniformly distributed output, thereby increasing output entropy and information. The second term of (4.4), which is of second order in the noise correlation parameter, can be either positive or negative in general. However, in the particular case of noise correlations independent of the stimulus, $\rho^{\mathrm{n}}(\mathbf{s}) = \rho^{\mathrm{n}}$, the Cauchy-Schwarz inequality guarantees its non-negativity (see App. S1 for a proof). This implies that noise correlations may be beneficial even when the sign rule is violated and the noise correlations are constant (see last section of App. S4). Noise synergy can also be computed using the re-summed entropies of (4.2). The formulas are slightly more involved and are reported in App. S2.

4.4 Numerical test on synthetic data

To test our approximations (4.1) and (4.2), we built a generalized linear model to mimic the response of a small population of 12 retinal neurons with nearest-neighbor interactions (Fig. 4.2A) for which mutual informations could be estimated exactly. The stimulus is modeled as a random Gaussian field sampled at 100Hz, with varying spatial correlations, allowing us to tune the strength of stimulus correlations (App. S3). The stimulus is convolved with a linear filter consisting of a difference-of-Gaussians receptive field with biphasic temporal kernel [101] (App. S3). The mean spike rate is controled by the result of this convolution, to which the effect of its own spiking history is added, through a non-linear function. In addition, the past activities of its neighbors control the stochastic part of firing, through coupling filters (the mean effect of which is subtracted from the average rate, see App. S3). This strategy allows us to tune noise correlations while keeping stimulus correlation constant. Importantly, this model is mathematically inconsistent with the maximum entropy assumption. It thus allows us to test for both the appropriateness of the maximum entropy approximation in the context of a realistic spiking model, and the accuracy of the small-correlation expansion.

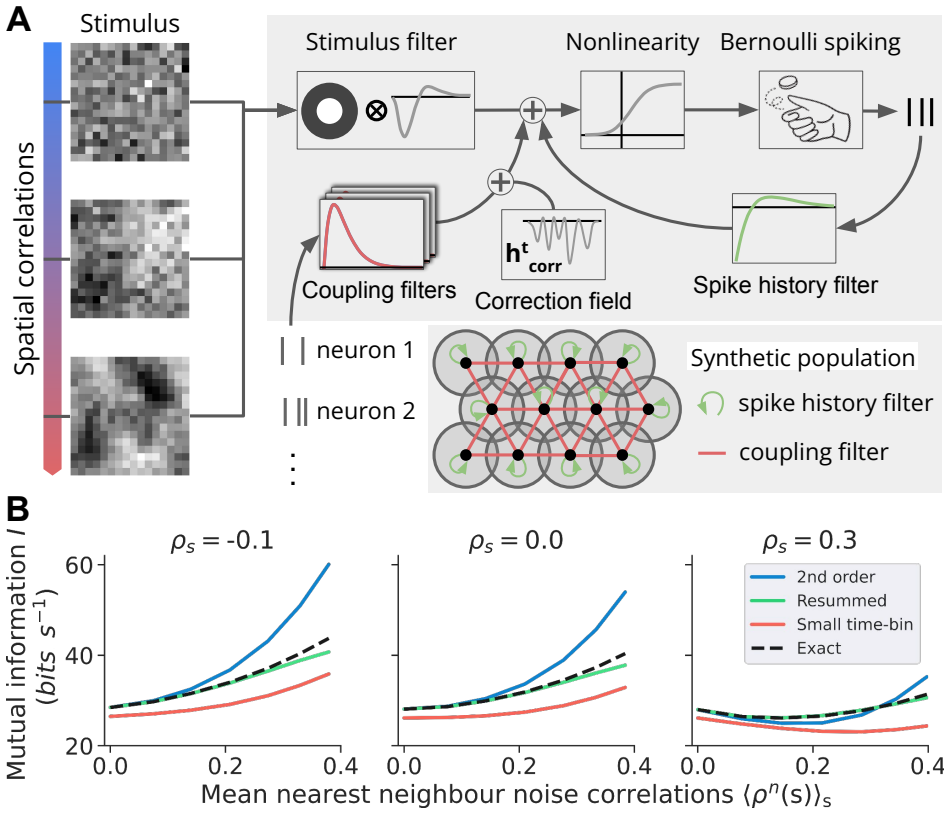

Figure 4.2: A) A spatially correlated random stimulus activates a network of 12 neurons according to a generalized linear model defined by stimulus, coupling, and spike-history filters. B) Exact, second order (4.1), re-summed (4.2), and small time bin expansion [129] of the mutual information for various strengths of the noise and stimulus correlations (averaged over all pairs of neighbors). Note that since mutual information is a difference of entropies, the error may be a non monotonic function of ρ^n and ρ^s.

After binning at $15ms$, we computed the exact mutual information between stimulus and response using exhaustive numerical simulations, and compared it with the predictions of our approximations, as well as the state-of-the-art small time bin expansion of Ref. [129] (Fig. 4.2B). We observed an excellent agreement between numerical calculations and analytical expressions, in particular for the re-summed mutual information (4.2), in contrast to the small time bin approximation, which yields inaccurrate results even in absence of noise correlations. Although less accurate, the second order approximation (4.1) still provided fair estimates for a wide range of correlation strengths. We further checked that the error did not blow up with the system's size, by analyzing networks subsampled from the full population with sizes 3 through 12 for various values of the stimulus and noise correlations (Fig. S1).

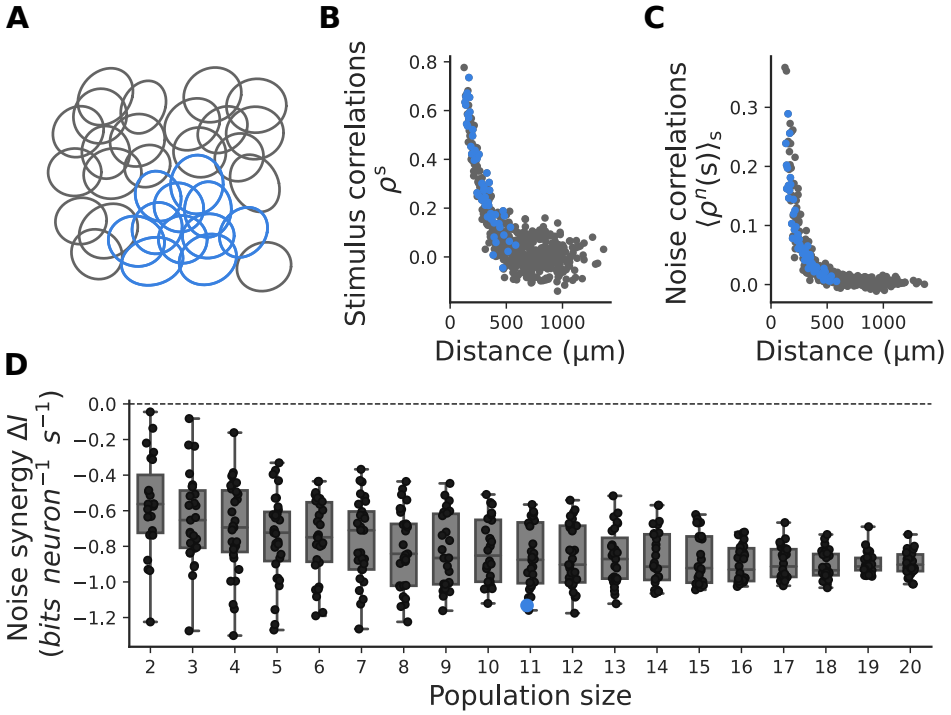

Figure 4.3: **Application on retinal population response to visual stimulation.** A) A mosaic of a population of off alpha cells in the rat retina. B) Stimulus correlation (ρ^{s}_{ij}) plotted against the distance between pair of cells stimulated with a white noise movie. C) Same as (B) but for noise correlations $\langle \rho^{\mathrm{n}}_{ij}(\mathrm{s}) \rangle_{\mathrm{s}}$. D) Noise synergy for subset populations of nearby cells. Each boxplot corresponds to the noise synergy of many subgroup of ganglion cells. Only nearby cells are considered.

4.5 Application to retinal data

We applied our formulas to *ex-vivo* multi-electrode array recordings of rat retinal ganglion cells in response to black and white checkerboard stimulation [40, 167]. The receptive

fields of the cells have a mosaic structure (Fig. 4.3A), so that neuronal responses show strong stimulus correlations between neighbors, which decay with the distance between the receptive field centers (Fig. 4.3B). Due to network effects [25], nearby cells also show strong noise correlations that decay with distance on a similar length scale (Fig. 4.3C).

We computed the noise synergy using our re-summed approximation (4.2) for many subgroups of nearby cells of different sizes (Fig. 4.3D). In this case it is not possible to estimate mutual informations exactly because of limited data, making it a good test case for the usefulness of our analytical formulas. Spike trains were binned at $15ms$ and, to correct for the bias stemming from noise in estimating correlations, we subtracted the value obtained after shuffling individual cell activities across repetitions. We observe that noise correlations impede information transmission, by the order of 1 bit per neuron per second, for a total information of around 10 bits per neuron per second. It should be stressed however that this result is specific to the white-noise stimulus statistics considered here, and may not be a general feature of retinal processing, as other stimulus statistics would change both the nature of stimulus correlations and the input-output relationship as the network adapts.

We also used our method to study the effect of spiking memory in single neurons, by treating the spike activity of the same neuron in N consecutive $4ms$ time bins as our activity vector $(n_1, \ldots, n_N)$ (treating time bins as we treated individual neurons previously, see Fig. 4.4A). Stimulus temporal auto-correlations are positive for about 50ms (Fig. 4.4B), then become negative and go to zero for longer times (not shown). Noise temporal correlations are driven by refractoriness, which suppresses activity immediately following a spike, and by burstiness, which induces rippling effects up to 50 ms (Fig. 4.4C). We find that these correlations improve information transmission by up to 8 bits per second (Fig. 4.4D), almost doubling it for some cells. This suggests that information is encoded not just in the average spike rate, but also through the control of inter-spike timing, consistent with previous findings [121, 52, 24, 41].

4.6 Discussion

Despite being based on a small-correlation expansion, our analytical predictions, especially (4.2), work well even in the presence of strong correlations, which are ubiquitous in neuroscience [25, 153, 116]. We showed how our results can be applied to same-time correlations between neurons, or to neuron autocorrelations, and they can readily be used on general spatial-temporal correlations.

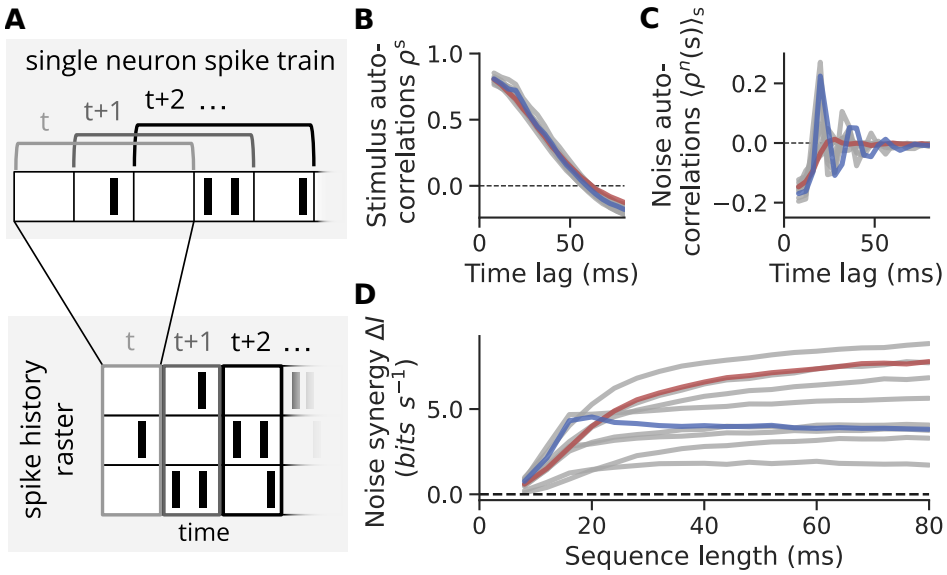

Figure 4.4: **Application on retinal temporal response to visual stimulation.** A) We build a pseudo population of neurons to describe the spiking history of single neurons. B) Stimulus autocorrelations for different cells responding to a white noise stimulation. Highlighted lines correspond to two example cells. C) Same as (B) but for noise autocorrelations. D) Noise synergy for different cells plotted against increasing temporal integration length.

Our work shares some connections with previous efforts to estimate or interpret information in population codes [26, 129, 166, 137, 75, 192, 8]. Ref. [137] proposes decompositions of the mutual information with different interpretations, but does not provide ways to estimate it. Refs. [26] and [75, 192] are mostly based on the Fisher information, which in some limit can be related to the mutual information. While the first term of our simpler expression (4.4) recovers one of their main results—the so-called sign rule—second and higher order terms in the noise correlation parameter provide important corrections when correlations are high, as can be seen from deviations from the initial slope in Fig. 4.2B. In [129] the authors developed a small time bin expansion of the mutual information. Expanding their results for small correlations (and further assuming Poisson distributed spike counts, see App. S4) gives back our second-order expression (4.1). Our method however does not need to assume small time bins, and still works well for large correlations. Ref. [166] provides estimate of the mutual information when the neuronal responses are correlated but have only small fluctuations around a large mean activity, which is not appropriate for small time bins or for low spike rates as in the retina.

Our results are based on the small correlation expansion developed in [155]. In order to apply this theoretical tool, we assumed that both the stimulus-conditioned and the marginal responses follow a pairwise maximum entropy distribution. These models are characterized by many unknown parameters that in principle need to be inferred from

data. However the final expressions for the mutual information contain only quantities that can be directly estimated from data, without needing any inference. This makes our approximations ready and easy to use, without requiring much computational efforts. We showed that it works well even when the data was generated with a very different model. Maximum entropy distributions are actually a series of approximations which, just like Taylor expansions, can be refined by adding higher-order correlations. A future direction could be to compute corrective terms to the mutual information corresponding to third- and higher-order correlation functions, rather than just pairwise correlations as we did in this work. At the same time, the pairwise approximation has proven very accurate for both marginal [153, 59, 177, 58, 54] and conditional [157, 64, 51, 39] responses of populations of neurons, and is only expected to break down for very large densely correlated populations [146]. We thus expect our results to be applicable to a wide array of neuronal contexts.

Acknowledgments

We thank Stéphane Deny for sharing the data and Tobias Kühn for useful discussions. This work was supported by the Agence Nationale de la Recherche (ANR-21-CE37-0024 NatNetNoise, ANR-20-CE37-0018-04 ShootingStar, ANR-18-CE37-0011 DECORE and ANR-22-CE37-0016-02 PerBaCo), by LabEx LIFESENSES (ANR-10-LABX-65), by IHU FOReSIGHT (ANR-18-IAHU-01), by Sorbonne Université with the Emergence program (CrInforNet), by Sorbonne Center for Artificial Intelligence - Sorbonne University - IDEX SUPER 11-IDEX-0004, ERC No 101045253 and by Retina France.

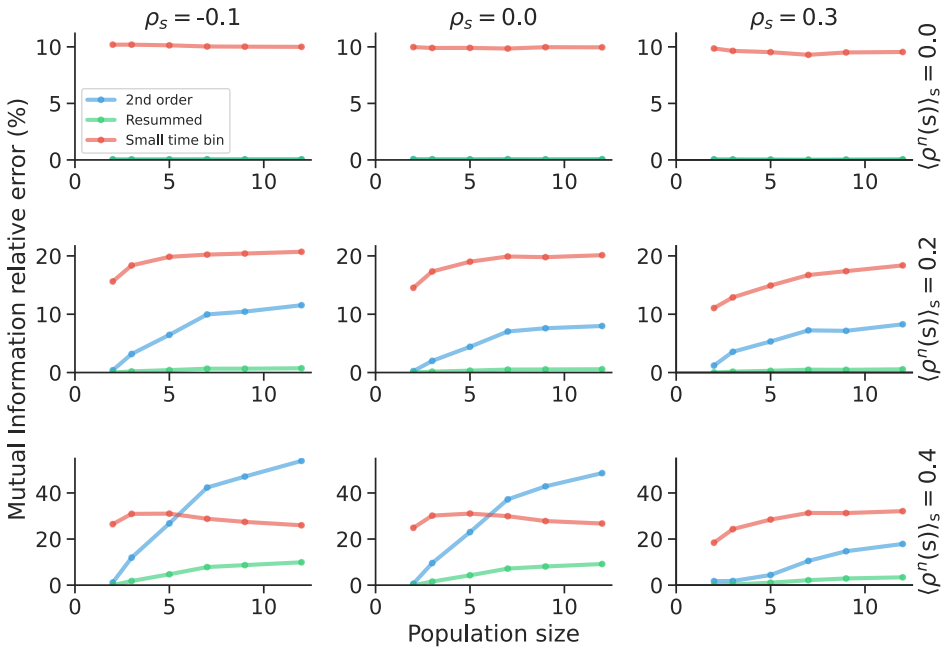

Figure S1: **Relative error $|\delta I|/I$ in the estimate of the mutual information as a function of population size.** Each subplot corresponds to different values of average nearest neighbour noise $\langle \rho^{\mathrm{n}}(s) \rangle_s$ and stimulus ρ_{s} correlations. For population size N, we selected compact groups of N cells out of the network of 12 cells of Fig. 2 (where by compact we mean that their topology minimizes their average distance to the group's center of mass). The error is computed as the difference between the various approximations and the true value estimated by the histogram method, and is then averaged over groups of cells. The resummed approximation systematically outperforms the second order and small time bin expansions for all values of noise and stimulus correlations.

S1 Second order approximation

Sessak-Monasson expansion of the entropy

Sessak and Monasson [155] proposed a small-correlation expansion of the entropy of Ising systems in order to develop an approximated inference method for the model parameters. We used that expansion to obtain approximations of entropies and mutual information in stimulus driven systems.

Here we start by summarising [155] to obtain an estimation of the Ising model entropy in term of mean activities and correlations. We consider a population of N binary neurons $\mathbf{n} = (n_1 ..., n_N)$ described by the following Hamiltonian:

$$\mathcal{H}[\mathbf{n}] = -\sum_{i<j} J_{ij} n_i n_j - \sum_i h_i n_i, \tag{S1.1}$$

where the values of the couplings J_{ij} and fields h_i are determined by the pairwise covariances c_{ij} and means μ_i of the neurons. Noting $Z\left[\{J_{ij}\},\{h_i\}\right]$ the partition function of the system for a given set of couplings and fields,

$$Z\left[\{J_{ij}\},\{h_i\}\right] = \sum_{\mathbf{n}} e^{-\mathcal{H}[\mathbf{n}]}, \tag{S1.2}$$

the neurons means' can be expressed as:

$$\mu_i = \left.\frac{\partial \log\left(Z\left[\{J_{ij}\},\{h_i\}\right]\right)}{\partial h_i}\right|_{J^*_{ij},h^*_i}, \tag{S1.3}$$

and their covariances as:

$$c_{ij} = \left.\frac{\partial \log\left(Z\left[\{J_{ij}\},\{h_i\}\right]\right)}{\partial J_{ij}}\right|_{J^*_{ij},h^*_i} - \mu_i\mu_j. \tag{S1.4}$$

By inverting these relations, we find the parameters of the model that match the observables.

Following [155], we introduce the entropy:

$$H\left[\mathbf{n}\right] = \log\left(Z(\{J_{ij}\},\{h_i\})\right) - \sum_{ij} J_{ij}(\alpha c_{ij} + \mu_i\mu_j) - \sum_i h_i\mu_i. \tag{S1.5}$$

where a parameter α has been introduced to rescale the observed covariances. Minimizing this entropy with respect to h_i and J_{ij} yields the parameters that match the μ_i and c_{ij} of the data, Eqs. S1.3 and S1.4 (up to α). When $\alpha = 1$, this minimization gives the true entropy of the system, a quantity that is in general difficult to compute exactly. The approach proposed by Sessak and Monasson [155] consists of expanding the entropy Eq. S1.5 at small α. This allows one to express the entropy, couplings and fields of the system as power series in α, where each term is a function of the means and covariances of the population. The evaluation of this series in $\alpha = 1$ then yields approximations for the entropy, couplings and fields, that involve only the covariances and means of the population. Following this approach, we can expand the entropy of system up to order 2 in α:

$$H\left[\mathbf{n}\right] = H_{0\text{th}}\left[\mathbf{n}\right] + \alpha H_{1\text{st}}\left[\mathbf{n}\right] + \alpha^2 H_{2\text{nd}}\left[\mathbf{n}\right] + O(\alpha^3). \tag{S1.6}$$

Evaluating the truncated series at $\alpha = 1$ gives the following approximation for the population's entropy:

$$H\left[\mathbf{n}\right] \approx H_{0\text{th}}\left[\mathbf{n}\right] + H_{1\text{st}}\left[\mathbf{n}\right] + H_{2\text{nd}}\left[\mathbf{n}\right]. \tag{S1.7}$$

The first term of this expansion $H_{0\text{th}}$ is the entropy of the neurons in the total absence of correlations and is thus the sum of the entropies of the single neurons:

$$H_{0\text{th}}\left[\mathbf{n}\right] = -\sum_i (1-\mu_i)\log\left(1-\mu_i\right) + \mu_i \log\left(\mu_i\right). \tag{S1.8}$$

The first order contribution $H_{1\text{st}}$ vanishes, and the second order contribution is given in terms of Pearson's correlation coefficients as:

$$H_{2\text{nd}}\left[\mathbf{n}\right] = -\frac{1}{2}\sum_{i<j}\rho_{ij}^2. \tag{S1.9}$$

Mutual Information

We can perform this expansion for the entropy of the marginal distribution $H\left[\mathbf{n}\right]$ as well as for the entropies of the conditional distributions $H\left[\mathbf{n}|\mathbf{s}\right]$, and hence calculate the mutual information to the second order in the correlations:

$$I \approx I_{0\text{th}} + I_{2\text{nd}}, \tag{S1.10}$$

where $I_{0\text{th}}$ is the sum of information carried by the neurons individually $I_{0\text{th}} = \sum_i \left[H[n_i] - \langle H[n_i|\mathbf{s}]\rangle_{\mathbf{s}}\right]$:

$$\begin{aligned} I_{0\text{th}} = &-\sum_i (1-\mu_i)\log\left(1-\mu_i\right) + \mu_i\log\left(\mu_i\right) \\ &+\left\langle \sum_i (1-\mu_i(\mathbf{s}))\log\left(1-\mu_i(\mathbf{s})\right) + \mu_i(\mathbf{s})\log\left(\mu_i(\mathbf{s})\right)\right\rangle_{\mathbf{s}}, \end{aligned} \tag{S1.11}$$

and where the first non-zero contribution from the pairwise correlations in the response is $I_{2\text{nd}}$:

$$I_{2\text{nd}} = -\frac{1}{2}\sum_{i<j}\left(\rho_{ij}^{\text{tot}\,2} - \left\langle \rho_{ij}^{\text{n}}(\mathbf{s})^2\right\rangle_{\mathbf{s}}\right). \tag{S1.12}$$

Noise synergy

Correlations in the marginal response $\rho_{ij}^{\text{tot}} = r_{ij}^{\text{s}} + r_{ij}^{\text{n}}$ boil down to their stimulus contribution in the conditionally independent case $\rho_{ij}^{\text{tot, CI}} = r_{ij}^{\text{s}}$. Thus the noise synergy $\Delta I \equiv I - I_{\text{CI}}$ is given at second oder by:

$$\Delta I \approx \sum_{i<j} \left[-r_{ij}^{\mathrm{n}} r_{ij}^{\mathrm{s}} + \frac{1}{2} \left(\left\langle \rho_{ij}^{\mathrm{n}}(\mathbf{s})^2 \right\rangle_{\mathbf{s}} - {r_{ij}^{\mathrm{n}}}^2 \right) \right]. \tag{S1.13}$$

In general the second term under the sum in the noise synergy, $\Delta I_{ij}^{\mathrm{quad}} = \frac{1}{2}\left(\langle \rho_{ij}^{\mathrm{n}}(\mathbf{s})^2 \rangle_{\mathbf{s}} - {r_{ij}^{\mathrm{n}}}^2\right)$ can be positive or negative depending on the level of correlation between $\rho_{ij}^{\mathrm{n}}(\mathbf{s})$ and $\sqrt{C_{ii}^{\mathrm{n}}(\mathbf{s}) C_{jj}^{\mathrm{n}}(\mathbf{s})}$, as $r_{ij}^{\mathrm{n}} = \langle \rho_{ij}^{\mathrm{n}}(\mathbf{s}) \sqrt{C_{ii}^{\mathrm{n}}(\mathbf{s}) C_{jj}^{\mathrm{n}}(\mathbf{s})} \rangle_{\mathbf{s}} / \sqrt{C_{ii} C_{jj}}$. However, if we assume noise correlations are independent from the stimulus $\rho_{ij}^{\mathrm{n}}(\mathbf{s}) = \rho_{ij}^{\mathrm{n}}$, we can show that $\Delta I_{ij}^{\mathrm{quad}} = \frac{1}{2}\left({\rho_{ij}^{\mathrm{n}}}^2 - {r_{ij}^{\mathrm{n}}}^2\right)$ is non-negative. First, the formulation of r_{ij}^{n} in terms of ρ_{ij}^{n} becomes:

$$r_{ij}^{\mathrm{n}} = \rho_{ij}^{\mathrm{n}} \frac{\left\langle \sqrt{C_{ii}^{\mathrm{n}}(\mathbf{s}) C_{jj}^{\mathrm{n}}(\mathbf{s})} \right\rangle_{\mathrm{s}}}{\sqrt{C_{ii} C_{jj}}}, \tag{S1.14}$$

which then gives:

$$\Delta I_{ij}^{\mathrm{quad}} = \frac{1}{2} {\rho_{ij}^{\mathrm{n}}}^2 \left(1 - \frac{\left\langle \sqrt{C_{ii}^{\mathrm{n}}(\mathbf{s}) C_{jj}^{\mathrm{n}}(\mathbf{s})} \right\rangle_{\mathrm{s}}^2}{C_{ii} C_{jj}} \right). \tag{S1.15}$$

According to the Cauchy-Schwartz inequality we have that $\left\langle \sqrt{C_{ii}^{\mathrm{n}}(\mathbf{s}) C_{jj}^{\mathrm{n}}(\mathbf{s})} \right\rangle_{\mathrm{s}}^2 \leq \langle C_{ii}^{\mathrm{n}}(\mathbf{s}) \rangle_{\mathrm{s}} \left\langle C_{jj}^{\mathrm{n}}(\mathbf{s}) \right\rangle_{\mathrm{s}}$. Besides, from the law of total variance we have that $C_{ii} = C_{ii}^{\mathrm{s}} + \langle C_{ii}^{\mathrm{n}}(\mathbf{s}) \rangle_{\mathrm{s}}$, thus $\langle C_{ii}^{\mathrm{n}}(\mathbf{s}) \rangle_{\mathrm{s}} \leq C_{ii}$ and finally $\left\langle \sqrt{C_{ii}^{\mathrm{n}}(\mathbf{s}) C_{jj}^{\mathrm{n}}(\mathbf{s})} \right\rangle_{\mathrm{s}}^2 \leq C_{ii} C_{jj}$. Altogether this gives that $\Delta I_{ij}^{\mathrm{quad}} \geq 0$.

S2 Resummed expansion

Entropy

In [155] it is shown that some of the terms in the small-correlation expansion of the couplings can be resummed to yield a better approximation. We can proceed in the exact same way for the entropy and resum some of the diagrams in the small-correlation expansion. All the terms of the second order approximation above are contained in the resummed expansion we detail here. Note that this resummed expansion is equivalent to a cluster expansion truncated to second order, with mean-field reference entropy [35]. It reads:

$$H[\mathbf{n}] \approx H_{0\mathrm{th}}[\mathbf{n}] + H_{\mathrm{pairs}}[\mathbf{n}] + H_{\mathcal{G}}[\mathbf{n}] - H_{\mathrm{dbl}}[\mathbf{n}]. \tag{S2.1}$$

The single site contribution $H_{0\mathrm{th}}[\mathbf{n}]$ is the same as above, and $H_{\mathrm{pairs}}[\mathbf{n}]$ corresponds to the entropy gain of all pairs in the population taken independently compared to the single

site contribution. Interestingly, taking the sum of these two first contributions amounts to making an independent pair approximation, which would be exact in the case of a tree-like network topology. $H_{\text{pairs}}[\mathbf{n}]$ is a sum over all pairs of neurons in the population:

$$H_{\text{pairs}}[\mathbf{n}] = \sum_{i<j} H[n_i, n_j], \tag{S2.2}$$

with $H[n_i, n_j]$ the entropy gain of pair (ij) compared to the single neurons case:

$$\begin{aligned} H[n_i, n_j] = &- (C_{ij} + \mu_i\mu_j) \log\left(1 + \frac{C_{ij}}{\mu_i\mu_j}\right) \\ &+ (C_{ij} + \mu_i(\mu_j - 1)) \log\left(1 + \frac{C_{ij}}{\mu_i(\mu_j - 1)}\right) \\ &+ (C_{ij} + \mu_j(\mu_i - 1)) \log\left(1 + \frac{C_{ij}}{\mu_j(\mu_i - 1)}\right) \\ &- (C_{ij} + (1 - \mu_i)(1 - \mu_j)) \log\left(1 + \frac{C_{ij}}{(1 - \mu_i)(1 - \mu_j)}\right). \end{aligned} \tag{S2.3}$$

The second resummed term of this expansion $H_{\mathcal{G}}[\mathbf{n}]$ corresponds to the contribution of interactions to the entropy in the mean-field approximation. It contains the resummation of all loop diagrams in the expansion and amounts to assuming the entropic contribution of pairwise correlations is Gaussian. Noting ρ the correlation matrix of $\mathbf{n}$ we get:

$$H_{\mathcal{G}}[\mathbf{n}] = \frac{1}{2} \log(|\rho|). \tag{S2.4}$$

Finally there are some terms in the expansion that are resummed both in $H_{\text{pairs}}[\mathbf{n}]$ and $H_{\mathcal{G}}[\mathbf{n}]$, therefore we need to substract them once from the expansion through $H_{\text{dbl}}[\mathbf{n}]$. The double counted terms simply correspond to the Gaussian approximation applied to each pair of neurons:

$$H_{\text{dbl}}[\mathbf{n}] = \frac{1}{2} \sum_{i<j} \log\left(1 - \rho_{ij}^2\right). \tag{S2.5}$$

Mutual Information

Applying the resummed approximation to the marginal and conditional responses results in a resummed approximation for the mutual information:

$$I \approx I_{0\text{th}} + I_{\text{pairs}} + I_{\mathcal{G}} - I_{\text{dbl}}. \tag{S2.6}$$

With I_{pairs} the contribution of pairwise correlations to the mutual information in the independent pairs approximation. If we write $C_{ij} \equiv \text{Cov}(n_i, n_j)$ the total covariance across stimuli and $C^{\text{n}}_{ij}(\mathbf{s}) \equiv \text{Cov}(n_i, n_j|\mathbf{s})$ the covariance at a given stimulus $\mathbf{s}$:

$$\begin{aligned}
I_{\text{pairs}} = \sum_{i<j} \Bigg[& -(C_{ij} + \mu_i\mu_j) \log\left(1 + \frac{C_{ij}}{\mu_i\mu_j}\right) \\
& + (C_{ij} + \mu_i(\mu_j - 1)) \log\left(1 + \frac{C_{ij}}{\mu_i(\mu_j - 1)}\right) \\
& + (C_{ij} + \mu_j(\mu_i - 1)) \log\left(1 + \frac{C_{ij}}{\mu_j(\mu_i - 1)}\right) \\
& - (C_{ij} + (1 - \mu_i)(1 - \mu_j)) \log\left(1 + \frac{C_{ij}}{(1 - \mu_i)(1 - \mu_j)}\right) \\
& - \Bigg\langle -(C^{\text{n}}_{ij}(\mathbf{s}) + \mu_i(\mathbf{s})\mu_j(\mathbf{s})) \log\left(1 + \frac{C^{\text{n}}_{ij}(\mathbf{s})}{\mu_i(\mathbf{s})\mu_j(\mathbf{s})}\right) \\
& + (C^{\text{n}}_{ij}(\mathbf{s}) + \mu_i(\mathbf{s})(\mu_j(\mathbf{s}) - 1)) \log\left(1 + \frac{C^{\text{n}}_{ij}(\mathbf{s})}{\mu_i(\mathbf{s})(\mu_j(\mathbf{s}) - 1)}\right) \\
& + (C^{\text{n}}_{ij}(\mathbf{s}) + \mu_j(\mathbf{s})(\mu_i(\mathbf{s}) - 1)) \log\left(1 + \frac{C^{\text{n}}_{ij}(\mathbf{s})}{\mu_j(\mathbf{s})(\mu_i(\mathbf{s}) - 1)}\right) \\
& - (C^{\text{n}}_{ij}(\mathbf{s}) + (1 - \mu_i(\mathbf{s}))(1 - \mu_j(\mathbf{s}))) \log\left(1 + \frac{C^{\text{n}}_{ij}(\mathbf{s})}{(1 - \mu_i(\mathbf{s}))(1 - \mu_j(\mathbf{s}))}\right) \Bigg\rangle_{\mathbf{s}} \Bigg].
\end{aligned} \tag{S2.7}$$

If we denote by ρ^{tot} the total correlation matrix across stimuli and $\rho^{\text{n}}(\mathbf{s})$ the correlation matrix at given stimulus $\mathbf{s}$, the Gaussian (i.e. mean-field) contribution of correlations to the mutual information takes the simple form:

$$I_{\mathcal{G}} = \frac{1}{2} \log\left(|\rho^{\text{tot}}|\right) - \frac{1}{2} \left\langle \log\left(|\rho^{\text{n}}(\mathbf{s})|\right) \right\rangle_{\mathbf{s}}, \tag{S2.8}$$

while the double counting correction becomes:

$$I_{\text{dbl}} = \frac{1}{2} \sum_{i<j} \left[\log\left(1 - \rho^{\text{tot}\,2}_{ij}\right) - \left\langle \log\left(1 - \rho^{\text{n}}_{ij}(\mathbf{s})^2\right) \right\rangle_{\mathbf{s}} \right]. \tag{S2.9}$$

Noise synergy

Likewise we can write a resummed approximation of the noise synergy:

$$\Delta I \approx \Delta I_{\text{pairs}} + \Delta I_{\mathcal{G}} - \Delta I_{\text{dbl}}. \tag{S2.10}$$

Noting that the total covariance across stimuli can be decomposed in terms of noise covariance and stimulus covariance $C_{ij} = C^{\text{s}}_{ij} + C^{\text{n}}_{ij}$, with $C^{\text{n}}_{ij} = \langle C^{\text{n}}_{ij}(\mathbf{s})\rangle_{\text{s}}$, the independent pairs approximation of the noise synergy ΔI_{pairs} would be given by:

$$\begin{aligned}
\Delta I_{\text{pairs}} = \sum_{i<j} \Bigg[& -C^{\text{n}}_{ij} \log(1 + \frac{C_{ij}}{(\mu_i(\mu_j - 1) + C_{ij})(\mu_j(\mu_i - 1) + C_{ij})}) \\
& - (C^{\text{s}}_{ij} + \mu_i\mu_j) \log(1 + \frac{C^{\text{n}}_{ij}}{\mu_i\mu_j + C^{\text{s}}_{ij}}) \\
& + (C^{\text{s}}_{ij} + \mu_i(\mu_j - 1)) \log(1 + \frac{C^{\text{n}}_{ij}}{\mu_i(\mu_j - 1) + C^{\text{s}}_{ij}}) \\
& + (C^{\text{s}}_{ij} + \mu_j(\mu_i - 1)) \log(1 + \frac{C^{\text{n}}_{ij}}{\mu_j(\mu_i - 1) + C^{\text{s}}_{ij}}) \\
& - (C^{\text{s}}_{ij} + (1 - \mu_i)(1 - \mu_j)) \log(1 + \frac{C^{\text{n}}_{ij}}{(1 - \mu_i)(1 - \mu_j) + C^{\text{s}}_{ij}}) \\
& - \Bigg\langle -(C^{\text{n}}_{ij}(\mathbf{s}) + \mu_i(\mathbf{s})\mu_j(\mathbf{s})) \log\left(1 + \frac{C^{\text{n}}_{ij}(\mathbf{s})}{\mu_i(\mathbf{s})\mu_j(\mathbf{s})}\right) \\
& + (C^{\text{n}}_{ij}(\mathbf{s}) + \mu_i(\mathbf{s})(\mu_j(\mathbf{s}) - 1)) \log\left(1 + \frac{C^{\text{n}}_{ij}(\mathbf{s})}{\mu_i(\mathbf{s})(\mu_j(\mathbf{s}) - 1)}\right) \\
& + (C^{\text{n}}_{ij}(\mathbf{s}) + \mu_j(\mathbf{s})(\mu_i(\mathbf{s}) - 1)) \log\left(1 + \frac{C^{\text{n}}_{ij}(\mathbf{s})}{\mu_j(\mathbf{s})(\mu_i(\mathbf{s}) - 1)}\right) \\
& - (C^{\text{n}}_{ij}(\mathbf{s}) + (1 - \mu_i(\mathbf{s}))(1 - \mu_j(\mathbf{s}))) \log\left(1 + \frac{C^{\text{n}}_{ij}(\mathbf{s})}{(1 - \mu_i(\mathbf{s}))(1 - \mu_j(\mathbf{s}))}\right) \Bigg\rangle_{\text{s}} \Bigg].
\end{aligned} \tag{S2.11}$$

The Gaussian contribution to the noise synergy takes again a simple form:

$$\Delta I_{\mathcal{G}} = \frac{1}{2} \log\left(\frac{|\rho^{\text{tot}}|}{|\rho^{\text{tot,CI}}|}\right) - \frac{1}{2} \left\langle \log\left(|\rho^{\text{n}}(\mathbf{s})|\right)\right\rangle_{\mathbf{s}}, \tag{S2.12}$$

where we recall that correlations in the marginal response can be expressed as the sum of stimulus and noise contributions $\rho^{\text{tot}} = r^{\text{s}} + r^{\text{n}}$. In the conditionally independent case we have $\rho^{\text{tot,CI}} = r^{\text{s}} + \nu^{\text{n}}$ with ν^{n} the diagonal matrix containing the diagonal elements of r^{n}.

Finally, the double counting correction to the noise synergy is given by:

$$\Delta I^{dbl} = \frac{1}{2}\sum_{i<j}\left[\log\left(\frac{1-\rho_{ij}^{\mathrm{tot}\,2}}{1-r_{ij}^{\mathrm{s}\,2}}\right) - \left\langle \log\left(1-\rho_{ij}^{\mathrm{n}}(\mathbf{s})^2\right)\right\rangle_{\mathrm{s}}\right]. \tag{S2.13}$$

S3 Generalized linear model simulations

The model used to generate the synthetic data is a Generalized Linear Model [135] with sigmoidal nonlinearity. The number of spikes emitted by cell i in time bin t of size $dt = 1$ ms follows a Bernoulli distribution with mean $\lambda_i(t)$, given by:

$$\lambda_i(t) = \left(1+\exp\left(-h_i(t)\right)\right)^{-1}, \quad h_i(t) = h_i^{\mathrm{bias}} + h_i^{\mathrm{stim}}(t) + h_i^{\mathrm{int}}(t) + h_i^{\mathrm{corr}}(t), \tag{S3.1}$$

where h_i^{bias} sets the baseline firing rate of the cell, $h_i^{\mathrm{stim}}(t)$ accounts for how the stimulus drives the cell's activity, $h_i^{\mathrm{int}}(t)$ accounts for the effect of couplings and self-coupling, while $h_i^{\mathrm{corr}}(t)$ is here to correct for the contribution of the neuron-neuron couplings to the firing rate.

The stimulus S is a movie of dimensions (N_x, N_y, N_t) where N_x and N_y correspond to the two spatial dimensions in pixels, and N_t is the temporal length of the stimulus in number of time bins. We simulated 12 cells organized on a triangular lattice as represented in Fig. S2B, spaced by $\xi = 4$ pixels. Here the stimulus consisted of N_t zero-mean, 2D Gaussian frames of size (N_x, N_y) with covariance function $C(\mathbf{u},\mathbf{v}) = \delta_{\mathbf{u},\mathbf{v}} + (1-\delta_{\mathbf{u},\mathbf{v}})c_0 \exp\left(-\|\mathbf{u}-\mathbf{v}\|/\lambda\right)$, where $\lambda = 2\xi$ and c_0 is varied to change the level of stimulus correlation in the response. There was no correlations between frames and they were refreshed at 100Hz.

The cells' stimulus filters, of size $(N_x, N_y, N_t^{\mathrm{stim}})$ were built from the product of a temporal and spatial component which were chosen to mimick the properties of natural retinal ganglion cells. The temporal component $K^{\mathrm{time}}(\tau)$ is illustrated Fig. S2D and consists of a difference of two raised cosine functions:

$$\mathrm{rc}(\tau, s, c) = \begin{cases} \cos((\log(\tau+s)-c)\pi)/2 + 1/2 & \text{when } -1 \le \log(\tau+s)-c \le 1 \\ \\ 0 & \text{otherwise,} \end{cases} \tag{S3.2}$$

such that:

$$K^{\mathrm{time}}(\tau) = a_1 \mathrm{rc}(\tau, s, c_1) - a_2 \mathrm{rc}(\tau, s, c_2), \tag{S3.3}$$

with $s = 50$, $a_1 = 0.35$, $c_1 = 5.3$, $a_2 = 1.15$ and $c_2 = 4.8$. The spatial component is made of a difference of Gaussian functions and exhibits a positive center and a negative surround,

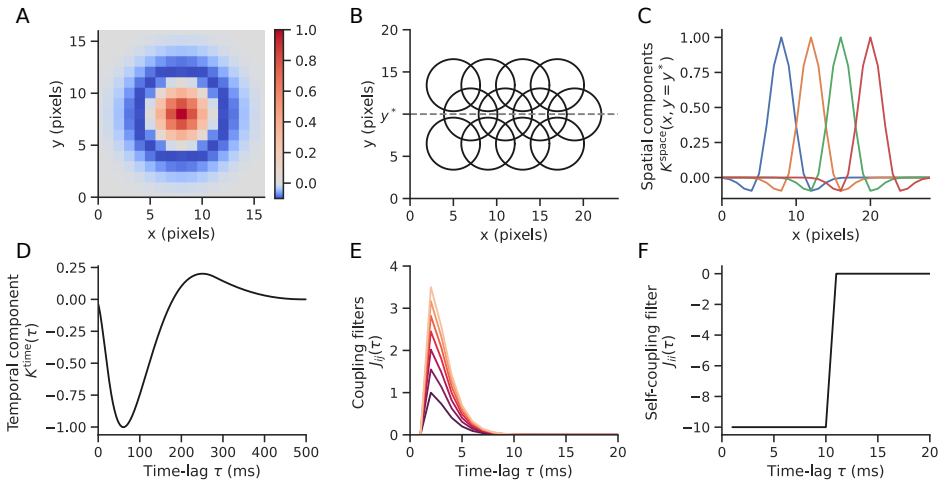

Figure S2: **Parameters of the GLM model.** The different parameters of the model were chosen to be biologically plausible and mimick those of retinal ganglion cells. A) The spatial component of the spatio-temporal stimulus filter K^{space} is made of a difference of Gaussians Eq. S3.4. B) The 12 cells are arranged according to a triangular lattice. We represented here the receptive fields of the cells by circles that correspond to the contour where the center and the surround of the spatial components compensate exactly. Only nearest neighbours in this population are coupled by non-zero J_{ij} coupling filters. C) Profile view of the spatial components cut in $y = y^*$ on panel B. D) The temporal component of the stimulus filter has a biphasic profile and consists of a difference of two raised cosine functions Eq. S3.3. E) Coupling filters are non zero for nearest neighbours and are defined by Eq. S3.8. The increasing coupling amplitude shown here induces increasing noise correlations in the response of the cells. F) The self coupling filter, accounting for the effect of the neuron's own spiking history, induced refractory effects over $10\,$ms in the past.

as shown on Fig. S2A and Fig. S2C. Noting $\mathbf{u_i^c} = (x_i^c, y_i^c)$ the spatial position of the receptive field center of cell i:

$$K_i^{\text{space}}(\mathbf{u}) = \frac{1}{r-1}\left[r\exp\left(-\frac{\|\mathbf{u}-\mathbf{u_i^c}\|^2}{2\sigma_{\text{center}}}\right) - \exp\left(-\frac{\|\mathbf{u}-\mathbf{u_i^c}\|^2}{2\sigma_{\text{surround}}}\right)\right] \tag{S3.4}$$

where $\sigma_{\text{center}} = 2$ pixel, $\sigma_{\text{surround}} = 2.1$ pixel, and $r = 1.12$. The firing rate variance is fixed through a parameter α_{s} (set to 0.5 in our synthetic experiments) such that $h_i^{\text{stim}}(t) = \alpha_{\text{s}} \cdot z\text{-score}\left(\tilde{h}_i^{\text{stim}}(t)\right)$, where $\tilde{h}_i^{\text{stim}}(t)$ is given by the temporal convolution of stimulus S by the spatio-temporal stimulus filter:

$$\tilde{h}_i^{\text{stim}}(t) = \sum_{\tau>0}\sum_{\mathbf{u}} K_i^{\text{space}}(\mathbf{u}) K^{\text{time}}(\tau) S(\mathbf{u}, t-\tau). \tag{S3.5}$$

Likewise, the spiking history contribution of cell i itself as well as that of the other cells in the network are accounted for by linear convolutions of the spiking histories by a set of temporal coupling filters:

$$h_i^{\text{int}}(t) = \sum_j \sum_{\tau>0} J_{ij}(\tau)\, n_j(t-\tau), \tag{S3.6}$$

where the self-coupling filters $J_{ii}(\tau)$, shown Fig. S2F, are given by:

$$J_{ii}(\tau) = \begin{cases} J^0_{\text{self}} & \text{if } \tau \leq \tau_{\text{refr,}} \\ \\ 0 & \text{otherwise,} \end{cases} \tag{S3.7}$$

with $J^0_{\text{self}} = -10$ and $\tau_{\text{refr}} = 10$ms. The neuron-neuron (i.e. $i \neq j$) couplings J_{ij} follow:

$$J_{ij}(\tau) = \begin{cases} J^0_{\text{coupl}}\, \tau \exp(-\tau) & \text{if i and j are nearest neighbours,} \\ \\ 0 & \text{otherwise,} \end{cases} \tag{S3.8}$$

where the coupling strength J^0_{coupl} can be varied (as illustrated on Fig. S2E) to change the amount of noise correlations in the response.

In order to vary independently stimulus and noise correlations in the neurons' response we introduced the field h_i^{corr} that corrects for the contribution of neuron-neuron couplings to the firing rates of the cells. This correction is needed because changing the strength of the couplings will not only change the amount of noise correlations in the reponse, but also the firing rates of the cells. This corrective field is computed via an iterative inference approach built upon the 2-step inference method [101]. The first step of this iterative

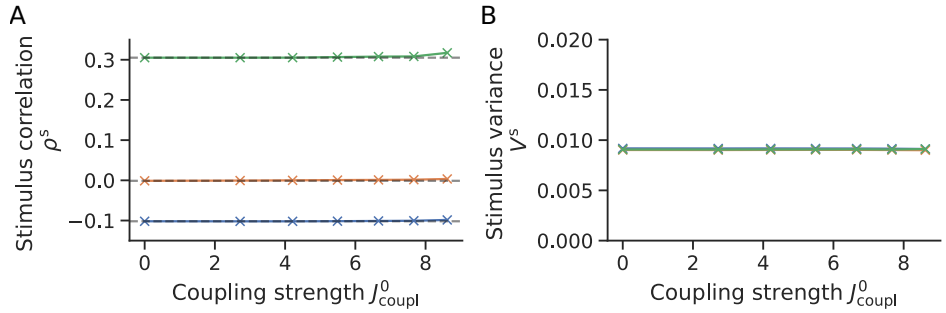

Figure S3: **Results of the iterative inference approach.** A) For the three test stimuli chosen we see that ρ^{s}, the stimulus correlation averaged over all pairs of neighbouring cells, is constant with respect to the coupling strength J^0_{coupl}. B) Likewise, the stimulus variance averaged across the population $V^{\text{s}} = \langle C^{\text{s}}_{ii} \rangle_i$ is independent of J^0_{coupl}.

procedure is the following: for a given stimulus movie, we simulate N_{repe} times the response of the conditionally-independent neurons (i.e. with $J^0_{\text{coupl}} = 0$). We then fix the strength of neuron-neuron couplings J^0_{coupl} to the desired value so as to induce noise correlations in the response of the cells. From here we approximate the interaction field $\hat{h}^{\text{int}}_i$ using the simulated response and the chosen couplings amplitude J^0_{coupl}. Finally we infer the corrective field h^{corr}_i on the previously simulated response similarly to the 2-step inference approach by minimzing the following single neuron negative log-likelihood for each neuron in the population:

$$nLL_i = -\sum_{r=1}^{N_{\text{repe}}} \sum_{t=1}^{N_{\text{t}}} h^{\text{corr}}_i(t) n_i(t,r) + \log\Big(1 - \hat{\lambda}_i(t,r)\Big), \tag{S3.9}$$

where $n_i(t,r)$ denotes the simulated response of neuron i in time bin t and repetition r and where $\hat{\lambda}_i(t,r) = \Big(1 + \exp\Big(-h^{\text{bias}}_i + h^{\text{stim}}_i(t) + \hat{h}^{\text{int}}_i(t,r) + h^{\text{corr}}_i(t)\Big)\Big)^{-1}$. The second step of the procedure consists in simulating the response of the cells using the previously chosen coupling amplitude and the infered corrective field $h^{\text{corr}}_i(t)$. We can then re-estimate the interaction field $\hat{h}^{\text{int}}_i$ on these simulated data and infer again h^{corr}_i to get a better approximation of the couplings' contribution to the firing rate. This second step is repeated as many times as needed to match the firing rates of the conditionally-dependent model to those of the conditionally-independent model within the desired precision. For the generation of the synthetic data used for testing the approximations we systematically performed 20 successive inference steps. This approach resulted for each given stimuli in constant stimulus correlations Fig. S3A as well as constant stimulus variance Fig. S3B across coupling strengths.

The simulation of the model described above was repeated $N_{\text{repe}} = 10^6$ times for each stimulus movie, then the mutual information was computed using the second order and

the resummed approximations as well as by the histogram (or “exact") method. In order to be able to compare the mutual information computed via the approximations to that computed via the histogram method, we first need to correct for the effect of sampling bias on these quantities. From [99] the bias in the entropy of a maximum entropy model with N_{cells} and thus $\frac{N_{\text{cells}}(N_{\text{cells}}+1)}{2}$ constraints, evaluated on N_{samples} can be approximated by $b_{\text{H}}^{\text{maxent}} = -\frac{N_{\text{cells}}(N_{\text{cells}}+1)}{4N_{\text{samples}}}$. In the case of the marginal entropy we have $N_{\text{samples}} = N_{\text{repe}} \times N_t$, while in the case of the noise entropy we have $N_{\text{samples}} = N_{\text{repe}}$. The sampling bias on the mutual information of a maximum entropy model is thus given by:

$$b_{\text{I}}^{\text{maxent}} = \frac{N_{\text{cells}}(N_{\text{cells}}+1)}{4N_{\text{repe}}}(1 - \frac{1}{N_t}), \tag{S3.10}$$

which has then to be subtracted from the raw results of the 2nd order and the resummed approximations. To evaluate the bias on the entropies evaluated via the histogram method we used a shuffling approach similar to [115]: we estimate the bias on the entropy $H[\mathbf{n}]$ as $b_{\text{H}}^{\text{exact}} = H^{\text{shuffle}}[\mathbf{n}] - (H^{0\text{th}}[\mathbf{n}] - b^{0th})$ where $H^{\text{shuffle}}[\mathbf{n}]$ is the entropy computed on the data shuffled so that correlations between cells are destroyed, and where $H^{0\text{th}}[\mathbf{n}]$ is the single site entropy contribution given by Eq.S1.8. $H^{0\text{th}}[\mathbf{n}]$ is also biased, so we need to correct it by $b_H^{0\text{th}} = -\frac{N_{\text{cells}}}{2N_{\text{samples}}}$ (as here we have only one constraint per neuron). Applying this to the marginal and conditional entropies gives the following bias for the mutual information computed via the exact method:

$$b_{\text{I}}^{\text{exact}} = \left(H^{\text{shuffle}}[\mathbf{n}] - \langle H^{\text{shuffle}}[\mathbf{n}|\mathbf{s}]\rangle_{\text{s}}\right) - \left(H^{0\text{th}}[\mathbf{n}] - \langle H^{0\text{th}}[\mathbf{n}|\mathbf{s}]\rangle_{\text{s}}\right) + \frac{N_{\text{cells}}}{2N_{\text{repe}}}(1 - \frac{1}{N_t}), \tag{S3.11}$$

which has to be subtracted from the raw result of the histogram method.

We applied a similar shuffling approach to the small time bin expansion of Ref. [129] in order to correct for the sampling bias in that context (see section S4).

S4 Link to the small time bin expansion (Panzeri *et al.* 1999)

Small time bin expansion

Panzeri et al. [129] introduced a small time bin t expansion of the mutual information:

$$I = tI^t + \frac{t^2}{2}I^{tt} + ... \tag{S4.1}$$

They quantify noise correlations and stimulus correlations in terms of $n_i(\mathbf{s})$ the spike count variable at fixed stimulus s, $\mu_i(\mathbf{s}) = \overline{n_i(\mathbf{s})}$ its average over the noise, $\mu_i = \langle \mu_i(\mathbf{s}) \rangle_{\mathbf{s}}$ the average firing rate across stimuli and use the according rates $n_i(\mathbf{s})/t$, $\mu_i(\mathbf{s})/t$ and μ_i/t in the expression of the mutual information. They introduce the "noise correlation density":

$$\gamma_{ij}(\mathbf{s}) = \begin{cases} \frac{\overline{n_i(\mathbf{s}) n_j(\mathbf{s})}}{\mu_i(\mathbf{s})\mu_j(\mathbf{s})} - 1 & \text{if } i \neq j \\ \\ \frac{\overline{n_i(\mathbf{s})^2} - \mu_i(\mathbf{s})}{\mu_i(\mathbf{s})^2} - 1 & \text{if } i = j, \end{cases} \tag{S4.2}$$

and the "stimulus correlation density":

$$\nu_{ij} = \frac{\langle \mu_i(\mathbf{s}) \mu_j(\mathbf{s}) \rangle_s}{\mu_i \mu_j} - 1. \tag{S4.3}$$

The first order contribution (sum over the single cell contributions) to the mutual information in the small time bin expansion is:

$$t I^t = \sum_i \left\langle \mu_i(\mathbf{s}) \ln \frac{\mu_i(\mathbf{s})}{\mu_i} \right\rangle_s, \tag{S4.4}$$

The second order contribution contains the correlations contributions:

$$\frac{t^2}{2} I^{tt} = \frac{t^2}{2} \sum_{ij} I_{ij}^{tt,(1)} + I_{ij}^{tt,(2)} + I_{ij}^{tt,(3)}, \tag{S4.5}$$

with

$$I_{ij}^{tt,(1)} = \frac{1}{t^2} \mu_i \mu_j \left[\nu_{ij} + (1 + \nu_{ij}) \ln \left(\frac{1}{1 + \nu_{ij}} \right) \right], \tag{S4.6}$$

$$I_{ij}^{tt,(2)} = \frac{1}{t^2} \langle \mu_i(\mathbf{s}) \mu_j(\mathbf{s}) \gamma_{ij}(\mathbf{s}) \rangle_s \ln \left(\frac{1}{1 + \nu_{ij}} \right), \tag{S4.7}$$

$$I_{ij}^{tt,(3)} = \frac{1}{t^2} \left\langle \mu_i(\mathbf{s}) \mu_j(\mathbf{s}) \left(1 + \gamma_{ij}(\mathbf{s})\right) \times \ln \left(\frac{\left(1 + \gamma_{ij}(\mathbf{s})\right) \langle \mu_i(\mathbf{s}') \mu_j(\mathbf{s}') \rangle_{s'}}{\langle \mu_i(\mathbf{s}') \mu_j(\mathbf{s}') \left(1 + \gamma_{ij}(\mathbf{s}')\right) \rangle_{s'}} \right) \right\rangle_s. \tag{S4.8}$$

The three contributions in I^{tt} render the effects of stimulus and noise correlations as well as interactions thereof [129]: $I_{ij}^{tt,(1)}$ contains the effect of signal correlations, while $I_{ij}^{tt,(2)}$ accounts for how (stimulus independent) noise correlations interact with stimulus correlations and affect information, and $I_{ij}^{tt,(3)}$ contains information carried by the stimulus-dependency of noise correlations.

These expressions are also biased by finite sampling. The bias, which should be subtracted from the estimate, can be computed as the contribution of noise correlations to the mutual information, evaluated on the shuffled data. In the decomposition of Eq. S4.5, this

reads:

$$b_{\mathrm{I}}^{\text{small-t}} = \frac{t^2}{2} \sum_{i \neq j} I_{ij}^{tt,(2),\text{shuffle}} + I_{ij}^{tt,(3),\text{shuffle}}, \tag{S4.9}$$

where the two terms under the sum are defined in Eqs. S4.7 and S4.8 and computed on the shuffled data.

Link to the small correlation expansion

The single site contribution of the small time bin approximation tI_{t} amounts to assuming information is conveyed only by spikes rather than by spikes and silences together. To illustrate this, we can rewrite the expression obtained previously for $I_{0\mathrm{th}}$ by regrouping the noise entropy and marginal entropy contributions separately for spikes on one hand, and silences on the other hand:

$$I_{0\mathrm{th}} = \sum_i \left\langle \mu_i(\mathbf{s}) \log \left(\frac{\mu_i(\mathbf{s})}{\mu_i} \right) \right\rangle_{\mathbf{s}} + \sum_i \left\langle (1-\mu_i(\mathbf{s})) \log \left(\frac{1-\mu_i(\mathbf{s})}{1-\mu_i} \right) \right\rangle_{\mathbf{s}}. \tag{S4.10}$$

We see the first term in this rewriting of $I_{0\mathrm{th}}$ corresponds to tI^t. Further corrections to the single site information I_{t} are found in the diagonal terms under the sum in I^{tt}. In the small rates (i.e. Poisson) limit however, this correction is simplified as $\gamma_{ii}(\mathbf{s})$ vanishes for $C_{ii}^{\mathrm{n}}(\mathbf{s}) = \mu_i(\mathbf{s})$. The single site correction coming from I^{tt} is therefore given by:

$$\frac{t^2}{2} \sum_i I_{ii}^{tt,(1)} = \frac{1}{2} \sum_i {\mu_i}^2 \left[\nu_{ii} + (1+\nu_{ii}) \ln \left(\frac{1}{1+\nu_{ii}} \right) \right]. \tag{S4.11}$$

The out of diagonal terms ($i \neq j$) in I^{tt} account for the effect of noise and stimulus cross-correlations. In the small rates and small correlations limit, we show here that we recover the main result of this paper. First we notice that in this limit:

$$\nu_{ij} = \frac{r_{ij}^{\mathrm{s}}}{\sqrt{C_{ii}C_{jj}}} = \frac{r_{ij}^{\mathrm{s}}}{\sqrt{\mu_i \mu_j}}, \tag{S4.12}$$

$$\gamma_{ij}(\mathbf{s}) = \frac{\rho_{ij}^{\mathrm{n}}(\mathbf{s})}{\sqrt{C_{ii}^{\mathrm{n}}(\mathbf{s}) C_{jj}^{\mathrm{n}}(\mathbf{s})}} = \frac{\rho_{ij}^{\mathrm{n}}(\mathbf{s})}{\sqrt{\mu_i(\mathbf{s})\mu_j(\mathbf{s})}}. \tag{S4.13}$$

Replacing these expressions in $I_{ij}^{tt,(1)}$, $I_{ij}^{tt,(2)}$ and $I_{ij}^{tt,(3)}$ then expanding the logarithms at small r_{ij}^{s} and $\rho_{ij}^{\mathrm{n}}(\mathbf{s})$ (and thus r_{ij}^{n}) and truncating at second order gives:

$$\begin{aligned}\frac{t^2}{2}I_{ij}^{tt,(1)} &= \mu_i\mu_j\left[\frac{r_{ij}^{\mathrm{s}}}{\sqrt{\mu_i\mu_j}} - \left(1+\frac{r_{ij}^{\mathrm{s}}}{\sqrt{\mu_i\mu_j}}\right)\ln\left(1+\frac{r_{ij}^{\mathrm{s}}}{\sqrt{\mu_i\mu_j}}\right)\right] \\ &\approx -\frac{1}{2}{r_{ij}^{\mathrm{s}}}^2, \end{aligned} \tag{S4.14}$$

$$\begin{aligned}\frac{t^2}{2}I_{ij}^{tt,(2)} &= -\left\langle \rho_{ij}^{\mathrm{n}}(\mathbf{s})\sqrt{C_{ii}^{\mathrm{n}}(\mathbf{s})C_{jj}^{\mathrm{n}}(\mathbf{s})}\right\rangle_s \ln\left(1+\frac{r_{ij}^{\mathrm{s}}}{\sqrt{\mu_i\mu_j}}\right) \\ &= -C_{ij}^{\mathrm{n}}\ln\left(1+\frac{r_{ij}^{\mathrm{s}}}{\sqrt{\mu_i\mu_j}}\right) \\ &\approx -r_{ij}^{\mathrm{n}}r_{ij}^{\mathrm{s}}, \end{aligned} \tag{S4.15}$$

$$\begin{aligned}\frac{t^2}{2}I_{ij}^{tt,(3)} &= \left\langle \left(\mu_i(\mathbf{s})\mu_j(\mathbf{s}) + \rho_{ij}(\mathbf{s})\sqrt{C_{ii}^{\mathrm{n}}(\mathbf{s})C_{jj}^{\mathrm{n}}(\mathbf{s})}\right)\times\left[\ln\left(1+\frac{\rho_{ij}^{\mathrm{n}}(\mathbf{s})}{\sqrt{\mu_i(\mathbf{s})\mu_j(\mathbf{s})}}\right)\right.\right. \\ &\quad \left.\left. - \ln\left(1+\frac{\langle\rho_{ij}(\mathbf{s}')\sqrt{C_{ii}^{\mathrm{n}}(\mathbf{s}')C_{jj}^{\mathrm{n}}(\mathbf{s}')}\rangle_{\mathbf{s}'}}{\langle\mu_i(\mathbf{s}')\mu_j(\mathbf{s}')\rangle_{\mathbf{s}'}}\right)\right]\right\rangle_{\mathbf{s}} \\ &= \left\langle \left(\mu_i(\mathbf{s})\mu_j(\mathbf{s}) + C_{ij}^{\mathrm{n}}(\mathbf{s})\right)\times\left[\ln\left(1+\frac{\rho_{ij}^{\mathrm{n}}(\mathbf{s})}{\sqrt{\mu_i(\mathbf{s})\mu_j(\mathbf{s})}}\right) - \ln\left(1+\frac{C_{ij}^{\mathrm{n}}}{C_{ij}^{\mathrm{s}}+\mu_i\mu_j}\right)\right]\right\rangle_{\mathbf{s}} \\ &= \left\langle \left(\mu_i(\mathbf{s})\mu_j(\mathbf{s}) + C_{ij}^{\mathrm{n}}(\mathbf{s})\right)\times\ln\left(1+\frac{\rho_{ij}^{\mathrm{n}}(\mathbf{s})}{\sqrt{\mu_i(\mathbf{s})\mu_j(\mathbf{s})}}\right)\right\rangle_{\mathbf{s}} \\ &\quad - \mu_i\mu_j\left(1+\frac{r_{ij}^{\mathrm{s}}}{\sqrt{\mu_i\mu_j}}+\frac{r_{ij}^{\mathrm{n}}}{\sqrt{\mu_i\mu_j}}\right)\times\ln\left(1+\frac{r_{ij}^{\mathrm{n}}/\sqrt{\mu_i\mu_j}}{r_{ij}^{\mathrm{s}}/\sqrt{\mu_i\mu_j}+1}\right) \\ &\approx \frac{1}{2}\left\langle\rho_{ij}^{\mathrm{n}}(\mathbf{s})^2\right\rangle_{\mathbf{s}} - \frac{1}{2}{r_{ij}^{n}}^2. \end{aligned} \tag{S4.16}$$

Summing up these contributions gives:

$$\begin{aligned}\frac{t^2}{2}I^{tt} &= \frac{t^2}{2}\sum_{ij} I_{ij}^{tt,(1)} + I_{ij}^{tt,(2)} + I_{ij}^{tt,(3)} \\ &= \frac{t^2}{2}\sum_{i} I_{ii}^{tt,(1)} + t^2\sum_{i<j} I_{ij}^{tt,(1)} + I_{ij}^{tt,(2)} + I_{ij}^{tt,(3)} \\ &\approx \frac{1}{2}\sum_i {\mu_i}^2\left[\nu_{ii} + (1+\nu_{ii})\ln\left(\frac{1}{1+\nu_{ii}}\right)\right] - \frac{1}{2}\sum_{i<j}\left((r_{ij}^{\mathrm{s}}+r_{ij}^{n})^2 - \left\langle\rho_{ij}^{\mathrm{n}}(\mathbf{s})^2\right\rangle_{\mathbf{s}}\right) \\ &\approx \frac{1}{2}\sum_i {\mu_i}^2\left[\nu_{ii} + (1+\nu_{ii})\ln\left(\frac{1}{1+\nu_{ii}}\right)\right] - \frac{1}{2}\sum_{i<j}\left({\rho_{ij}^{\mathrm{tot}}}^2 - \left\langle\rho_{ij}^{\mathrm{n}}(\mathbf{s})^2\right\rangle_{\mathbf{s}}\right), \end{aligned} \tag{S4.17}$$

and finally:

$$\begin{aligned} I &\approx tI^t + \frac{t^2}{2} I^{tt} \\ &\approx I_{\text{single}} + I_{\text{2nd}}, \end{aligned} \tag{S4.18}$$

with the single site term given by:

$$I_{\text{single}} = \sum_i \left\langle \mu_i(\mathbf{s}) \ln \frac{\mu_i(\mathbf{s})}{\mu_i} \right\rangle_s + \frac{1}{2} \sum_i {\mu_i}^2 \left[\nu_{ii} + (1+\nu_{ii}) \ln \left(\frac{1}{1+\nu_{ii}} \right) \right], \tag{S4.19}$$

and the second order cross-correlations contribution by:

$$I_{\text{2cd}} = -\frac{1}{2} \sum_{i<j} \left({\rho_{ij}^{\text{tot}}}^2 - \left\langle {\rho_{ij}^{\text{n}}(\mathbf{s})}^2 \right\rangle_{\mathbf{s}} \right). \tag{S4.20}$$

In the small correlations and small rates (or Poisson) limit, the contribution to the mutual information of cross-correlations as described by the small time bin expansion coincides with the second order approximation derived in this paper.

Interpretation of the small time bin and small correlation approximations

The agreement between the small time bin and small correlation approximations raises questions about the interpretation of how noise correlations contribute to the mutual information. On the basis of the small time bin expansion [129] and its generalization [137], it is often stated in the litterature that noise correlations increase mutual information in two cases: when they are of opposite sign than stimulus correlations and follow the “sign rule", as seen in the expression of $I_{ij}^{tt,(2)}$ (Eq. S4.7); or when they fluctuate with the stimulus, as seen in the expression of $I_{ij}^{tt,(3)}$ (Eq. S4.8). Therefore, constant noise correlations of the same sign than stimulus correlations should not benefit the mutual information in any way. However, that interpretation relies on quantifying noise correlations and their variations through the correlation density $\gamma_{ij}(\mathbf{s})$, which do not correspond to correlation coefficients in the classical sense.

By contrast, as we showed in App. S1, where results are expressed in terms of the more familiar Pearson correlations $\rho_{ij}^{\text{n}}(\mathbf{s})$, constant noise correlations contribute positively to the mutual information (Eq. S1.15). This different interpretation is not a consequence of the incompatibility of the two approaches (as shown by Eq. S4.16), but rather of the two definitions of “constant noise correlations”. In the case of constant Pearson noise correlations $\rho_{ij}^{\text{n}}(\mathbf{s})$, the noise correlation density $\gamma_{ij}(\mathbf{s})$ defined in Eq. S4.2 actually varies

with the stimulus, thereby positively contributing to the mutual information according to the original interpretation. Therefore, it is important in these discussions to specify how noise correlations are defined: $\gamma_{ij}(\mathbf{s})$, $\rho^n_{ij}(\mathbf{s})$, or something else.

Chapter 5

Strong but not weak noise correlations are beneficial for population coding

We have seen in the introductory chapters of this thesis that the general picture emerging from the literature regarding how noise correlations affect sensory coding can be summarized through the "sign-rule". This rule suggests that noise correlations will be beneficial when of opposite sign than stimulus correlations, and detrimental otherwise. Decades of theoretical investigations led to the widely admitted idea that positive noise correlations, such as the ones observed in ganglion cells populations (see chapter 1), will be detrimental if they correlate neurons that also tend to be correlated by the stimulus (see chapter 2). These results are at odds with the structure of noise and signal correlations that are observed in many sensory systems, where noise correlations are often found to be strong and positive between cells that also share sensitivity to the same stimuli.

In the previous part, we addressed the fact that assessing the impact of noise correlations on sensory coding from the mutual information standpoint is intrinsically difficult. We overcame this issue by deriving an approximation for the mutual information that allows us to relate directly the structure of pairwise noise and signal correlations to their impact on stimulus information encoding.

In this chapter, we build on the approximation and insights derived previously to investigate how strong positive noise correlations as those observed in the retina affect sensory coding. Our approximation — which has been shown in the previous chapter to provide accurate measures of mutual information for retina-like populations of intermediate size, even under significant stimulus and noise correlation strengths — is applied here to experimental recordings of OFF-α ganglion cells from the rodent retina. We show that the impact of positive noise correlations strongly depends on the stimulus, but most importantly

that strong noise correlations can lead to violations of the sign-rule. In contrast to prior studies, which have mostly focused on Fisher information [45, 166, 75, 57] or optimality considerations [178, 75, 117], we derive new quantitative criteria for the value of noise correlations that predict their impact on mutual information between stimulus and response. We expand the insight derived from our approximation to a Gaussian population model, demonstrating the existence of a critical value beyond which noise correlations become beneficial. We further investigate the impact of noise correlations on the encoding of different features of the stimulus and show that, both in model and data, positive noise correlations favor the encoding of small details of the stimulus, at the expense of large-scale features which are already well encoded by the system. This chapter has been submitted for peer review and is currently available as a preprint:

Gabriel Mahuas, Thomas Buffet, Olivier Marre, Ulisse Ferrari, and Thierry Mora. "Strong, but not weak, noise correlations are beneficial for population coding". In: *bioRxiv* (2024), pp. 2024–06

Abstract. Neural correlations play a critical role in sensory information coding. They are of two kinds: signal correlations, when neurons have overlapping sensitivities, and noise correlations from network effects and shared noise. It is commonly thought that stimulus and noise correlations should have opposite signs to improve coding. However, experiments from early sensory systems and cortex typically show the opposite effect, with many pairs of neurons showing both types of correlations to be positive and large. Here, we develop a theory of information coding by correlated neurons which resolves this paradox. We show that noise correlations are always beneficial if they are strong enough. Extensive tests on retinal recordings under different visual stimuli confirm our predictions. Finally, using neuronal recordings and modeling, we show that for high dimensional stimuli noise correlation benefits the encoding of fine-grained details of visual stimuli, at the expense of large-scale features, which are already well encoded.

5.1 Introduction

Neurons from sensory systems encode information about incoming stimuli in their collective spiking activity. This activity is noisy: repetitions of the very same stimulus can drive different responses [199, 112, 60, 165, 73, 36]. It has been shown that the noise is shared among neurons and synchronizes them, an effect called *noise correlations*, as opposed to *signal correlations* induced by the stimulus [6, 36, 86, 8, 127]. Noise correlations have been observed since the first synchronous recordings of multiple neurons [133, 109] and at all levels of sensory processing, from the retina [112, 124, 161, 135, 186, 187, 57, 200, 149, 167] to the visual cortex [199, 56, 87, 165, 47, 73, 97] and other brain areas [183, 6, 94, 13, 36, 8, 70]

Strong noise correlations have been measured mostly between nearby neurons with similar stimulus sensitivity [107, 109, 199, 94, 13, 87, 67, 7, 73]. This behaviour is particularly evident in the retina between nearby ganglion cells of the same type [112, 161, 186, 187, 167]. This observation is however surprising, since previously it was thought that these correlations are detrimental to information coding: a theoretical argument [199, 26, 129, 137, 166] suggests that noise correlations are detrimental to information transmission if they have the same sign as signal correlations [6, 8, 127]. This rule is sometimes called the sign rule [75], and is related to the notion of information-limiting correlations [116]. Since nearby neurons with similar tuning are positively correlated by the signal, the theory would predict that their positive noise correlations should be detrimental, making the code less efficient. However, a large body of literature has reported the beneficial effects of noise correlations on coding accuracy [1, 135, 178, 45, 63, 162, 57, 149, 22, 117]. Because of these contradictions, the effect of shared variability on information transmission is still unclear,

and remains a largely debated topic in neuroscience [6, 86, 8].

Here we aim to resolve these tensions by developing a general framework that builds on previous theoretical work [102] and is grounded on the analysis of multi-electrode array recordings of rat and mouse retinas. While previous studies have considered the impact of noise correlations either for particular stimuli [199, 135, 57, 70], or for particular models [166, 178, 45], our approach is general and covers both low and high dimensional stimuli. We show that the sign rule can be broken in a specific regime that we observed in retinal responses: when noise correlations are strong enough compared to signal correlations, they have a beneficial effect on information transmission. Our results unravel the complex interplay between signal and noise correlations, and predict when and how noise correlations are beneficial or detrimental. In the case of high dimensional stimuli, like images or videos, our theory predicts different effects of noise correlations depending on stimulus features. In particular, it explains how large noise correlations between neurons with similar stimulus sensitivity help encode fine details of the stimulus.

We study theoretically the different regimes for pairs of spiking neurons, and illustrate them in the correlated activity of rat retinal ganglion cells. We then extend our analysis to large populations of sensory neurons, and propose a spectral analysis suggesting that local noise correlations enhance information by favoring the accurate encoding of fine-grained details. We validate this last prediction combining data from the mouse retina with accurate convolutional neural network (CNN) models.

5.2 Results

Strong pairwise noise correlations enhance information transmission

We start with a simple model of a pair of spiking neurons encoding an angle θ, for instance the direction of motion of a visual stimulus, in their responses r_1 and r_2. These responses are correlated through two sources: signal correlations ρ_{s} due to an overlap of the tuning curves (Fig. 5.1A); and noise correlations ρ_{n} due to shared noise (see Methods for mathematical definitions). We asked how this shared noise affects the encoded information, for a fixed level of noise in neurons.

To quantify the joint coding capacity of the 2 neurons, we computed the mutual information $I(\theta; r_1, r_2)$ between their activities and the stimulus θ. For fixed tuning curves, we find that the mutual information depends non monotonously on the noise correlation ρ_{n} (Fig. 5.1B). For small abolute values of ρ_{n}, the sign rule is satisfied, meaning that negative

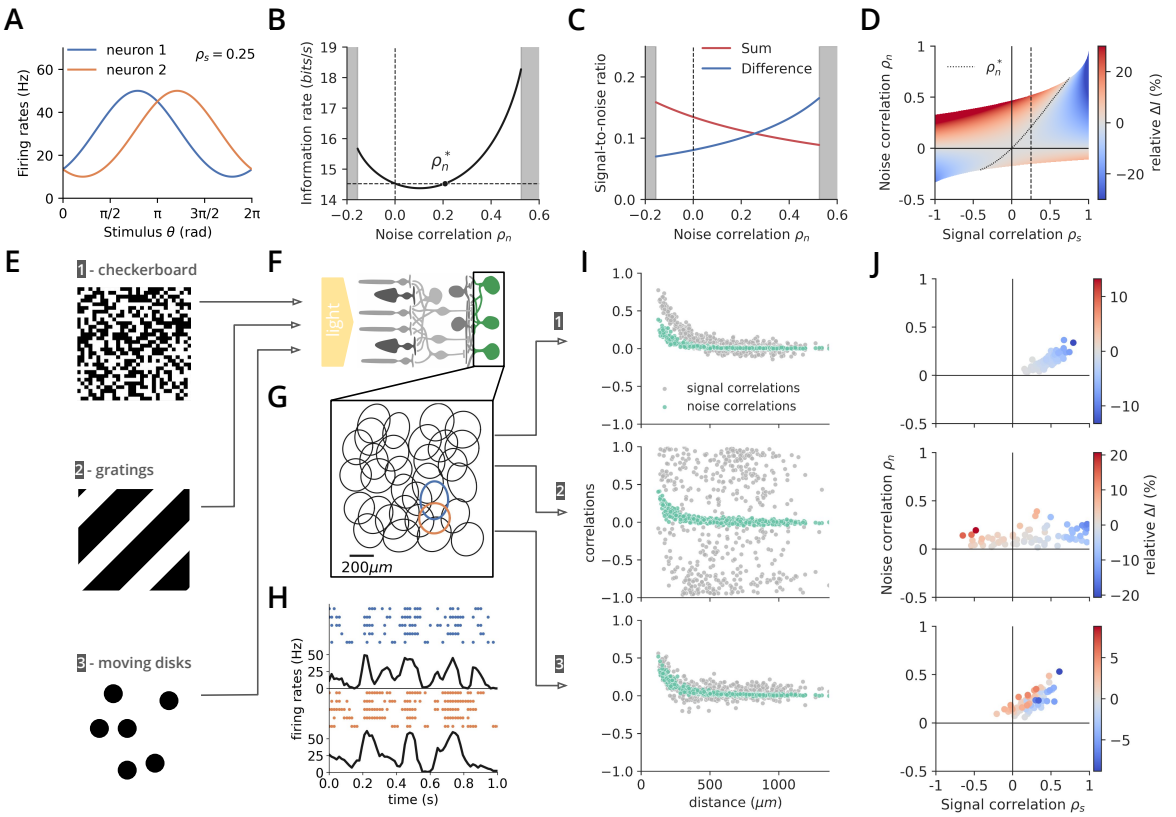

Figure 5.1: **The effects of noise correlations on information coding depends on the stimulus**. **A.** Example pair of Von Mises tuning curves with moderate signal correlation level ($\rho_s = 0.25$). **B.** Mutual information between stimulus and response for the example pair of A, vs the strength of noise correlations. Grey areas correspond to forbidden correlations zones. **C.** The non-monoticity of B may be explained by examining how well the stimulus is represented by the sum and difference of the two neurons' activities, as measured by their signal-to-noise ratios. Noise correlations enhance noise in the sum, but reduce it in the difference. **D.** Heatmap representing the noise synergy, defined as the relative gain of mutual information induced by noise correlations compared to the uncorrelated case. The dotted vertical line corresponds to the example pair of A and B. **E.** Three stimuli with different spatiotemporal statistics were presented to a rat retina. **F.** Retinal ganglion cells (RGCs) were recorded using a multi-electrode array (MEA). **G.** We isolated a nearly complete population of OFF-α cells, with receptive fields (RFs) that tile the visual field following approximately a triangular lattice. **H.** Example raster plots and firing rates for two cells with neighboring RFs. **I.** Signal and noise correlations for each pairs of neurons in the population, versus their distance. Each plot corresponds to 1 of the 3 stimuli of E. **J.** Noise synergy induced by noise correlations for all pairs of nearby neurons ($\geq 300\mu m$), for each stimulus of E.

noise correlations are beneficial, and weak positive ones are detrimental [75, 6, 1, 137]. However, the mutual information increases again and noise correlations become beneficial if they are larger than a certain threshold ρ_{n}^*, violating the sign rule. This non monotonous dependency may be intuitively explained as the interplay between two opposite effects (Fig. 5.1C). Negative noise correlations are beneficial because they reduce noise in the total activity of the neurons. By contrast, positive noise correlations reduce noise in their differential activity, but this effect only dominates when they are strong enough.

We call "noise synergy" the gain in information afforded by noise correlations, $\Delta I = I(\rho_{\mathrm{n}}) - I(\rho_{\mathrm{n}} = 0)$. Fig. 5.1D shows how noise synergy depends on both the noise and signal correlation, where the latter is varied in the model by changing the overlap between the tuning curves. Very generally, and beyond the cases predicted by the sign-rule, noise correlations are beneficial also when they are stronger than the signal correlations. We can gain insight into this behaviour by computing an approximation of the mutual information that is valid for small correlations, following [102] (see Methods). The noise synergy can be expressed as:

$$\Delta I \approx \frac{\alpha}{2}\rho_{\mathrm{n}}(\rho_{\mathrm{n}} - \rho_{\mathrm{n}}^*), \tag{5.1}$$

where $\alpha \leq 1$ is prefactor that grows with the signal-to-noise ratio (SNR) of the neurons. Eq. 5.1 captures the behaviour of Fig. 5.1B, in particular the observation that noise correlations are beneficial if $\rho_{\mathrm{n}}\rho_{\mathrm{s}} < 0$, as the sign rule predicts, or if they are strong enough, $|\rho_n| > |\rho_n^*|$. We can show (see Methods) that the threshold ρ_{n}^* scales with the signal correlation strength ρ_{s}:

$$\rho_n^* = \beta\, \rho_{\mathrm{s}}. \tag{5.2}$$

This result holds in the case of Gaussian neurons (see Methods) and the prefactor $\beta \leq 1$ gets smaller and even approaches 0 as the SNR increases. It is also smaller when these SNR are dissimilar between cells, consistent with previous reports [45]. When the SNRs are weak and similar, we have $\beta \approx 1$. This analysis indicates that noise correlations are beneficial when they are of the same strength as signal correlations, but also that this benefit is enhanced when neurons are reliable.

Our definition of the noise synergy relies on comparing the noise-correlated and uncorrelated cases at fixed noise level (SNR). However, increasing noise correlations at constant SNR decreases the effective variability of the response, as measured by the noise entropy of the joint response of the pair (see Methods). This means that high noise correlations imply a more precise response, which could explain the gain in information. To study this possible confounding factor, we also computed ΔI at equal noise entropy, instead of equal SNR, and found that strong noise correlations are still beneficial, with modified $\rho^* = 2\rho_{\mathrm{s}}/(1+\rho_{\mathrm{s}}^2) \leq 1$

(see Methods).

Benefit of noise correlations in pairs of retinal ganglion cells

The theory predicts that noise correlations may be beneficial when they are of the same sign and magnitude as the signal correlations. To see whether real neurons fall in that physiological regime, we recorded *ex vivo* the joint spiking activity of rat retinal ganglion cells (RGCs, see Methods) [40, 24]. We subjected the same retinal preparation to 3 stimuli with distinct spatio-temporal patterns: a random flickering checkerboard, drifting gratings, and randomly moving disks (Fig. 5.1E). The activity of RGCs was recorded using a multi-electrode array (Fig. 5.1F), and data was processed to assign spikes to each neuron [196]. We identified cells belonging to a nearly complete OFF-α population forming a regular mosaic pattern of their receptive fields (Fig. 5.1G).

Each of the 3 stimulus movies was repeated multiple times (Fig. 5.1H), which allowed us to compute the noise and signal correlation functions ρ_{n} and ρ_{s} (Fig. 5.1I), see Methods. All three stimuli produced similar structures of noise correlations across the network, with positive correlations between cells with nearby receptive fields. This is consistent with the fact that noise correlations are a property of the network, independent of the stimulus [167, 101], and likely come here from gap junctions coupling neighbouring RGCs [25, 186]. In contrast, signal correlations strongly depend on the statistical structure of the presented stimulus, and may be positive or negative, with varying strengths.

To test the predictions of our theory, we computed the mutual information between stimulus and response for all pairs of cells whose receptive fields were closer than $300\,\mu\mathrm{m}$ (Fig. 5.1J). The case of the drifting gratings with fixed orientation offers an illustration of the sign rule. That stimulus induces strong negative signal correlations between many cells, depending on their relative positions relative to the gratings direction. Since noise correlations are positive, they are of opposite sign and therefore beneficial. In the case of the checkerboard stimulus, noise correlations were found to be generally detrimental. This again agrees with the sign rule since they have the same sign as signal correlations, but are too weak to surpass the critical correlation value ρ_n^*. Finally, the case of the moving disks provides an example of the third regime, which violates the sign rule: noise correlations are of the same sign as the signal correlations, but also of comparable magnitude. As a result, many pairs fall above the threshold ρ_n^*, making noise correlations beneficial.

Overall, the 3 stimuli illustrate the 3 possible regimes predicted by the theory when noise correlations are positive: a beneficial effect when signal correlations are negative,

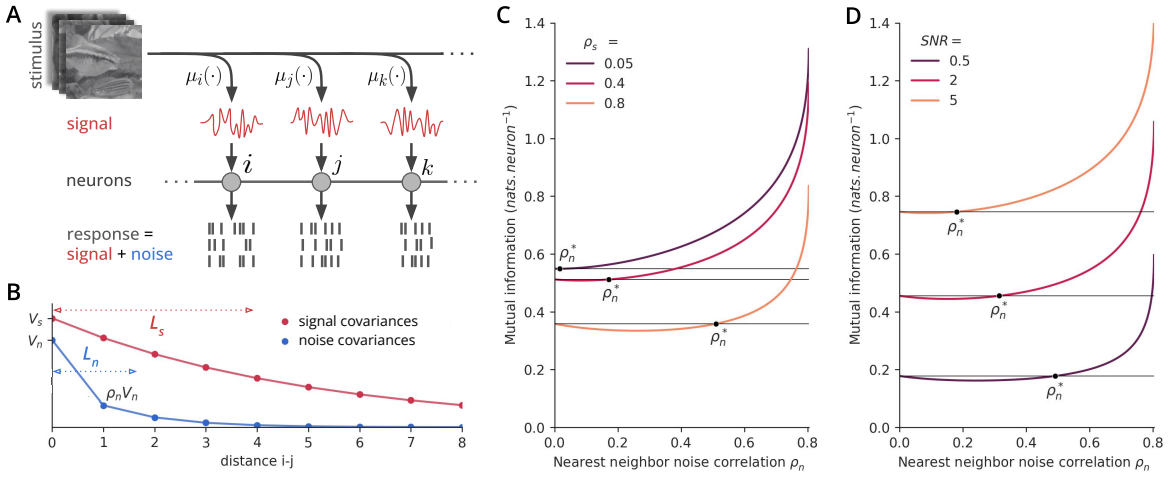

Figure 5.2: **Population analysis**. **A.** Neurons are assumed to be spatially arranged along sensory space. They combine features of the stimulus through a response function μ_i. Noise is added to the neural responses. **B.** Signal and noise covariances versus distance between neurons. Signal and noise covariances decay exponentially with distance with spatial scales L_s and L_n. **C.-D.** Mutual information as a function of the noise correlation between neighbors for: (**C**) varying levels of signal correlations, with fixed $V_\mathrm{s} = 2$, $V_\mathrm{n} = 1$, and $L_\mathrm{n} = 2$; and (**D**) varying levels of signal-to-noise ratio (SNR$= V_\mathrm{s}/V_\mathrm{n}$), with $L_\mathrm{s} = L_\mathrm{n} = 2$ ($\rho_\mathrm{s} \approx 0.6$).

a detrimental effect if signal correlations are positive and noise correlation weaker, and a beneficial effect when noise and signal correlations are both positive and of the same magnitude.

Large sensory populations in high dimension

We then asked how these results extend from pairs to large populations, by considering a large number of neurons tiling sensory space (Fig. 5.2A). To go beyond neurons tuned to a single stimulus dimension, and account for the ability of neurons to respond to different stimuli in a variety of natural contexts, we assume that each neuron responds to high-dimensional stimulus, like a whole image, a temporal sequence, or a movie. As different stimuli are shown, the spike rate of each neuron will vary. For computational ease, we take these fluctuations to be Gaussian of variance V_s.

To account for the empirical observation that nearby neurons tend to have close receptive fields, we correlate the responses of any two neurons with a strength that decreases as a function of their distance in sensory space, with characteristic decay length L_s (Fig. 5.2B). The value of the correlation between nearest neighbours quantifies the signal correlation,

ρ_{s}. For simplicity the response noise is also assumed to be Gaussian of variance V_{n}. To model positive noise correlations between nearby neurons observed in both the retina [112, 161, 186, 187, 167] and cortex [107, 109, 199, 94, 13, 87, 67, 7, 73], we assume that they also decay with distance, but with a different length L_{n} (Fig. 5.2B). The noise correlation between nearest neighbors, defined as ρ_{n}, quantifies their strength.

In this setting, both signal and noise correlations are positive, and the sign rule alone would predict a detrimental effect of noise correlations. The mutual information can be computed analytically in terms of simple linear algebra operations over the neurons' covariance matrices (see Methods) [3]. Using these exact formulas, we examined how the mutual information changes as a function of the noise correlation ρ_{n} for different values of the signal correlation ρ_{s} (Fig. 5.2C) and of the SNR $V_{\mathrm{s}}/V_{\mathrm{n}}$ (Fig. 5.2D).

The results qualitatively agree with the case of pairs of neurons considered previously. Weak noise correlations impede information transmission, in accordance with the sign rule. However, they become beneficial as they increase past a critical threshold (ρ_n^*), and this threshold grows with the signal correlation strength. It also decreases and even vanishes as the SNR is increased (Fig. 5.2D and Methods for a discussion of the large SNR limit). This means that more reliable neurons imply an enhanced benefit of noise correlations. We further proved that, even at low SNR, there always exists a range of noise correlation strengths where noise correlations are beneficial (see Methods). The general dependency of ρ_{n}^* on the correlation ranges L_{s} and L_{n} is shown in Fig. S1.

Based on the analysis of pairs of neurons, we expect inhomogeneities in the SNR $V_{\mathrm{s}}/V_{\mathrm{n}}$ of neurons to enhance the benefit of noise correlations. To study this effect, we let the power of the signal V_{s} vary between cells, while the noise level V_{n} is kept constant. Assuming that each cell is assigned a random value of V_{s}, we can compute the correction to the critical noise correlation ρ_n^*. We find that ρ_n^* decreases at leading order with the magnitude of the inhomogeneity (see Methods). This result confirms that, in large populations of neurons as well, variability among neurons makes it more likely for noise correlations to have a beneficial effect.

Spectral decomposition

Mutual information is a single number that provides a global quantification of coding efficiency, but says nothing about what is being transmitted. Likewise, a positive noise synergy indicates that noise correlations are beneficial overall, but it doesn't tells us what feature of the stimulus are better encoded, nor which specific interactions between signal

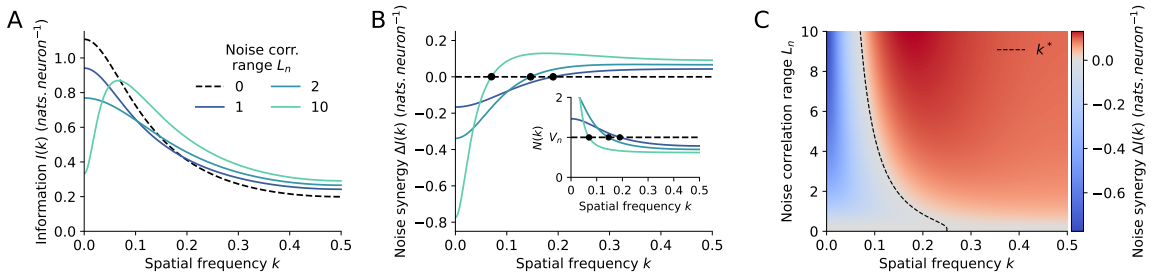

Figure 5.3: **Spectral analysis of stimulus information encoding**. **A.** Spatial spectral decomposition $I(k)$ of the mutual information between stimulus and response for a system with SNR$= 2$, $L_\mathrm{s} = 2$ and $\rho_\mathrm{n} = 0.4$, for various ranges of the noise correlations ($L_\mathrm{n} = 0$ corresponds to the absence of noise correlations). **B.** Spectral decomposition of the noise synergy $\Delta I(k) = \log[(1 + S(k)/N(k))/(1 + S(k)/V_\mathrm{n})]$. The inset shows the power spectrum of the noise. **C.** Heatmap showing the noise synergy spectral decomposition as a function of the noise correlation range L_n. The critical spatial frequency k^* above which noise correlations are beneficial is shown as a black dotted line.

and noise allow for that benefit. We wondered what features of the signal were enhanced by strong positive noise correlations in our population encoding model.

Thanks to the translation-invariant structure of the model, the mutual information and noise synergies may be decomposed spectrally as a sum over spatial frequencies k (expressed in units of inverse distance between nearest neighbors):

$$\Delta I = \frac{n}{2} \int_{-1/2}^{1/2} \mathrm{d}k \, \log \left(\frac{1 + S(k)/N(k)}{1 + S(k)/V_\mathrm{n}} \right), \tag{5.3}$$

where $S(k)$ is the power spectrum of the stimulus, and $N(k)$ that of the noise (see Methods), and $n \to \infty$ is the total number of neurons. In this decomposition, low frequencies correspond to long-range collective modes, while high frequencies correspond to fine-grain features.

Natural stimuli involve spatially extended features impacting many neurons. This causes neural responses to exhibit strong long-range signal correlations between neurons, corresponding in our model to large L_s (Fig. 5.2B). Most information is then carried by low frequency modes of the response (Fig. 5.3A).

Noise correlations concentrate noise power at low frequencies and decrease noise power at high frequencies for a fixed noise level V_n (inset of Fig. 5.3B). As a result, noise correlations enhance information in the high frequency modes of the signal ($k \geq k^*$), at the expense of the low frequencies features (Fig. 5.3B), which are already well represented. Fig. 5.3C shows the spectral decomposition of the noise synergy as a function of the noise correlation range

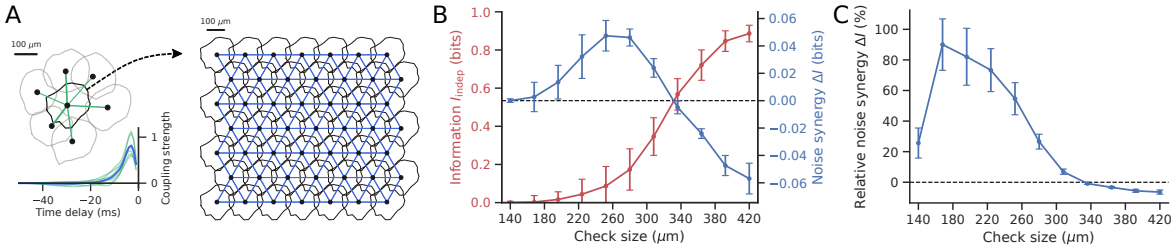

Figure 5.4: **Noise correlations benefit small scale features at the detriment of large scale ones**. **A.** We built a large population of 49 RGCs based on 7 neurons recorded from the mouse retina. A deep-GLM [101] was fit to the experimental population and its central neuron model was tilled on a triangular lattice to create a large RGC network. Couplings between the central experimental cell and its neighbors were symmetrized (green links in the population plot; green lines in the inset plot) and averaged to obtain the coupling filters between nearest neighbors in the synthetic population (blue links in the synthetic mosaic; blue line in the inset plot). **B.** Information in the absence of noise correlations I_{indep} and noise synergy ΔI per pixel for stimulus features of increasing scales. These quantities were computed via a decoding approach applied to a binary flashed checkerboard stimuli with various check sizes. Error bars are the standard error obtained by repeating the analysis on bootstrapped data. In the absence of noise correlations, little information is transmitted about small stimulus features. By contrast, large scale features are well encoded and information per pixel saturates towards 1 bit as check size grows. The noise synergy is positive for small and intermediate check sizes while negative for larger checks, in line with the theoretical results highlighted in Fig. 5.3. **C.** Noise correlations nearly double the amount of information encoded about stimulus features of small and intermediate sizes, while only decreasing information for the largest checks by less than 10%.

L_{n}. The critical frequency $k^* = (1/2\pi)\arccos(e^{-1/L_{\mathrm{n}}})$ above which noise correlations are beneficial only depends on L_{n} (Fig. 5.3C). However, the relative information gains in each frequency domain depends on the strengths of the signal and noise correlations.

In summary, noise correlations enhance fine details of the stimulus to the detriment of its broad features, which are already sufficiently well encoded. This redistribution of the noise across the spectrum drives the gain in information. This effect is generic to any choice of the correlation lengths, and we expect it to hold for other forms of the power spectra and receptive field geometries.

Noise correlations in the retina favor the encoding of fine stimulus details

To test our predictions, we studied experimentally the impact of noise correlations on the encoding of features at different spatial scales in the retina. We recorded *ex vivo* the spiking

activity of 7 OFF-α retinal ganglion cells from a mouse retina using the same experimental technique as described before. We presented the retina with a multi-scale checkerboard stimulus composed of frames made of random black and white checkers, flashed at 4 Hz. Each frame was made of a checkerboard with a given spatial resolution (checks of sizes 12, 24, 36, 72 and 108 μm). From the recorded activity, we infered a deep generalized linear model [101] and used the inferred model to build a large synthetic population of 49 cells organized on a triangular lattice (Fig. 5.4A). We then generated a large dataset of repeated responses to regular black and white checker flashes. Each checker was composed of checks of a given size (sizes ranging from 140 to 420 μm, with 28 μm increments) and for each check size, 50 spatially offset versions of the checker were showed. We trained a linear decoder of each pixel value (black or white) on this synthetic dataset, and a second decoder on the synthetic data in which the activity of each cell was shuffled across repetitions to destroy noise correlations (see Methods).

The two decoders were then applied to the testing datasets, synthetically generated in the same way as the training sets, to decode each pixel from the response. For a fair comparison, the second decoder was applied to data in which noise correlations were removed by shuffling, as in the training. The mutual information carried by the decoders was then estimated separately for each checker size. To limit border effects, the mutual information was estimated for each pixel within a small hexagon centered on the central cell of the synthetic population, of size (distance between opposite sides of the hexagon) equal to the distance between cells.

We found that the gain in mutual information afforded by noise correlations is large and positive for small and intermediate check sizes, while moderately negative for large checks (Fig. 5.4B and C). These results suggest that noise correlation benefit the encoding of small-scale features of the stimulus, at the expense of the large-scale ones, which are easier to encode. Noise correlations can therefore trade the encoding power of large-scale features to improve sensitivity to the small-scale ones.

5.3 Discussion

Many experimental works have shown that neurons with the strongest positive noise correlations are similarly tuned to the stimulus [107, 109, 199, 94, 13, 87, 67, 7, 73]. Here the sign rule [6, 75, 8] would predict a detrimental effect of shared variability, at odds with the efficient coding hypothesis [16], which is supported by a large body of work showing that noise correlations are indeed beneficial [1, 135, 45, 162, 57, 149]. Our work resolves

this inconsistency by showing that beyond a critical value ρ_{n}^*, noise correlations become beneficial to information encoding regardless of their sign. We experimentally demonstrated this effect in recordings of retinal neurons subject to stimuli with different statistics, and showed that it generalizes to large populations of sensory neurons.

Pairwise correlations build up to strong network effects for large populations [153]. This large scale synchronization should be detrimental for coding because it impedes denoising by pooling the signal of multiple neurons [199, 166]: the information gain saturates compared to a population of independent neurons. In contrast, other studies focusing on the stimulus response of large sensory populations have observed a positive gain [135, 45, 63, 57, 149]. Our study proposes a solution to this dispute: when the neural population encodes a low dimensional stimulus, as the angle of a drifting gratings, similarly tuned nearby neurons become strongly signal-correlated, and their noise correlations are detrimental [166]. In the case of high dimensional stimuli, like naturalistic images or videos, signal correlations between them are positive but weak, so that noise correlations become larger than the threshold ρ_n^* and therefore beneficial. We analyzed the impact of shared variability depending of the stimulus spatial frequency: large scale (low dimensional) modes give rise to strong signal correlations, making positive noise correlations detrimental, while small scale (high dimensional) modes benefit from positive signal correlations since their signal correlations are small.

Previous theoretical work assessed the potential benefits of noise correlations violating the sign-rule [45, 75], and studied the interplay of noise and signal correlations in special cases with specific correlation structures [166, 7, 57, 200]. Previous decompositions of the mutual information [128, 137] suggested that variations of the noise correlations with the stimulus may be beneficial, with additionnal information encoded in these variations. However, these results relied on a non-standard definition of noise correlations, making a direct comparison to our results intricate (see appendix D in [102]). Nonetheless, we considered the impact of such fluctuations within our framework, by relaxing the assumption of constant noise correlations in the second order derivation of the noise synergy (see Methods). The computation shows that these fluctuations can improve the noise synergy in two ways: by being large, and by being synchronized to the noise level $V_{\mathrm{n}}(\theta)$, also assumed to be stimlulus dependent. Our results thus extend and clarify previous theoretical work under a common information-theoretic framework.

Our work also shares some similarity with Ref. [178], where the authors predicted the optimal patterns of noise correlations maximizing information transmission by a population of neurons. They showed that at high SNR, optimal noise correlations follow the sign rule. This result does not rule out that high levels of noise correlations violating the sign

rule could be beneficial—albeit not optimal—in agreement with our theory. However, a direct comparison with our results is difficult because in Ref. [178] noise correlations were tuned through inter-neuron couplings that affect the mean response of each neuron to the stimulus, which is kept constant in our analysis. In fact, this effect leads to optimal noise correlations of the same sign as stimulus correlations in the low SNR regime [178]. It was also shown to improve positional coding in the hippocampus through the sharpening of stimulus tuning [117]. This apparent violation of the sign rule is however indirect and distinct from the direct beneficial effect of strong noise correlations that we discuss in this work.

Several studies have focused on the effect of noise correlations on the Fisher information [166, 45, 57, 1, 165]. While our main results are based on the mutual information, they equivalently apply to the Fisher information in the Gaussian case [26] (see Supplementary Appendix). To further test the robustness of our conclusions, we demonstrated that our results are model independent, and hold both for binary and Gaussian neurons. In addition, empirical results from the retinal recordings (Fig. 5.1J) were obtained without any approximation or model choice, and agree with the theory.

Also building on the Fisher information, another line of work [116, 81] suggested that noise correlations are detrimental when aligned to the signal direction in each point of response space. The structure of this type of noise correlations, called "differential" or "information-limiting" correlations, can be intuited from the definition of the Fisher information [116]. Although an in-depth discussion is beyond the scope of this paper, we have performed an additional numerical analysis (see Fig. S2) to demonstrate that information-limiting correlations become increasingly beneficial to the mutual information as their strength increases, while they are always detrimental to the Fisher information.

We validated our theoretical predictions experimentally on recordings of neurons from the retina. Applying our approach to data in sensory cortical areas where similar noise correlation structures have been observed [73, 165] could lead to new understanding of the role of noise correlations in sensory information processing. Another key open question is what stimulus ensembles most benefit from noise correlations, and where naturalistic stimuli stand in that regard. We have further shown that noise correlations benefit the encoding of high-frequency features of the stimulus, which correspond to fine-grained neural activity patterns. Extending these results to higher cortical areas would require understanding which features from the stimulus drive such activity patterns.

5.4 Methods

Covariance and correlation measures

The average responses of two neurons 1 and 2 are given as function of the stimulus θ by the tuning curves $\mu_1(\theta) = \langle r_1 \rangle_\theta$ and $\mu_2(\theta) = \langle r_2 \rangle_\theta$. Signal correlations are defined as $\rho_\mathrm{s} = \mathrm{Corr}_\theta(\mu_1, \mu_2)$, and noise correlations as: $\rho_\mathrm{n}(\theta) = \mathrm{Corr}(r_1, r_2|\theta)$. The sum of these two coefficients does not have a simple interpretation in terms of total correlation or covariance, but we can also decompose the total correlation coefficient between r_1 and r_2 as $\mathrm{Corr}(r_1, r_2) = r_\mathrm{s} + r_\mathrm{n}$, with $r_\mathrm{s} = \mathrm{Cov}_\theta(\mu_1, \mu_2)/\sqrt{\mathrm{Var}(r_1)\mathrm{Var}(r_2)}$, and $r_\mathrm{n} = \langle \mathrm{Cov}(r_1, r_2|\theta) \rangle_\theta / \sqrt{\mathrm{Var}(r_1)\mathrm{Var}(r_2)}$.

Pairwise analysis

Tuning curves. We consider a pair of neurons encoding an angle θ. The responses of the two neurons, r_1 and r_2, are assumed to be binary (spike or no spike in a 10 ms time window) and correlated. Their average responses $\mu_1(\theta)$ and $\mu_2(\theta)$ are given by Von Mises functions (Fig. 5.1A):

$$\mu_i(\theta) = a \frac{\exp\left(\cos\left(\theta - \theta_\mathrm{c}^i\right)/w\right) - \exp\left(-1/w\right)}{\exp\left(1/w\right) - \exp\left(-1/w\right)} + b. \tag{5.1}$$

Signal correlations between the two neurons can be tuned by varying the distance between the center of the two tuning curves θ_c^1 and θ_c^2. The tuning curve width w was set arbitrarily to 5, the amplitude a to 0.4 and the baseline b to 0.1. The strength of noise correlations is set to a constant of θ, $\rho_\mathrm{n}(\theta) = \rho_\mathrm{n}$.

Small correlation expansion. When noise correlations ρ_n are constant, the noise synergy may be expanded as [102]:

$$\Delta I = -r_\mathrm{s} r_\mathrm{n} + \frac{1}{2}\left({\rho_\mathrm{n}}^2 - {r_\mathrm{n}}^2\right) = \frac{\alpha}{2}\rho_\mathrm{n}\left(\rho_\mathrm{n} - \rho_\mathrm{n}^*\right), \tag{5.2}$$

where the second equality highlights the dependency on ρ_n. The critical ρ_n^* may be written as

$$\rho_\mathrm{n}^* = \beta\, \rho_\mathrm{s}, \quad \text{with} \quad \beta = \frac{2 V_\mathrm{s} V_\mathrm{n}}{{V_\mathrm{tot}}^2 - {V_\mathrm{n}}^2}, \tag{5.3}$$

and the prefactor $\alpha = 1 - V_{\rm n}^2/V_{\rm tot}^2$, with the shorthands $V_{\rm tot} = \sqrt{{\rm Var}(r_1){\rm Var}(r_2)}$, $V_{\rm n} = \langle\sqrt{{\rm Var}(r_1|\theta){\rm Var}(r_2|\theta)}\rangle_\theta$, and $V_{\rm s} = \sqrt{{\rm Var}(\mu_1(\theta)){\rm Var}(\mu_2(\theta))}$ corresponding to measures of total, noise, and signal variances in the two cells.

By Cauchy-Schwartz we have:

$$ {V_{\rm n}}^2 \leq \langle{\rm Var}(r_1|\theta)\rangle_\theta\langle{\rm Var}(r_2|\theta)\rangle_\theta, \tag{5.4} $$

which entails

$$ \beta \lesssim \frac{1}{\cosh\frac{\Delta\ln R}{2} + R/2} \leq 1, \tag{5.5} $$

where $R = \sqrt{R_1 R_2}$ and $\Delta\ln R = \ln(R_1/R_2)$, with $R_i = {\rm Var}(\mu_i)/\langle{\rm Var}(r_i|\theta)\rangle_\theta$ the signal-to-noise ratio of the cells. R measures the overall strength of signal-to-noise ratios, while x measures their dissimilarity. The last inequality implies that noise correlations are always beneficial for $\rho_{\rm n} > \rho_{\rm s}$.

Varying noise correlations. When $\rho_{\rm n}(\theta)$ depends on θ, the noise synergy becomes [102]:

$$ \Delta I = \Delta I_{\rm c} + \Delta I_{\rm f,1} + \Delta I_{\rm f,2}, \tag{5.6} $$

where ΔI_c is given by Eq. 5.2, and $\Delta I_{\rm f,1} = (1/2){\rm Var}_\theta(\rho_{\rm n}(\theta)) \geq 0$ accounts for the effect of fluctuations of $\rho_{\rm n}(\theta)$. $\Delta I_{\rm f,2}$ is given by:

$$ \begin{aligned} \Delta I_{\rm f,2} = -\left\langle (\rho_{\rm n}(\theta) - \bar\rho_{\rm n})\,\frac{V_{\rm n}(\theta)}{V_{\rm tot}}\right\rangle_\theta \times \\ \left(\frac{1}{2}\left\langle (\rho_{\rm n}(\theta) + \bar\rho_{\rm n})\,\frac{V_{\rm n}(\theta)}{V_{\rm tot}}\right\rangle_\theta + r_{\rm s}\right), \end{aligned} \tag{5.7} $$

where $\bar\rho_{\rm n} = \langle\rho_{\rm n}(\theta)\rangle_\theta$. This contribution can be positive or negative, depending on how noise correlations $\rho_{\rm n}(\theta)$ co-vary with the noise variance of the pair $V_{\rm n}(\theta)$.

Gaussian case. To test the theory's robustness to modeling choices, we also considered a continuous rather than binary neural response: $r_i = \mu_i(\theta) + \delta r_i$, where both μ_i and δr_i are Gaussian variables defined by their covariance matrices $\Sigma_{{\rm s},ij} = {\rm Cov}_\theta(\mu_i, \mu_j)$, and $\Sigma_{{\rm n},ij} = \langle{\rm Cov}(r_i, r_j|\theta)\rangle_\theta$. The noise synergy can be calculated through classic formulas for the entropy for Gaussian variables, yielding:

$$ \Delta I = \frac{1}{2}\log\left(\frac{|\Sigma_{\rm s} + \Sigma_{\rm n}||V_{\rm n}|}{|\Sigma_{\rm s} + V_{\rm n}||\Sigma_{\rm n}|}\right), \tag{5.8} $$

where $|X|$ denotes the determinant of matrix X, and where $V_{\rm n}$ is the diagonal matrix containing the noise variances of the cells $V_{{\rm n},ii} = \Sigma_{{\rm n},ii}$. Note that this formula is general for an arbitrary number of correlated neurons. In the pairwise case considered here matrices are of size 2×2. The condition for beneficial noise correlations $\Delta I \geq 0$ is satisfied for $\rho_{\rm n} \geq \rho^*_{\rm n}$, with

$$\rho^*_{\rm n} = \beta \rho_{\rm s} \quad \text{with } \beta = \frac{1}{\cosh \frac{\Delta \ln R}{2} + (1-\rho_s^2)R/2} \leq 1, \tag{5.9}$$

which has a similar form as Eq. 5.5.

Noise synergy at constant noise entropy. Increasing noise correlations at constant $V_{\rm n}$ decreases the effective variability of the response, as measured by the noise entropy, $H(\{r_1, r_2\}|\theta) = \ln(2\pi e|\Sigma_{\rm n}|^{1/2})$, with $|\Sigma_{\rm n}| = V_{\rm n}^2(1-\rho_{\rm n}^2)$ in the case two neurons with the same noise level. To correct for this effect we also computed ΔI at constant noise entropy, by rescaling the noise variances in the correlated and uncorrelated cases, $V_{\rm n,c}$ and $V_{\rm n,u}$, so that their resulting noise entropies are equal $|\Sigma_{\rm n}| = V_{\rm n,c}^2(1-\rho_{\rm n}^2) = V_{\rm n,u}^2$.

The critical noise correlation at which $\Delta I \geq 0$ is then given by:

$$\rho^*_{\rm n} = 2\frac{\rho_{\rm s}}{1+\rho_{\rm s}^2} \leq 1, \tag{5.10}$$

where the last inequality implies that strong enough noise correlations are always beneficial.

Retinal data. Retinal data were recorded ex-vivo from a rat retina using a microelectrode array [40] and sorted using SpyKING CIRCUS [196] to isolate single neuron spike trains. From the ensemble of single cells we could isolate a population of 32 OFF-α ganglion cells. Three stimuli movie with different spatio-temporal statistics were presented to the retina: a checkboard movie consisting of black and white checks changing color randomly at 40 Hz and repeated 79 times; a drifting grating movie consisting of black and white stripes of width $333\,\mu$m moving in a fixed direction relatively to the retina, at speed 1 mm/s, and repeated 120 times; and finally a movie composed of 10 black disks jittering according to a Brownian motion on a white background, repeated 54 times.

Gaussian population and spectral analysis

We consider a population of n neurons organized along a 1D lattice with constant interneuron spacing. Their mean response and noise are assumed to be Gaussian, with their noise and signal covariances given by an exponentially decaying function of their pairwise

distances:

$$\Sigma_{\mathrm{s},ij} = V_\mathrm{s} e^{-|i-j|/L_\mathrm{s}}, \tag{5.11}$$

$$\Sigma_{\mathrm{n},ij} = V_\mathrm{n}(\delta_{ij} + \rho_\mathrm{n}^0 e^{-|i-j|/L_\mathrm{n}}). \tag{5.12}$$

V_s and V_n are the signal and noise variance of the single cells. The parameter ρ_n^0 sets the strength of noise correlations such that nearest neighbors have noise correlation $\rho_\mathrm{n} \equiv \rho_\mathrm{n}^0 \exp(-1/L_\mathrm{n})$. When n is large and boundary effects can be ignored, the system is invariant by translation and we can diagonalize Σ_s and Σ_n in the Fourrier basis $\nu_{k,l} = \frac{1}{\sqrt{n}} \exp(-i2\pi kl/n)$. Denoting the spectra of Σ_s and Σ_n by $S(l/n)$ and $N(l/n)$, the expression of the noise synergy, Eq. 5.8, can then be written as a sum over modes:

$$\Delta I = \frac{1}{2} \sum_{l=-\frac{(n-1)}{2}}^{\frac{(n-1)}{2}} \log\left(\frac{1 + S(l/n)/N(l/n)}{1 + S(l/n)/V_\mathrm{n}}\right), \tag{5.13}$$

which simplifies in the $n \to \infty$ limit to:

$$\frac{\Delta I}{n} = \frac{1}{2} \int_{-1/2}^{1/2} \log\left(\frac{1 + S(k)/N(k)}{1 + S(k)/V_\mathrm{n}}\right) \mathrm{d}k, \tag{5.14}$$

with

$$S(k) = V_\mathrm{s} \frac{1 - {\rho_\mathrm{s}}^2}{1 - 2\rho_\mathrm{s}\cos(2\pi k) + {\rho_\mathrm{s}}^2}, \tag{5.15}$$

$$N(k) = V_\mathrm{n}\left(1 - \rho_\mathrm{n}^0 + \rho_\mathrm{n}^0 \frac{1 - {\lambda_\mathrm{n}}^2}{1 - 2\lambda_\mathrm{n}\cos(2\pi k) + {\lambda_\mathrm{n}}^2}\right), \tag{5.16}$$

where $\rho_\mathrm{s} = \exp(-1/L_\mathrm{s})$ is the nearest-neighbors signal correlation, and $\lambda_\mathrm{n} = \exp(-1/L_\mathrm{n})$. k is a wave vector interpretable as a spatial frequency in units of the system's size, up to a 2π factor. Examining Eq. 5.14, we see that noise correlations are beneficial for frequencies for which $N(k) \leq V_\mathrm{n}$, which happens for $k \geq k^*$ where $k^* = (1/2\pi)\arccos(e^{-1/L_\mathrm{n}})$.

In the low noise regime, $R = V_\mathrm{s}/V_\mathrm{n} \gg 1$, the noise synergy reduces to:

$$\frac{\Delta I}{n} \approx -\frac{1}{2} \int_{-1/2}^{1/2} \log(N(k)/V_\mathrm{n}) dk \geq 0, \tag{5.17}$$

where the inequality stems from Jensen's inequality, because $-\log$ is a convex function, and $-\log(\int_{-1/2}^{1/2} dk N(k)/V_\mathrm{n}) = 0$. Therefore in that regime noise correlations are always beneficial.

In the high noise limit, $R \ll 1$, the noise synergy becomes:

$$\frac{\Delta I}{n} \approx \frac{1}{2}\int_{-1/2}^{1/2}\left[\frac{S(k)}{N(k)} - \frac{S(k)}{V_{\rm n}}\right]dk. \tag{5.18}$$

Computing this integral gives the critical noise correlation:

$$\rho_{\rm n}^* = \rho_{\rm s}\frac{(1-\lambda_{\rm n}^2)}{1-2\lambda_{\rm n}\rho_{\rm s}+{\rho_{\rm s}}^2} \leq \rho_{\rm n}^{\max}, \tag{5.19}$$

where $\rho_{\rm n}^{\max} = (1+\lambda_{\rm n})/2$ is the maximum possible value of $\rho_{\rm n}$ (ensuring that the noise spectrum $N(k)$ is non-negative for all k). The last inequality in Eq. 5.19 implies that there always exists a regime in which strong noise correlations are beneficial.

Non-identical neurons. To study the effect of nonhomogeneities among neurons, we considered the case where the signal variance of each cell is different, and drawn at random as $\sqrt{V_{\rm s}^i} = \mu + \eta_i$, where η_i is normally distributed with zero mean and variance ν^2. The noise synergy can be rewritten in the high noise regime ($R \ll 1$) as:

$$\Delta I \approx \frac{1}{2}{\rm Tr}\left(\Sigma_{\rm s}\Sigma_{\rm n}^{-1} - \Sigma_{\rm s}V_{\rm n}^{-1}\right). \tag{5.20}$$

Averaging this expression over η_i yields:

$$\Delta I \approx \Delta I_{\rm u} + \frac{1}{2}\frac{\nu^2}{(\mu^2+\nu^2)}{\rm Tr}\big(\Sigma_{\rm n}^{-1}\left(\bar{R}\mathbb{I} - \Sigma_{\rm s}\right)\big), \tag{5.21}$$

where $\Delta I_{\rm u}$ is the noise synergy in a uniform population (with $V_{\rm s} = \mu^2+\nu^2$), and where the second term is always positive, with $\bar{R} = \langle V_{\rm s}^i\rangle/V_{\rm n} = (\mu^2+\nu^2)/V_{\rm n}$.

Taking the continuous limit ($n \to \infty$) in Eq. 5.21, similarly to the integral limit of Eq. 5.14, allows us to write the critical noise correlation $\rho_{\rm n}^*$ as:

$$\rho_{\rm n}^* = \frac{\rho_{\rm n}^{*,{\rm u}}}{1+\gamma} + \frac{1-\rho_{\rm n}^{*,{\rm u}}}{2}\left(1 - \sqrt{1 + \frac{4\gamma\rho_{\rm s}^2(1-\lambda_{\rm n}^2)}{(1+\gamma)^2(1-\lambda_{\rm n}\rho_{\rm s})^2}}\right), \tag{5.22}$$

where $\gamma = \nu^2/\mu^2$ quantifies the relative magnitude of nonhomogeneities, and $\rho_{\rm n}^{*,{\rm u}}$ is the critical noise correlation value in a uniform population (Eq. 5.19). This modified critical noise correlation value is always smaller than in the uniform case, and scales linearly at leading order with the inhomogeneity parameter γ:

$$\rho_{\rm n}^* = \rho_{\rm n}^{*,{\rm u}}\left(1 - \gamma\frac{1-\rho_{\rm s}^2}{1-\rho_{\rm s}\lambda_{\rm n}}\right) + o(\gamma). \tag{5.23}$$

Decoding analysis

Experimental and synthetic data. We presented a mouse retina with a stimulus consisting of a black and white random checkerboard flashed at 4Hz, each frame with a given spatial resolution (checks of sizes 12, 24, 36, 72 and 108 μm). Retinal ganglion cell activity was recorded ex-vivo using a micro-electrode array and single neuron activity isolated via spike sorting using SpyKING CIRCUS [196]. We isolated a population of $N_{\text{cells}} = 7$ OFF-α retinal ganglion cells which presents strong noise correlations in their response [187]. The original recording contained a $15\,s$ checkerboard movie repeated 90 times as well as 90 different $22.5\,s$ long unrepeated movies.

We inferred a deep Generalized Linear Model (GLM) of the central cell among 7 from the experimental population (Fig. 5.4A), consisting of a stimulus-processing filter, and filters for the spiking history of the cell as well as its 6 neighbors (couplings). The stimulus-processing part of the model consisted of a deep neural network composed of two spatio-temporal convolutional layers followed by a readout layer. The whole model was fit to the data using the 2-step inference approach [101].

A synthetic population of 49 OFF-α ganglion cells was then constructed by arranging them on a triangular lattice of 7 by 7 points. Each cell responds according to the inferred GLM with translated receptive fields. Nearest neighbors were coupled with the average of the GLM couplings inferred between the central cell from the experiment and its neighbors.

To stimulate this synthetic population, we generated a synthetic stimulus ensemble from 550 regular black and white checker frames, each with a given check size ranging from 140 to 420 μm (with increments of 28 μm). Every checker of a given size was presented for 5 different regularly spaced offsets ranging from 0 to 224 μm both in the horizontal and vertical directions, resulting in 25 different frames per size. To further ensure that the color of each pixel in the stimulus ensemble is black or white with equal probability, each checker frame also had its color-reversed version in the set, resulting in 50 different frames for a given check size. A single snippet from the synthetic stimulus ensemble consisted of a 250ms white frame followed by one of the 550 aforementioned checker frames.

We built a training, a vadliation, and testing set for the dependent and independent decoders by simulating the synthetic population for sets of 3750, 1250, and 5000 repetitions (respectively) of each synthetic stimulus snippets.

Decoders. The binary decoders are logistic regressors taking in the integrated response of the population over the $N_\tau = 5$ past bins of 50 ms to predict the ongoing stimulus frame.

The predicted stimulus at time t and repeat k is given by:

$$\hat{X}(x, y, t, k) = f\left(A_{x,y} r(t, k) + \beta_{x,y}\right), \tag{5.24}$$

where x, y are the pixel indices along the two dimensions of the stimulus, $f(x) = (1+e^{-x})^{-1}$ is the sigmoidal function, $A_{x,y}$ is a matrix of size $(N_\tau, N_{\text{cells}})$, $r(t, k)$ is a matrix of size $(N_{\text{cells}}, N_\tau)$ containing the spike history of the population at time t and repeat k, and $\beta_{x,y}$ a pixel-wise bias. Each decoder was trained by minimizing the average binary cross entropy (BCE) between predicted stimulus $\hat{X}(x, y, t, k)$ and the true stimulus $X(x, y, t)$, $\langle \text{BCE}(x, y, t, k) \rangle_{x,y,t,k}$, where

$$\begin{aligned} \text{BCE}(x, y, t, k) = &-X(x, y, t) \ln\left(\hat{X}(x, y, t, k)\right) \\ &-(1 - X(x, y, t)) \ln\left(1 - \hat{X}(x, y, t, k)\right). \end{aligned} \tag{5.25}$$

Training was done by stochastic gradient descent on the synthetic datasets using the training (3750 repetitions) and validation (1250 repetitions) sets. Optimization was done using stochastic gradient descent with momentum, with early stopping when the validation loss did not improve over 6 consecutive epochs. During that procedure the learning rate was divided by 4 whenever the validation loss did not improve for 3 consecutive epochs.

We probed the decoders' abilities to decode features of different spatial scales by decoding the simulated responses of the synthetic population to the checker stimuli with varying check size from the testing set. Performances of the decoders were assessed by computing the mutual information between each pixel's color X and it's decoded value $\hat{X}$, separately for the different sizes of checks. The noise synergy was then computed as the difference between the mutual information averaged over pixels for the dependent and independent decoders.

Error bars were computed as follows. We infered 10 deep GLMs on bootstraps of the original training set, obtained by re-sampling with replacement the simulus-response pairs used for training. These 10 models were used to generate 10 surrogate training sets, from which 10 separate decoders were infered with noise correlations, and another 10 without noise correlations. Then synthetic test sets for the checker decoding task were generated from each of the 10 models, and the performance of each decoder computed separately with and without noise correlations, yielding 10 values of the mutual information, and 10 values of the noise synergy (both averaged over pixels). The error bars are the standard deviations of the resulting information, noise synergy and synergy-to-information ratios (i.e. relative noise synergy) over the 10 bootstraps.

Data availibility

Part of the data utilized in this work have been used or published in previous studies [40, 24]. The remaining data and codes will be shared upon publication of this study. The code used to generate the synthetic data is available at `https://github.com/gmahuas/noisecorr`.

Acknowledgements

We thank Stéphane Deny for help with the experimental data used in the paper and Matthew Chalk and Simone Azeglio for useful discussions. This work was supported by ANR grants n. ANR-22-CE37-0023 “LOCONNECT” and n. ANR-21-CE37-0024 NatNetNoise, and by IHU FOReSIGHT (ANR-18-IAHU-01) and by Sorbonne Center for Artificial Intelligence-Sorbonne University- IDEX SUPER 11-IDEX-0004. This work was also supported by the Bettencourt Schueller Foundation.

5.5 Supplementary information

Strong noise correlation and Fisher information

We consider a pair of neurons encoding an arbitrary variable θ. The responses r of these neurons is assumed to be Gaussian of mean $\mu(\theta)$ and covariance matrix $\Sigma_{\mathrm{n}})$, where $\mu\left(\theta\right)$ are the tuning curves. In this context the Fisher information is defined as:

$$F_{\mathrm{dep}}\left(\theta\right)=\mu'\left(\theta\right)^{\top}\Sigma_{\mathrm{n}}^{-1}\mu'\left(\theta\right). \tag{5.1}$$

Expanding this expression for a pair of neurons with equal noise variance, $\Sigma_{\mathrm{n}}=V_{\mathrm{n}}\begin{pmatrix}1 & \rho_{\mathrm{n}}\\ \rho_{\mathrm{n}} & 1\end{pmatrix}$, yields:

$$F_{\mathrm{dep}}\left(\theta\right)=\frac{\mu_1'\left(\theta\right)^2+\mu_2'\left(\theta\right)^2}{V_{\mathrm{n}}\left(1-\rho_{\mathrm{n}}^2\right)}\left(1-\rho_{\mathrm{n}}\frac{2\mu_1'\left(\theta\right)\mu_2'\left(\theta\right)}{\mu_1'\left(\theta\right)^2+\mu_2'\left(\theta\right)^2}\right). \tag{5.2}$$

In the absence of noise correlations ($\rho_{\mathrm{n}}=0$), the Fisher information simplifies to:

$$F_{\mathrm{indep}}\left(\theta\right)=\frac{\mu_1'\left(\theta\right)^2+\mu_2'\left(\theta\right)^2}{V_{\mathrm{n}}}. \tag{5.3}$$

To quantify the overall Fisher improvement we introduce the quantity $\Delta R=\langle(F_{\mathrm{dep}}(\theta)/F_{\mathrm{indep}}(\theta)-1)\rangle_\theta$. Defining $\xi(\theta)=\frac{2\mu_1'(\theta)\mu_2'(\theta)}{\mu_1'(\theta)^2+\mu_2'(\theta)^2}$, the Fisher improvement becomes:

$$\Delta R=\frac{\rho_{\mathrm{n}}\left(\rho_{\mathrm{n}}-\langle\xi(\theta)\rangle_\theta\right)}{1-{\rho_{\mathrm{n}}}^2}. \tag{5.4}$$

Therefore, strong positive noise correlations will benefit the Fisher information whenever they exceed the critical value $\rho_{\mathrm{n}}^{*,\mathrm{F}}=\langle\xi(\theta)\rangle_\theta$.

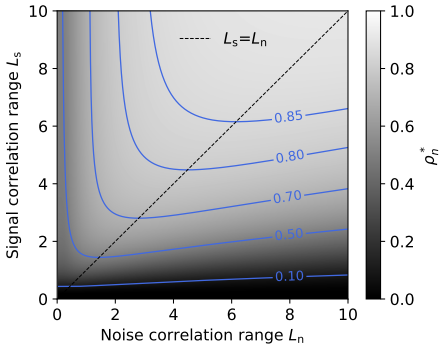

Figure S1: **Behavior of ρ_{n}^*.** ρ_{n}^* changes non-monotically with the signal L_{s} and noise L_{n} correlation ranges, and is concave with respect to these parameters. The maximum value of ρ_{n}^* at a given L_{s} is achieved when $L_{\mathrm{n}} = L_{\mathrm{s}}$.

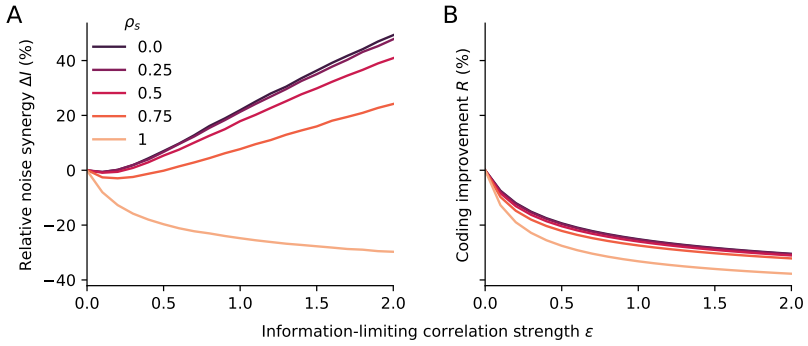

Figure S2: **Impact of information limiting correlations on stimulus information.** An angular stimulus θ is encoded by a pair of neurons characterized by Von Mises tuning curves (with parameters $a = 40$, $b = 10$, and $w = 5$). Their response is Gaussian of means $\mu_1(\theta)$ and $\mu_2(\theta)$. Information-limiting correlations are defined by a covariance of the form: $\Sigma_{\mathrm{n}}(\theta) = V_0\mathbb{I} + \epsilon\mu'(\theta)\mu'(\theta)^\top$, were ϵ controls their strength, and where we set $V_0 = V_{\mathrm{s}}/2$. Note that Σ_{n} now depends on θ. **A.** Relative noise synergy $\Delta I = I_{\mathrm{dep}}/I_{\mathrm{indep}} - 1$ as a function of ϵ, where I_{dep} and I_{indep} quantify the mutual information with and without (off diagonal terms of Σ_{n} set to 0) noise correlations, for different levels of signal correlation ρ_{s}. Mutual information was computed via Monte-Carlo integration. Information-limiting noise correlations become beneficial to the mutual information if they are strong enough, except when cells are perfectly signal-correlated. **B.** By contrast, information-limiting correlations are always detrimental to the Fisher information $F(\theta) = \mu'(\theta)^\top \Sigma_{\mathrm{n}}^{-1}(\theta)\mu'(\theta)$. The relative Fisher improvement $\Delta R = \langle (F_{\mathrm{dep}}(\theta)/F_{\mathrm{indep}}(\theta) - 1)\rangle_\theta$, where $F_{\mathrm{dep}}(\theta)$ and $F_{\mathrm{indep}}(\theta)$ denote the Fisher information with and without noise correlations, is always negative and decreases with ϵ and ρ_s.

Chapter 6

Discussion and perspectives

Sensory neurons encode stimulus information in the collective activity that is shaped both by stimulus input as well as network interactions [25, 163, 51, 135, 64, 187]. These two sources to the collective behavior have very different impacts on neural responses: while the stimulus drives the deterministic part of neural activity, interactions also reshape the noise inherent to neural responses and correlate it across the network. Experimental evidence highlighted how, across a wide variety of sensory systems, noise correlations tend to be positive and strong for pairs of neurons that shared stimulus preferences [112, 161, 187, 167, 107, 94, 13, 87, 73, 7]. By contrast, a wide body of theoretical and computational literature argued that such noise correlations structure should decrease stimulus information encoding compared to the hypothetical case of uncorrelated noise [128, 137, 8, 1, 166, 193, 45, 195].

The main results discussed in this thesis suggest a potential resolution to this discrepancy. Positive noise correlations do not necessarily decrease the information conveyed by the collective response of sensory neurons. In fact, we demonstrated that positive noise correlations can benefit sensory coding even when neurons are positively correlated by stimulus inputs, as long as noise correlations are sufficiently strong. More precisely, we derived a "rule-of-thumb" from the small correlation approximation of the mutual information, that shows how noise correlations will benefit coding when larger than signal correlations. Furthermore, our results suggest that this benefit goes hand in hand with a change in the informative content of neural responses: positive short range noise correlations benefit the encoding of small details of the stimulus, while they decrease the information conveyed about large scale stimulus features.

To reach these conclusions, we developed and employed various computational methods, and used the retina as a model sensory system. These aspects of our approach inevitably involve certain assumptions and limitations. In this section, we will discuss these constraints

and explore the potential insights and future directions emerging from our findings.

6.1 Stimulus dependency of noise correlations

In chapter 1, we showed through a new model based approach that the interaction network underlying noise correlations between α cells from the rat retina is robust to strong changes in statistics of the stimulus ensemble. This result implies that noise correlations in the marginal response distribution (which we denoted r_{ij}^{n} in chapter 2) should be minimally affected by changes in the stimulus ensemble. Beyond this result, a point that we didn't investigate in detail in this thesis concerns the stimulus dependence of noise correlations in the response conditioned to the stimulus (denoted $\rho_{ij}^{\mathrm{n}}(S)$ in chapter 2). Indeed, although they are relatively constant when average over sufficiently large durations or stimulus ensembles, noise correlations could still fluctuate significantly with firing rates or with the stimulus over fine time scales.

Investigating further this point could be of prominent interest as findings from the literature show that, beyond their strength and relation to stimulus correlations, the dependency of noise correlations on the stimulus will also impact sensory coding [57, 200, 137]. When deriving the "rule-of-thumb" in chapter 5, we assume for simplicity that pairwise noise correlations in the conditional response are constant. In reality, these correlations can be modulated by firing rates of the cells [38] as well as by the stimulus [180].

Nonetheless, the intuition derived from the "rule-of-thumb" does predict qualitatively the sign-rule violations observed in experimental data, where the mutual information is computed exactly and accounts for stimulus dependent noise correlations (see chapter 5 Fig. 5.1 panel J). This means that the stimulus dependency of noise correlations between OFF-α ganglion cells from the rat retina does not affect significantly their impact on stimulus encoding.

Studying the implications of the detailed stimulus-dependency of noise correlations and understanding whether our findings hold for different ganglion cell types or even different systems remains an open question, that is moreover difficult to tackle. Indeed, to assess accurately their effect on the mutual information, stimulus dependent noise correlations need to be estimated with sufficient precision from experimental data. To do this, the response to many single stimulus trials have to be recorded experimentally, for a wide range of stimulus values. Here, classical models such as the GLM fall out of their use case, as their functional form will impose the structure of stimulus dependent noise correlations, however well they model the stimulus average of noise correlations (i.e. marginal noise

correlations). Approaches such as integrate-and-fire or biologically inspired models [38, 181] could be used to investigate the relationship between firing rates, stimulus correlations and noise correlations depending on the origin of noise correlations (i.e. shared noise or gap junctions). Powerful statistical methods can also be applied to this task. A recent and promising advance to model the stimulus dependency of noise correlations consists in using Wishart processes to reduce drastically the need for experimental data [120].

6.2 Beyond pairwise instantaneous correlations

The goal of this thesis was to investigate the impact of the structure of neural noise on sensory coding in the retina. To keep this problem tractable and obtain interpretable results, we had to make significant simplifications to the most general formulation of this problem. The first major (but common) simplification we did was to discretize neural responses in time and only considered the population activity within one time bin. By doing so, we bias our estimate of information by imposing an arbitrary temporal resolution to the neural code, and we further neglect all (signal and noise) correlations between successive time bins. In this thesis, we mostly focused on studying the coding impact of the fast noise correlations observed between mammalian α retinal ganglion cells. To mitigate the effect of the aforementioned simplification, we binned experimental data at around $15\ ms$ to integrate most noise correlations in single bins, leaving almost no such correlations between successive time bins. As a consequence, we neglect the temporal structure of correlations within one time bin, while it as been shown in different systems including the retina that stimulus information can be carried by spikes with millisecond precision [121, 11]. A potential solution to circumvent this issue and account for the temporal structure of noise correlations has been illustrated in chapter 4, where we applied our approach to compute the information conveyed by subsequent small time bins in single neurons. Future work could directly extend this approach to populations of retinal neurons and their spiking history.

The second important simplification we performed was to reduce our description of collective neural activity in individual time bins to pairwise correlations between neurons. This approximation naturally stemmed from the use of pairwise maximum entropy models and the corresponding small correlation expansion. However, while the pairwise assumption for conditional responses agrees at least intuitively with the fact that nearby ganglion cells are coupled via gap junctions, one could argue that shared noise could induce higher order noise correlations that may not be predictable from pairwise interactions. This is even more striking for marginal distributions, as the strong spatio-temporal correlations present in the

stimulus could drive higher order correlations in neural responses, leading to the breakdown of the approximation [39]. That being said, maximum entropy models have been extensively used in the literature to model both conditional [64, 51, 39] and marginal [153, 176, 58, 54] response distributions. It has been shown, both in the retina and the cortex, that pairwise interactions are usually sufficient for up to a few tens of neurons [125, 176]. This is in agreement with our results discussed in chapter 3, where we validated our approach on non maximum entropy synthetic data. However, accounting for higher order interactions starts becoming important for larger populations [125, 176]. Extensions of our approach could be done by approximating the mutual information from small correlation expansions of maximum entropy models accounting not only for pairwise but also for triplet interactions. It is worth noting however that applying such approaches to experimental data will be hindered by the fact that three point correlations are notoriously hard to estimate from limited data.

6.3 Stimulus specific impact of noise correlations

We have shown that the impact of noise correlations on stimulus encoding depends heavily on the choice of the stimulus ensemble. In particular, we showed that strong positive noise correlations such as those observed in the retina would overall increase the mutual information for stimuli that only moderately correlate neural responses. Our experimental data analyses focused on a set of stimuli that had very different statistics, to illustrate the different effects predicted from our theoretical developments. A point we didn't address in our analyses but which could constitute an immediate next step for future work would be to investigate where natural stimuli stand in that picture.

One could question however, the understanding that such an analysis would provide. Indeed, the mutual information is but an average measure of the uncertainty reduction that one gets about the stimulus value, from observing neural responses. If two different populations, one noise correlated and the other uncorrelated, have the same mutual information value, does this imply that noise correlations have no impact on stimulus encoding and can be overlooked when studying sensory coding? Beyond their effect on the mutual information value, noise correlations could also redistribute the amount of information available in neural responses about each different stimulus composing the ensemble.

In order to gather intuition on this point, one can rely on the Fisher based mutual information approximation that we mentioned in chapter 2 [26, 166, 192]. In the small neural noise regime, the mutual information can be approximated as $I = \alpha\langle\log(J(\theta))\rangle_\theta + \beta$,

where α and β are constant and $J(\theta)$ is the Fisher information about stimulus parameter θ (see Eq. 2.12 for the linear Fisher expression). Coming back to the simplistic case of two neurons encoding a single stimulus parameter θ as illustrated in Fig. 2.1 from chapter 2, one would therefore expect positive noise correlations to benefit stimuli lying in between receptive fields, when direction of the local signal is orthogonal to that of the noise (see discussion on this point in chapter 2 and Fig. 2.2 C).

By contrast, a simple approximation for the mutual information in the high noise regime derived in ref. [166] yields a very different interpretation. Assuming constant noise covariance C^{n} and Gaussian conditional neural responses, noting $\mu(\theta)$ the mean population response vector to stimulus θ and introducing $\delta(\theta) = \mu(\theta) - \langle \mu(\theta) \rangle_\theta$, this approximation gives:

$$I = \frac{1}{2} \langle \delta(\theta)^\top C^{\mathrm{n}-1} \delta(\theta) \rangle_\theta. \tag{6.1}$$

Interestingly, this expression strongly resembles that of the linear Fisher information Eq. 2.12, at the difference that mean response derivatives $\mu'(\theta)$ are now replaced by their "global" counterpart $\delta(\theta) = \mu(\theta) - \langle \mu(\theta) \rangle_\theta$. The same reasoning applied to the Fisher information previously can be adapted to this new setting: noise correlations will be beneficial when they are orthogonal to the signal, which is now quantified by $\delta(\theta)$ instead of $\mu'(\theta)$. For the simple pairwise example re-invoked above, this implies that positive noise correlations will benefit the encoding of stimuli that activate one neuron more than average and the other less than average. By contrast to the previous low noise case where positive noise correlations would benefit stimulus encoding between tuning curves, noise correlations will now benefit the encoding of stimuli that are centered on one of the two receptive fields.

This transition between high and low noise regimes clearly illustrates how understanding the precise role of noise correlations in sensory coding will require going beyond the average information measure that the mutual information is. We note nonetheless that validating and extending the picture sketched here to the case of more complex and realistic stimuli may require using more general and theoretically grounded decompositions of the mutual information, such as the Stimulus Specific Information [27, 28].

6.4 Population coding of specific stimulus features

Retinal ganglion cells activity is often very nonlinear and sensitive to several features of the stimulus [50, 40]. How sensitivity to multiple features extends from single neurons to populations and more precisely how the encoding of different features interact at the population level is unclear. Such parallel feature encoding could give rise to flexible neural

codes, where information about different features is readily available from the collective activity of a given population. Moreover, all features that compose the visual stimulus may not have the same behavioral interest at any given time. To take an example, an animal actively engaging in prey capture may care more about detecting small moving objects in the visual field than other features such as global luminance changes. Understanding how information related to one "relevant" stimulus feature is impacted by the activity induced by other "irrelevant" features in the context of a given task could shed new light on population coding in the retina. In this thesis, we considered that noise in neural responses was entirely due to intrinsic neural variability, while the signal of interest to encode by the system consisted of the whole visual input. In light of the picture outlined above, one could consider an alternate view of the problem. Instead of treating the whole visual input as the signal to encode, one could rather consider that only specific features of the stimulus (like moving objects in the aforementioned example) ought to be treated as signal. As a result, the part of neural activity that is driven by other "irrelevant" features (i.e. global luminance in the example above) should be treated as noise when evaluating information in neural responses. This implies that there are now two noise sources to the collective behavior of sensory neurons: intrinsic noise that we already considered in the rest of the thesis, and extrinsic noise, that is induced by "irrelevant" stimulus features.

From there, one can first wonder about the effect of these total noise correlations on the encoding of the feature considered here as "relevant". An answer to this question can be provided by extending the main results derived in chapters 4 and 5 to this new setup: total noise correlations will benefit stimulus encoding whenever they are larger than or opposed to signal correlations. Interestingly, it has been shown that such extrinsic noise correlations can indeed lead to synergistic coding of specific stimulus features in the retina [91]. Specifically, salamander direction selective ganglion cells are both sensitive to stimulus motion direction and luminance changes, with similar sensitivity in luminance changes for neighboring cells. When considering motion direction as the signal to encode and luminance induced activity as noise, neural correlations increase the information encoded by collective neural responses about stimulus motion.

Another possible direction would be to focus on the impact of intrinsic noise correlations on a specific feature from the stimulus. This could be done by maintaining the distinction between extrinsic and intrinsic noise correlations and computing the noise synergy related to intrinsic neural noise, i.e. the difference between mutual information with and without intrinsic noise correlations, with conserved extrinsic correlations. Our main results could also be extended to this setting, by decomposing total noise covariances in two components $C^{\mathrm{n}}_{\mathrm{tot}}(s) = C^{\mathrm{n}}_{\mathrm{int}}(s) + C^{\mathrm{n}}_{\mathrm{ext}}(s)$ and inserting this decomposition directly in the second order

mutual information approximation derived in chapter 4. This would result in a modified value for the critical noise correlation value discussed in chapters 4 and 5, that would depend on the strength of signal and extrinsic noise correlations, as well as on the ratio between extrinsic and intrinsic noise variances. Applying this approach to experimental data could be challenging however, as it would imply getting repeated trials for each stimuli and each extrinsic noise realisation, making the use of generative models to palliate the lack of experimental time almost unavoidable.

6.5 Conclusion

The impact of noise correlations on stimulus encoding has been a longstanding research topic in sensory neuroscience. The main contributions of this thesis attempted to tackle this problem from three complementary angles. First, we developed a new inference approach for Generalized Linear Models, that describes accurately the contributions of noise and stimulus correlations to the collective behavior of rat retinal ganglion cells. This approach allowed us to assess the independence of the interaction network underlying noise correlations with respect to the stimulus ensemble. Further, this approach could be used to investigate the impact of noise correlations in simulated responses to new synthetic stimuli. Second, we developed a small-correlation approximation of the mutual information that can be directly applied to experimental recordings and which provides theoretical insight on the impact of noise correlations on sensory coding. The accuracy of this approximation was validated on synthetic data, and the approach was used to quantify the impact of noise correlations and refractory effects in retinal coding. Lastly, by investigating the impact of noise correlations on coding in experimental data from the retina, we illustrated the predictions of the aforementioned approximation. In particular, we showed that, beyond the "sign-rule", strong and positive noise correlation could still benefit sensory coding as long as they were larger than stimulus correlations. We further demonstrated that such effects could also arise in large neural populations, and that they are concomitant with beneficial effects of noise correlations on the encoding of small scale stimulus features.

Bibliography

[1] Larry F Abbott and Peter Dayan. "The effect of correlated variability on the accuracy of a population code". In: *Neural computation* 11.1 (1999), pp. 91–101.

[2] Petri Ala-Laurila, Martin Greschner, EJ Chichilnisky, and Fred Rieke. "Cone photoreceptor contributions to noise and correlations in the retinal output". In: *Nature neuroscience* 14.10 (2011), pp. 1309–1316.

[3] Joseph J Atick and A Norman Redlich. "Towards a theory of early visual processing". In: *Neural computation* 2.3 (), pp. 308–320.

[4] Joseph J Atick and A Norman Redlich. "What does the retina know about natural scenes?" In: *Neural computation* 4.2 (1992), pp. 196–210.

[5] Fred Attneave. "Some informational aspects of visual perception." In: *Psychological review* 61.3 (1954), p. 183.

[6] Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. "Neural correlations, population coding and computation". In: *Nature reviews neuroscience* 7.5 (), pp. 358–366.

[7] Bruno B Averbeck and Daeyeol Lee. "Effects of noise correlations on information encoding and decoding". In: *Journal of neurophysiology* 95.6 (2006), pp. 3633–3644.

[8] Rava Azeredo da Silveira and Fred Rieke. "The geometry of information coding in correlated neural populations". In: *Annual Review of Neuroscience* 44 (2021), pp. 403–424.

[9] Baktash Babadi, Alexander Casti, Youping Xiao, Ehud Kaplan, and Liam Paninski. "A generalized linear model of the impact of direct and indirect inputs to the lateral geniculate nucleus". In: *Journal of Vision* 10.10 (2010), pp. 22–22.

[10] Tom Baden, Philipp Berens, Katrin Franke, Miroslav Roman Roson, Matthias Bethge, and Thomas Euler. "The functional diversity of retinal ganglion cells in the mouse". In: *Nature* 529.7586 (2016), pp. 345–350.

[11] Tom Baden, Federico Esposti, Anton Nikolaev, and Leon Lagnado. “Spikes in retinal bipolar cells phase-lock to visual stimuli with millisecond precision”. In: *Current Biology* 21.22 (2011), pp. 1859–1869.

[12] Tom Baden, Thomas Euler, and Philipp Berens. “Understanding the retinal basis of vision across species”. In: *Nature Reviews Neuroscience* 21.1 (2020), pp. 5–20.

[13] Wyeth Bair, Ehud Zohary, and William T Newsome. “Correlated firing in macaque visual area MT: time scales and relationship to behavior”. In: *Journal of Neuroscience* 21.5 (2001), pp. 1676–1697.

[14] HB Barlow and William R Levick. “The mechanism of directionally selective units in rabbit’s retina.” In: *The Journal of physiology* 178.3 (1965), p. 477.

[15] HB Barlow and WR Levick. “Three factors limiting the reliable detection of light by retinal ganglion cells of the cat”. In: *The Journal of Physiology* 200.1 (1969), pp. 1–24.

[16] Horace B Barlow et al. “Possible principles underlying the transformation of sensory messages”. In: *Sensory communication* 1.01 (1961), pp. 217–233.

[17] Horace B Barlow. “Summation and inhibition in the frog’s retina”. In: *The Journal of physiology* 119.1 (1953), p. 69.

[18] Ramon Bartolo, Richard C Saunders, Andrew R Mitz, and Bruno B Averbeck. “Information-limiting correlations in large neural populations”. In: *Journal of Neuroscience* 40.8 (2020), pp. 1668–1678.

[19] Jeffrey Beck, Vikranth R Bejjanki, and Alexandre Pouget. “Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons”. In: *Neural computation* 23.6 (2011), pp. 1484–1502.

[20] Michael Berry and Markus Meister. “Refractoriness and neural precision”. In: *Advances in neural information processing systems* 10 (1997).

[21] Adam Bleckert, Gregory W Schwartz, Maxwell H Turner, Fred Rieke, and Rachel OL Wong. “Visual space is represented by nonmatching topographies of distinct mouse retinal ganglion cell types”. In: *Current Biology* 24.3 (2014), pp. 310–315.

[22] Juan C Boffi, Brice Bathellier, Hiroki Asari, and Robert Prevedel. “Effective sound localization coding by noisy populations of mouse inferior colliculus neurons revealed by fast volumetric imaging”. In: ().

[23] Bart G Borghuis, Peter Sterling, and Robert G Smith. “Loss of sensitivity in an analog neural circuit”. In: *Journal of Neuroscience* 29.10 (2009), pp. 3045–3058.

[24] Vicente Botella-Soler, Stéphane Deny, Georg Martius, Olivier Marre, and Gašper Tkačik. “Nonlinear decoding of a complex movie from the mammalian retina”. In: *PLoS computational biology* 14.5 (), e1006057.

[25] Iman H Brivanlou, David K Warland, and Markus Meister. “Mechanisms of concerted firing among retinal ganglion cells”. In: *Neuron* 20.3 (1998), pp. 527–539.

[26] Nicolas Brunel and Jean-Pierre Nadal. “Mutual information, Fisher information, and population coding”. In: *Neural computation* 10.7 (1998), pp. 1731–1757.

[27] Daniel A Butts. “How much information is associated with a particular stimulus?” In: *Network: Computation in Neural Systems* 14.2 (2003), p. 177.

[28] Daniel A Butts and Mark S Goldman. “Tuning curves, neuronal variability, and sensory coding”. In: *PLoS biology* 4.4 (2006), e92.

[29] Santiago A Cadena, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, and Alexander S Ecker. “Deep convolutional models improve predictions of macaque V1 responses to natural images”. In: *PLoS computational biology* 15.4 (2019), e1006897.

[30] Santiago A Cadena, Fabian H Sinz, Taliah Muhammad, Emmanouil Froudarakis, Erick Cobos, Edgar Y Walker, Jake Reimer, Matthias Bethge, Andreas Tolias, and Alexander S Ecker. “How well do deep neural networks trained on object recognition characterize the mouse visual system?” In: *Real Neurons $\{\backslash\&\}$ Hidden Units: Future directions at the intersection of neuroscience and artificial intelligence@ NeurIPS 2019* (Real Neurons & …). 2019.

[31] EJ Chichilnisky. “A simple white noise analysis of neuronal light responses”. In: *Network: computation in neural systems* 12.2 (2001), p. 199.

[32] François Chollet et al. *Keras*. `https://keras.io`. 2015.

[33] BG Cleland and WR Levick. “Brisk and sluggish concentrically organized ganglion cells in the cat’s retina”. In: *The Journal of Physiology* 240.2 (1974), pp. 421–456.

[34] Simona Cocco and Rémi Monasson. “Adaptive cluster expansion for inferring Boltzmann machines with noisy data”. In: *Physical review letters* 106.9 (), p. 090601.

[35] Simona Cocco and Rémi Monasson. “Adaptive cluster expansion for the inverse Ising problem: convergence, algorithm and tests”. In: *Journal of Statistical Physics* 147 (), pp. 252–314.

[36] Marlene R Cohen and Adam Kohn. “Measuring and interpreting neuronal correlations”. In: *Nature neuroscience* 14.7 (), pp. 811–819.

[37] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press, 2005.

[38] Jaime De La Rocha, Brent Doiron, Eric Shea-Brown, Krešimir Josić, and Alex Reyes. "Correlation between neural spike trains increases with firing rate". In: *Nature* 448.7155 (2007), pp. 802–806.

[39] Geoffroy Delamare and Ulisse Ferrari. "Time-dependent maximum entropy model for populations of retinal ganglion cells". In: *Physical Sciences Forum* (Physical Sciences Forum). Vol. 5. 1. MDPI. 2022, p. 31.

[40] Stephane Deny, Ulisse Ferrari, Emilie Mace, Pierre Yger, Romain Caplette, Serge Picaud, Gašper Tkačik, and Olivier Marre. "Multiplexed computations in retinal ganglion cells of a single type". In: *Nature communications* 8.1 (2017), p. 1964.

[41] Amadeus Dettner, Sabrina Münzberg, and Tatjana Tchumatchenko. "Temporal pairwise spike correlations fully capture single-neuron information". In: *Nature communications* 7.1 (), p. 13805.

[42] Steven H DeVries. "Correlated firing in rabbit retinal ganglion cells". In: *Journal of Neurophysiology* 81.2 (1999), pp. 908–920.

[43] Steven H Devries and Denis A Baylor. "Mosaic arrangement of ganglion cell receptive fields in rabbit retina". In: *Journal of neurophysiology* 78.4 (1997), pp. 2048–2060.

[44] Dawei W Dong and Joseph J Atick. "Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus". In: *Network: Computation in neural systems* 6.2 (1995), p. 159.

[45] Alexander Ecker, Philipp Berens, Andreas Tolias, and Matthias Bethge. "The effect of noise correlations in populations of diversely tuned neurons". In: *Nature Precedings* (2011), pp. 1–1.

[46] Alexander S Ecker, Philipp Berens, R James Cotton, Manivannan Subramaniyan, George H Denfield, Cathryn R Cadwell, Stelios M Smirnakis, Matthias Bethge, and Andreas S Tolias. "State dependence of noise correlations in macaque primary visual cortex". In: *Neuron* 82.1 (2014), pp. 235–248.

[47] Alexander S Ecker, Philipp Berens, Georgios A Keliris, Matthias Bethge, Nikos K Logothetis, and Andreas S Tolias. "Decorrelated neuronal firing in cortical microcircuits". In: *science* 327.5965 (), pp. 584–587.

[48] Jos J Eggermont. "Properties of correlated neural activity clusters in cat auditory cortex resemble those of neural assemblies". In: *Journal of neurophysiology* 96.2 (2006), pp. 746–764.

[49] Christina Enroth-Cugell and John G Robson. "The contrast sensitivity of retinal ganglion cells of the cat". In: *The Journal of physiology* 187.3 (1966), pp. 517–552.

[50] Adrienne L Fairhall, C Andrew Burlingame, Ramesh Narasimhan, Robert A Harris, Jason L Puchalla, and Michael J Berry. "Selectivity for multiple stimulus features in retinal ganglion cells". In: *Journal of neurophysiology* 96.5 (2006), pp. 2724–2738.

[51] Ulisse Ferrari, Stephane Deny, Matthew Chalk, Gašper Tkačik, Olivier Marre, and Thierry Mora. "Separating intrinsic interactions from extrinsic correlations in a network of sensory neurons". In: *Physical Review E* 98.4 (2018), p. 042410.

[52] Ulisse Ferrari, Stephane Deny, Olivier Marre, and Thierry Mora. "A simple model for low variability in neural spike trains". In: *Neural Computation* 30.11 (2018), pp. 3009–3036.

[53] Ulisse Ferrari, Stéphane Deny, Abhishek Sengupta, Romain Caplette, Francesco Trapani, José-Alain Sahel, Deniz Dalkara, Serge Picaud, Jens Duebel, and Olivier Marre. "Towards optogenetic vision restoration with high resolution". In: *PLoS computational biology* 16.7 (2020), e1007857.

[54] Ulisse Ferrari, Tomoyuki Obuchi, and Thierry Mora. "Random versus maximum entropy models of neural population activity". In: *Physical Review E* 95.4 (), p. 042321.

[55] Greg D Field, Alexander Sher, Jeffrey L Gauthier, Martin Greschner, Jonathon Shlens, Alan M Litke, and EJ Chichilnisky. "Spatial properties and functional organization of small bistratified ganglion cells in primate retina". In: *Journal of Neuroscience* 27.48 (2007), pp. 13261–13272.

[56] József Fiser, Chiayu Chiu, and Michael Weliky. "Small modulation of ongoing cortical dynamics by sensory input during natural vision". In: *Nature* 431.7008 (), pp. 573–578.

[57] Felix Franke, Michele Fiscella, Maksim Sevelev, Botond Roska, Andreas Hierlemann, and Rava Azeredo da Silveira. "Structures of neural correlation and how they favor coding". In: *Neuron* 89.2 (2016), pp. 409–422.

[58] Elad Ganmor, Ronen Segev, and Elad Schneidman. "A thesaurus for a neural population code". In: *Elife* 4 (), e06134.

[59] Elad Ganmor, Ronen Segev, and Elad Schneidman. "The architecture of functional interaction networks in the retina". In: *Journal of Neuroscience* 31.8 (), pp. 3044–3054.

[60] Timothy J Gawne and Barry J Richmond. "How independent are the messages carried by adjacent inferior temporal cortical neurons?" In: *Journal of Neuroscience* 13.7 (), pp. 2758–2771.

[61] Felipe Gerhard, Moritz Deger, and Wilson Truccolo. “On the stability and dynamics of stochastic spiking neuron models: Nonlinear Hawkes process and point process GLMs”. In: *PLoS computational biology* 13.2 (2017), e1005390.

[62] Tim Gollisch and Markus Meister. “Eye smarter than scientists believed: neural computations in circuits of the retina”. In: *Neuron* 65.2 (2010), pp. 150–164.

[63] Arnulf BA Graf, Adam Kohn, Mehrdad Jazayeri, and J Anthony Movshon. “Decoding the activity of neuronal populations in macaque primary visual cortex”. In: *Nature neuroscience* 14.2 (), pp. 239–245.

[64] Einat Granot-Atedgi, Gašper Tkačik, Ronen Segev, and Elad Schneidman. “Stimulus-dependent maximum entropy models of neural population codes”. In: *PLoS computational biology* 9.3 (2013), e1002922.

[65] Martin Greschner, Jonathon Shlens, Constantina Bakolitsa, Greg D Field, Jeffrey L Gauthier, Lauren H Jepson, Alexander Sher, Alan M Litke, and EJ Chichilnisky. “Correlated firing among major ganglion cell types in primate retina”. In: *The Journal of physiology* 589.1 (2011), pp. 75–86.

[66] Divyansh Gupta, Wiktor Młynarski, Anton Sumser, Olga Symonova, Jan Svatoň, and Maximilian Joesch. “Panoramic visual statistics shape retina-wide organization of receptive fields”. In: *Nature Neuroscience* 26.4 (2023), pp. 606–614.

[67] Diego A Gutnisky and Valentin Dragoi. “Adaptive coding of visual information in neural populations”. In: *Nature* 452.7184 (), pp. 220–224.

[68] Joshua Hahn, Aboozar Monavarfeshani, Mu Qiao, Allison H Kao, Yvonne Kölsch, Ayush Kumar, Vincent P Kunze, Ashley M Rasys, Rose Richardson, Joseph B Wekselblatt, et al. “Evolution of neuronal cell classes and types in the vertebrate retina”. In: *Nature* 624.7991 (2023), pp. 415–424.

[69] Haldan K Hartline. “The receptive fields of optic nerve fibers”. In: *American Journal of Physiology-Legacy Content* 130.4 (1940), pp. 690–699.

[70] Omer Hazon, Victor H Minces, David P Tomàs, Surya Ganguli, Mark J Schnitzer, and Pablo E Jercog. “Noise correlations in neural ensemble activity limit the accuracy of hippocampal spatial representations”. In: *Nature communications* 13.1 (2022), p. 4276.

[71] Alexander Heitman, Nora Brackbill, Martin Greschner, Alexander Sher, Alan M Litke, and EJ Chichilnisky. “Testing pseudo-linear models of responses to natural scenes in primate retina”. In: *BioRxiv* (2016), p. 045336.

[72] David Hocker and Il Memming Park. "Multistep inference for generalized linear spiking models curbs runaway excitation". In: *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)* (2017 8th International IEEE/EMBS ...). IEEE. 2017, pp. 613–616.

[73] Sonja B Hofer, Ho Ko, Bruno Pichler, Joshua Vogelstein, Hana Ros, Hongkui Zeng, Ed Lein, Nicholas A Lesica, and Thomas D Mrsic-Flogel. "Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex". In: *Nature neuroscience* 14.8 (2011), pp. 1045–1052.

[74] Edward H Hu and Stewart A Bloomfield. "Gap junctional coupling underlies the short-latency spike synchrony of retinal α ganglion cells". In: *Journal of Neuroscience* 23.17 (2003), pp. 6768–6777.

[75] Yu Hu, Joel Zylberberg, and Eric Shea-Brown. "The sign rule and beyond: boundary effects, flexibility, and noise correlations in neural population codes". In: *PLoS computational biology* 10.2 (2014), e1003469.

[76] Chang-Jin Jeon, Enrica Strettoi, and Richard H Masland. "The major cell populations of the mouse retina". In: *Journal of Neuroscience* 18.21 (1998), pp. 8936–8946.

[77] Na Young Jun, Greg D Field, and John Pearson. "Scene statistics and noise determine the relative arrangement of receptive field mosaics". In: *Proceedings of the National Academy of Sciences* 118.39 (2021), e2105115118.

[78] MohammadMehdi Kafashan, Anna W Jaffe, Selmaan N Chettih, Ramon Nogueira, Iñigo Arandia-Romero, Christopher D Harvey, Ruben Moreno-Bote, and Jan Drugowitsch. "Scaling of sensory information in large neural populations shows signatures of information-limiting correlations". In: *Nature communications* 12.1 (2021), p. 473.

[79] Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven Siegelbaum, A James Hudspeth, Sarah Mack, et al. *Principles of neural science*. Vol. 4. McGraw-hill New York, 2000.

[80] Ingmar Kanitscheider, Ruben Coen-Cagli, Adam Kohn, and Alexandre Pouget. "Measuring Fisher information accurately in correlated neural populations". In: *PLoS computational biology* 11.6 (2015), e1004218.

[81] Ingmar Kanitscheider, Ruben Coen-Cagli, and Alexandre Pouget. "Origin of information-limiting noise correlations". In: *Proceedings of the National Academy of Sciences* 112.50 (2015), E6973–E6982.

[82] Hilbert J. Kappen and FDB Rodrıguez. "Efficient learning in Boltzmann machines using linear response theory". In: *Neural Computation* 10.5 (1998), pp. 1137–1156.

[83] Prakash Kara, Pamela Reinagel, and R Clay Reid. “Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons”. In: *Neuron* 27.3 (2000), pp. 635–646.

[84] Dimokratis Karamanlis, Mohammad H Khani, Helene M Schreyer, Sören J Zapp, Matthias Mietsch, and Tim Gollisch. “Natural stimuli drive concerted nonlinear responses in populations of retinal ganglion cells”. In: *bioRxiv* (2023), pp. 2023–01.

[85] Ryota Kobayashi, Shuhei Kurita, Anno Kurth, Katsunori Kitano, Kenji Mizuseki, Markus Diesmann, Barry J Richmond, and Shigeru Shinomoto. “Reconstructing neuronal circuitry from parallel spike trains”. In: *Nature communications* 10.1 (2019), p. 4468.

[86] Adam Kohn, Ruben Coen-Cagli, Ingmar Kanitscheider, and Alexandre Pouget. “Correlations and neuronal population information”. In: *Annual review of neuroscience* 39 (2016), pp. 237–256.

[87] Adam Kohn and Matthew A Smith. “Stimulus dependence of neuronal correlation in primary visual cortex of the macaque”. In: *Journal of Neuroscience* 25.14 (2005), pp. 3661–3673.

[88] Subhodh Kotekal and Jason N MacLean. “Recurrent interactions can explain the variance in single trial responses”. In: *PLoS computational biology* 16.1 (2020), e1007591.

[89] Brenna Krieger, Mu Qiao, David L Rousso, Joshua R Sanes, and Markus Meister. “Four alpha ganglion cell types in mouse retina: Function, structure, and molecular signatures”. In: *PloS one* 12.7 (2017), e0180091.

[90] Stephen W Kuffler. “Discharge patterns and functional organization of mammalian retina”. In: *Journal of neurophysiology* 16.1 (1953), pp. 37–68.

[91] Norma Krystyna Kühn and Tim Gollisch. “Activity correlations between direction-selective retinal ganglion cells synergistically enhance motion decoding from complex visual scenes”. In: *Neuron* 101.5 (2019), pp. 963–976.

[92] Tobias Kühn and Frédéric van Wijland. “Diagrammatics for the inverse problem in spin systems and simple liquids”. In: *Journal of Physics A: Mathematical and Theoretical* 56.11 (), p. 115001.

[93] Peter E Latham and Sheila Nirenberg. “Synergy, redundancy, and independence in population codes, revisited”. In: *Journal of Neuroscience* 25.21 (2005), pp. 5195–5206.

[94] Daeyeol Lee, Nicholas L Port, Wolfgang Kruse, and Apostolos P Georgopoulos. “Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex”. In: *Journal of Neuroscience* 18.3 (1998), pp. 1161–1170.

[95] Jerome Y Lettvin, Humberto R Maturana, Warren S McCulloch, and Walter H Pitts. "What the frog's eye tells the frog's brain". In: *Proceedings of the IRE* 47.11 (1959), pp. 1940–1951.

[96] William R Levick. "Receptive fields and trigger features of ganglion cells in the visual streak of the rabbit's retina". In: *The Journal of physiology* 188.3 (1967), p. 285.

[97] I-Chun Lin, Michael Okun, Matteo Carandini, and Kenneth D Harris. "The nature of shared cortical variability". In: *Neuron* 87.3 (), pp. 644–656.

[98] Dmitry R Lyamzin, Samuel J Barnes, Roberta Donato, Jose A Garcia-Lazaro, Tara Keck, and Nicholas A Lesica. "Nonlinear transfer of signal and noise correlations in cortical networks". In: *Journal of Neuroscience* 35.21 (2015), pp. 8065–8080.

[99] Jakob H Macke, Iain Murray, and Peter Latham. "How biased are maximum entropy models?" In: *Advances in neural information processing systems* 24 ().

[100] Gabriel Mahuas, Thomas Buffet, Olivier Marre, Ulisse Ferrari, and Thierry Mora. "Strong, but not weak, noise correlations are beneficial for population coding". In: *bioRxiv* (2024), pp. 2024–06.

[101] Gabriel Mahuas, Giulio Isacchini, Olivier Marre, Ulisse Ferrari, and Thierry Mora. "A new inference approach for training shallow and deep generalized linear models of noisy interacting neurons". In: *Advances in neural information processing systems* 33 (2020), pp. 5070–5080.

[102] Gabriel Mahuas, Olivier Marre, Thierry Mora, and Ulisse Ferrari. "Small-correlation expansion to quantify information in noisy sensory systems". In: *Physical Review E* 108.2 (2023), p. 024406.

[103] O Marre, D Amodei, N Deshmukh, K Sadeghi, F Soo, T Holy, MJ Berry, et al. "Recording of a large and complete population in the retina". In: *Journal of Neuroscience* 32.43 (2012), p. 1485973.

[104] Olivier Marre, Dario Amodei, Nikhil Deshmukh, Kolia Sadeghi, Frederick Soo, Timothy E Holy, and Michael J Berry. "Mapping a complete neural population in the retina". In: *Journal of Neuroscience* 32.43 (2012), pp. 14859–14873.

[105] Olivier Marre, Sami El Boustani, Yves Fregnac, and Alain Destexhe. "Prediction of spatiotemporal patterns of neural activity from pairwise correlations". In: *Physical review letters* 102.13 (2009), p. 138101.

[106] Richard H Masland. "The neuronal organization of the retina". In: *Neuron* 76.2 (2012), pp. 266–280.

[107] DAVID N Mastronarde. “Correlated firing of cat retinal ganglion cells. I. Spontaneously active inputs to X-and Y-cells”. In: *Journal of Neurophysiology* 49.2 (1983), pp. 303–324.

[108] DAVID N Mastronarde. “Interactions between ganglion cells in cat retina”. In: *Journal of Neurophysiology* 49.2 (1983), pp. 350–365.

[109] David N Mastronarde. “Correlated firing of retinal ganglion cells”. In: *Trends in neurosciences* 12.2 (), pp. 75–80.

[110] Lane McIntosh, Niru Maheswaranathan, Aran Nayebi, Surya Ganguli, and Stephen Baccus. “Deep learning models of the retinal response to natural scenes”. In: *Advances in neural information processing systems* 29 (2016).

[111] Markus Meister and Michael J Berry. “The neural code of the retina”. In: *neuron* 22.3 (1999), pp. 435–450.

[112] Markus Meister, Leon Lagnado, and Denis A Baylor. “Concerted signaling by retinal ganglion cells”. In: *Science* 270.5239 (1995), pp. 1207–1210.

[113] Leenoy Meshulam, Jeffrey L Gauthier, Carlos D Brody, David W Tank, and William Bialek. “Collective behavior of place and non-place neurons in the hippocampal network”. In: *Neuron* 96.5 (2017), pp. 1178–1191.

[114] Fernando Montani, Adam Kohn, Matthew A Smith, and Simon R Schultz. “The role of correlations in direction and contrast coding in the primary visual cortex”. In: *Journal of Neuroscience* 27.9 (2007), pp. 2338–2348.

[115] Marcelo A Montemurro, Riccardo Senatore, and Stefano Panzeri. “Tight data-robust bounds to mutual information combining shuffling and model selection techniques”. In: *Neural Computation* 19.11 (), pp. 2913–2957.

[116] Ruben Moreno-Bote, Jeffrey Beck, Ingmar Kanitscheider, Xaq Pitkow, Peter Latham, and Alexandre Pouget. “Information-limiting correlations”. In: *Nature neuroscience* 17.10 (2014), pp. 1410–1417.

[117] Michele Nardin, Jozsef Csicsvari, Gašper Tkačik, and Cristina Savin. “The structure of hippocampal CA1 interactions optimizes spatial coding across experience”. In: *Journal of Neuroscience* 43.48 (2023), pp. 8140–8156.

[118] Jonathan J Nassi and Edward M Callaway. “Parallel processing strategies of the primate visual system”. In: *Nature reviews neuroscience* 10.5 (2009), pp. 360–372.

[119] Aran Nayebi, Daniel Bear, Jonas Kubilius, Kohitij Kar, Surya Ganguli, David Sussillo, James J DiCarlo, and Daniel L Yamins. “Task-driven convolutional recurrent models of the visual system”. In: *Advances in neural information processing systems* 31 (2018).

[120] Amin Nejatbakhsh, Isabel Garon, and Alex Williams. “Estimating Noise Correlations Across Continuous Conditions With Wishart Processes”. In: *Advances in Neural Information Processing Systems* 36 (2024).

[121] Ilya Nemenman, Geoffrey D Lewen, William Bialek, and Rob R de Ruyter van Steveninck. “Neural coding of natural stimuli: information at sub-millisecond resolution”. In: *PLoS computational biology* 4.3 (), e1000025.

[122] Trang-Anh Nghiem, Bartosz Telenczuk, Olivier Marre, Alain Destexhe, and Ulisse Ferrari. “Maximum-entropy models reveal the excitatory and inhibitory correlation structures in cortical neuronal activity”. In: *Physical Review E* 98.1 (2018), p. 012402.

[123] Sunny Nigam, Sorin Pojoga, and Valentin Dragoi. “Synergistic coding of visual information in columnar networks”. In: *Neuron* 104.2 (2019), pp. 402–411.

[124] Sheila Nirenberg, Steve M Carcieri, Adam L Jacobs, and Peter E Latham. “Retinal ganglion cells act largely as independent encoders”. In: *Nature* 411.6838 (2001), pp. 698–701.

[125] Valdemar Kargård Olsen, Jonathan R Whitlock, and Yasser Roudi. “The quality and complexity of pairwise maximum entropy models for large cortical populations”. In: *PLOS Computational Biology* 20.5 (2024), e1012074.

[126] Liam Paninski. “Maximum likelihood estimation of cascade point-process neural encoding models”. In: *Network: Computation in Neural Systems* 15.4 (2004), p. 243.

[127] Stefano Panzeri, Monica Moroni, Houman Safaai, and Christopher D Harvey. “The structures and functions of correlations in neural population codes”. In: *Nature Reviews Neuroscience* 23.9 (), pp. 551–567.

[128] Stefano Panzeri, Simon R Schultz, Alessandro Treves, and Edmund T Rolls. “Correlations and the encoding of information in the nervous system”. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266.1423 (1999), pp. 1001–1012.

[129] Stefano Panzeri, Alessandro Treves, Simon Schultz, and Edmund T Rolls. “On decoding the responses of a population of neurons from short time windows”. In: *Neural computation* 11.7 (), pp. 1553–1577.

[130] Il Memming Park, Evan W Archer, Nicholas Priebe, and Jonathan W Pillow. “Spectral methods for neural characterization using generalized quadratic models”. In: *Advances in neural information processing systems* 26 (2013).

[131] Il Memming Park, Miriam LR Meister, Alexander C Huk, and Jonathan W Pillow. “Encoding and decoding in parietal cortex during sensorimotor decision-making”. In: *Nature neuroscience* 17.10 (2014), pp. 1395–1403.

[132] Leo Peichl. “Alpha ganglion cells in mammalian retinae: common properties, species differences, and some comments on other ganglion cells”. In: *Visual neuroscience* 7.1-2 (1991), pp. 155–169.

[133] Donald H Perkel, George L Gerstein, and George P Moore. “Neuronal spike trains and stochastic point processes: II. Simultaneous spike trains”. In: *Biophysical journal* 7.4 (), pp. 419–440.

[134] Jonathan Pillow. *Likelihood-based approaches to modeling the neural code*. Vol. 70. 3. MIT press Cambridge, Massachusets, 2007.

[135] Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. “Spatio-temporal correlations and visual signalling in a complete neuronal population”. In: *Nature* 454.7207 (2008), pp. 995–999.

[136] Xaq Pitkow and Markus Meister. “Decorrelation and efficient coding by retinal ganglion cells”. In: *Nature neuroscience* 15.4 (2012), pp. 628–635.

[137] G Pola, A Thiele, KP Hoffmann, and S Panzeri. “An exact method to quantify the information transmitted by different mechanisms of correlational coding”. In: *Network: Computation in Neural Systems* 14.1 (2003), p. 35.

[138] Adrián Ponce-Alvarez, Alexander Thiele, Thomas D Albright, Gene R Stoner, and Gustavo Deco. “Stimulus-dependent variability and noise correlations in cortical MT neurons”. In: *Proceedings of the National Academy of Sciences* 110.32 (), pp. 13162–13167.

[139] James FA Poulet and Carl CH Petersen. “Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice”. In: *Nature* 454.7206 (2008), pp. 881–885.

[140] Jose C Principe and Jyh-Ming Kuo. “Dynamic modelling of chaotic time series with neural networks”. In: *Advances in neural information processing systems* 7 (1994).

[141] Floyd Ratliff and H KEFFER Hartline. “The responses of Limulus optic nerve fibers to patterns of illumination on the receptor mosaic”. In: *The Journal of general physiology* 42.6 (1959), pp. 1241–1255.

[142] Colleen E Rhoades, Nishal P Shah, Michael B Manookin, Nora Brackbill, Alexandra Kling, Georges Goetz, Alexander Sher, Alan M Litke, and EJ Chichilnisky. “Unusual physiological properties of smooth monostratified ganglion cell types in primate retina”. In: *Neuron* 103.4 (2019), pp. 658–672.

[143] Fred Rieke, David Warland, Rob de Ruyter Van Steveninck, and William Bialek. *Spikes: exploring the neural code*. MIT press, 1999.

[144] Rajeev V Rikhye, Aditya Gilra, and Michael M Halassa. "Thalamic regulation of switching between cortical representations enables cognitive flexibility". In: *Nature neuroscience* 21.12 (2018), pp. 1753–1763.

[145] Robert W Rodieck. "Quantitative analysis of cat retinal ganglion cell response to visual stimuli". In: *Vision research* 5.12 (1965), pp. 583–601.

[146] 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". In: *PLoS computational biology* 5.5 (), e1000380.

[147] Kaushambi Roy, Sandeep Kumar, and Stewart A Bloomfield. "Gap junctional coupling between retinal amacrine and ganglion cells underlies coherent activity integral to global object perception". In: *Proceedings of the National Academy of Sciences* 114.48 (2017), E10484–E10493.

[148] Suva Roy, Na Young Jun, Emily L Davis, John Pearson, and Greg D Field. "Intermosaic coordination of retinal receptive fields". In: *Nature* 592.7854 (2021), pp. 409–413.

[149] Kiersten Ruda, Joel Zylberberg, and Greg D Field. "Ignoring correlated activity causes a failure of retinal population codes". In: *Nature communications* 11.1 (2020), p. 4605.

[150] Daniel L Ruderman. "The statistics of natural images". In: *Network: Computation in Neural Systems* 5.4 (1994), pp. 517–548.

[151] Caroline A Runyan, Eugenio Piasini, Stefano Panzeri, and Christopher D Harvey. "Distinct timescales of population coding across cortex". In: *Nature* 548.7665 (2017), pp. 92–96.

[152] Cristina Savin and Gašper Tkačik. "Maximum entropy models as a tool for building precise neural controls". In: *Current opinion in neurobiology* 46 (2017), pp. 120–126.

[153] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. "Weak pairwise correlations imply strongly correlated network states in a neural population". In: *Nature* 440.7087 (2006), pp. 1007–1012.

[154] Greg Schwartz, Sam Taylor, Clark Fisher, Rob Harris, and Michael J Berry. "Synchronized firing among retinal ganglion cells signals motion reversal". In: *Neuron* 55.6 (2007), pp. 958–969.

[155] Vitor Sessak and Remi Monasson. "Small-correlation expansions for the inverse Ising problem". In: *Journal of Physics A: Mathematical and Theoretical* 42.5 (2009), p. 055001.

[156] Claude Elwood Shannon. “A mathematical theory of communication”. In: *ACM SIGMOBILE mobile computing and communications review* 5.1 (2001), pp. 3–55.

[157] Hideaki Shimazaki, Shun-ichi Amari, Emery N Brown, and Sonja Grün. “State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data”. In: *PLoS computational biology* 8.3 (), e1002385.

[158] Hideaki Shimazaki, Kolia Sadeghi, Tomoe Ishikawa, Yuji Ikegaya, and Taro Toyoizumi. “Simultaneous silence organizes structured higher-order interactions in neural populations”. In: *Scientific reports* 5.1 (), p. 9821.

[159] Jonathon Shlens, Greg D Field, Jeffrey L Gauthier, Martin Greschner, Alexander Sher, Alan M Litke, and EJ Chichilnisky. “The structure of large-scale synchronized firing in primate retina”. In: *Journal of Neuroscience* 29.15 (2009), pp. 5022–5031.

[160] Jonathon Shlens, Greg D Field, Jeffrey L Gauthier, Matthew I Grivich, Dumitru Petrusca, Alexander Sher, Alan M Litke, and EJ Chichilnisky. “The structure of multi-neuron firing patterns in primate retina”. In: *Journal of Neuroscience* 26.32 (2006), pp. 8254–8266.

[161] Jonathon Shlens, Fred Rieke, and EJ Chichilnisky. “Synchronized firing in the retina”. In: *Current opinion in neurobiology* 18.4 (2008), pp. 396–402.

[162] Rava Azeredo da Silveira and Michael J Berry. “High-fidelity coding with correlated neurons”. In: *PLoS computational biology* 10.11 (), e1003970.

[163] Kristina D Simmons, Jason S Prentice, Gašper Tkačik, Jan Homann, Heather K Yee, Stephanie E Palmer, Philip C Nelson, and Vijay Balasubramanian. “Transformation of stimulus correlations by the retina”. In: *PLoS computational biology* 9.12 (2013), e1003344.

[164] Eero P Simoncelli, Liam Paninski, Jonathan Pillow, Odelia Schwartz, et al. “Characterization of neural responses with stochastic stimuli”. In: *The cognitive neurosciences* 3.327-338 (2004), p. 1.

[165] Matthew A Smith and Adam Kohn. “Spatial and temporal scales of neuronal correlation in primary visual cortex”. In: *Journal of Neuroscience* 28.48 (2008), pp. 12591–12603.

[166] Haim Sompolinsky, Hyoungsoo Yoon, Kukjin Kang, and Maoz Shamir. “Population coding in neuronal systems with correlated noise”. In: *Physical Review E* 64.5 (2001), p. 051904.

[167] Oleksandr Sorochynskyi, Stephane Deny, Olivier Marre, and Ulisse Ferrari. “Predicting synchronous firing of large neural populations from sequential recordings”. In: *PLoS computational biology* 17.1 (2021), e1008501.

[168] Peter N Steinmetz, A Roy, PJ Fitzgerald, SS Hsiao, KO Johnson, and Ernst Niebur. "Attention modulates synchronized neuronal firing in primate somatosensory cortex". In: *Nature* 404.6774 (2000), pp. 187–190.

[169] Peter Sterling and Simon Laughlin. *Principles of neural design*. MIT press, 2015.

[170] Steven P Strong, Roland Koberle, Rob R De Ruyter Van Steveninck, and William Bialek. "Entropy and information in neural spike trains". In: *Physical review letters* 80.1 (), p. 197.

[171] Hidenori Tanaka, Aran Nayebi, Niru Maheswaranathan, Lane McIntosh, Stephen Baccus, and Surya Ganguli. "From deep learning to mechanistic understanding in neuroscience: the structure of retinal prediction". In: *Advances in neural information processing systems* 32 (2019).

[172] Toshiyuki Tanaka. "Mean-field theory of Boltzmann machine learning". In: *Physical Review E* 58.2 (1998), p. 2302.

[173] Aonan Tang, David Jackson, Jon Hobbs, Wei Chen, Jodi L Smith, Hema Patel, Anita Prieto, Dumitru Petrusca, Matthew I Grivich, Alexander Sher, et al. "A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro". In: *Journal of Neuroscience* 28.2 (), pp. 505–518.

[174] Gaia Tavoni, Ulisse Ferrari, Francesco P Battaglia, Simona Cocco, and Remi Monasson. "Functional coupling networks inferred from prefrontal cortex activity show experience-related effective plasticity". In: *Network Neuroscience* 1.3 (2017), pp. 275–301.

[175] David J Thouless, Philip W Anderson, and Robert G Palmer. "Solution of 'solvable model of a spin glass'". In: *Philosophical Magazine* 35.3 (1977), pp. 593–601.

[176] Gašper Tkačik, Olivier Marre, Dario Amodei, Elad Schneidman, William Bialek, and Michael J Berry. "Searching for collective behavior in a large network of sensory neurons". In: *PLoS computational biology* 10.1 (2014), e1003408.

[177] Gašper Tkačik, Olivier Marre, Dario Amodei, Elad Schneidman, William Bialek, and Michael J Berry II. "Searching for collective behavior in a network of real neurons". In: *arXiv preprint arXiv:1306.3061* ().

[178] Gašper Tkačik, Jason S Prentice, Vijay Balasubramanian, and Elad Schneidman. "Optimal population coding by noisy spiking neurons". In: *Proceedings of the National Academy of Sciences* 107.32 (2010), pp. 14419–14424.

[179] Francesco Trapani, Giulia Lia Beatrice Spampinato, Pierre Yger, and Olivier Marre. "Differences in nonlinearities determine retinal cell types". In: *Journal of Neurophysiology* 130.3 (2023), pp. 706–718.

[180] Stuart Trenholm, Amanda J McLaughlin, David J Schwab, Maxwell H Turner, Robert G Smith, Fred Rieke, and Gautam B Awatramani. “Nonlinear dendritic integration of electrical and chemical synaptic inputs drives fine-scale correlations”. In: *Nature neuroscience* 17.12 (2014), pp. 1759–1766.

[181] Philipp Khuc Trong and Fred Rieke. “Origin of correlated activity between parasol retinal ganglion cells”. In: *Nature neuroscience* 11.11 (2008), pp. 1343–1351.

[182] Wilson Truccolo, Uri T Eden, Matthew R Fellows, John P Donoghue, and Emery N Brown. “A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects”. In: *Journal of neurophysiology* 93.2 (2005), pp. 1074–1089.

[183] W Martin Usrey and R Clay Reid. “Synchronous activity in the visual system”. In: *Annual review of physiology* 61.1 (), pp. 435–456.

[184] Michael Vidne, Yashar Ahmadian, Jonathon Shlens, Jonathan W Pillow, Jayant Kulkarni, Alan M Litke, EJ Chichilnisky, Eero Simoncelli, and Liam Paninski. “Modeling the impact of common noise inputs on the network activity of retinal ganglion cells”. In: *Journal of computational neuroscience* 33 (2012), pp. 97–121.

[185] Anna L Vlasits, Thomas Euler, and Katrin Franke. “Function first: classifying cell types and circuits of the retina”. In: *Current opinion in neurobiology* 56 (2019), pp. 8–15.

[186] Bela Völgyi, Samir Chheda, and Stewart A Bloomfield. “Tracer coupling patterns of the ganglion cell subtypes in the mouse retina”. In: *Journal of Comparative Neurology* 512.5 (2009), pp. 664–687.

[187] Bela Völgyi, Feng Pan, David L Paul, Jack T Wang, Andrew D Huberman, and Stewart A Bloomfield. “Gap junctions are essential for generating the correlated spike activity of neighboring retinal ganglion cells”. In: *PloS one* 8.7 (2013), e69426.

[188] Fei Wang, E Li, Lei De, Qiwen Wu, and Yifeng Zhang. “OFF-transient alpha RGCs mediate looming triggered innate defensive response”. In: *Current Biology* 31.11 (2021), pp. 2263–2273.

[189] Heinz Wässle. “Parallel processing in the mammalian retina”. In: *Nature Reviews Neuroscience* 5.10 (2004), pp. 747–757.

[190] Heinz Wässle, Christian Puller, Frank Müller, and Silke Haverkamp. “Cone contacts, mosaics, and territories of bipolar cells in the mouse retina”. In: *Journal of Neuroscience* 29.1 (2009), pp. 106–117.

[191] Alison I Weber and Jonathan W Pillow. "Capturing the dynamical repertoire of single neurons with generalized linear models". In: *Neural computation* 29.12 (2017), pp. 3260–3289.

[192] Xue-Xin Wei and Alan A Stocker. "Mutual information, Fisher information, and efficient coding". In: *Neural computation* 28.2 (2016), pp. 305–326.

[193] Stefan D Wilke and Christian W Eurich. "Representational accuracy of stochastic neural populations". In: *Neural computation* 14.1 (2002), pp. 155–189.

[194] Si Wu, Shun-ichi Amari, and Hiroyuki Nakahara. "Population coding and decoding in a neural field: a computational study". In: *Neural Computation* 14.5 (2002), pp. 999–1026.

[195] Stuart Yarrow, Edward Challis, and Peggy Seriès. "Fisher and Shannon information in finite neural populations". In: *Neural computation* 24.7 (2012), pp. 1740–1780.

[196] Pierre Yger, Giulia LB Spampinato, Elric Esposito, Baptiste Lefebvre, Stéphane Deny, Christophe Gardella, Marcel Stimberg, Florian Jetter, Guenther Zeck, Serge Picaud, et al. "A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo". In: *Elife* 7 (2018), e34518.

[197] Anthony Zador. "Spikes: Exploring the neural code". In: *Science* 277.5327 (), pp. 772–773.

[198] Yifeng Zhang, In-Jung Kim, Joshua R Sanes, and Markus Meister. "The most numerous ganglion cell type of the mouse retina is a selective feature detector". In: *Proceedings of the National Academy of Sciences* 109.36 (2012), E2391–E2398.

[199] Ehud Zohary, Michael N Shadlen, and William T Newsome. "Correlated neuronal discharge rate and its implications for psychophysical performance". In: *Nature* 370.6485 (1994), pp. 140–143.

[200] Joel Zylberberg, Jon Cafaro, Maxwell H Turner, Eric Shea-Brown, and Fred Rieke. "Direction-selective circuits shape noise to ensure a precise population code". In: *Neuron* 89.2 (2016), pp. 369–383.