Retinal processing: insights from mathematical modelling

Bruno Cessac

Retinal processing: insights from mathematical modelling

Bruno Cessac
Université Côte d'Azur INRIA, France
INRIA Biovision team and Neuromod Institute
bruno.cessac@inria.fr

January 17, 2022

Abstract

The retina is the entrance of the visual system. Although based on common biophysical principles the dynamics of retinal neurons is quite different from their cortical counterparts, raising interesting problems for modellers. In this paper I address some mathematically stated questions in this spirit, discussing, in particular: (1) How could lateral amacrine cell connectivity shape the spatio-temporal spike response of retinal ganglion cells ? (2) How could spatio-temporal stimuli correlations and retinal network dynamics shape the spike train correlations at the output of the retina ? These questions are addressed, first, introducing a mathematically tractable model of the layered retina, integrating amacrine cells lateral connectivity and piecewise linear rectification, allowing to compute the retinal ganglion cells receptive field together with the voltage and spike correlations of retinal ganglion cells resulting from the amacrine cells networks. Then, I review some recent results showing how the concept of spatio-temporal Gibbs distributions and linear response theory can be used to characterize the collective spike response to a spatio-temporal stimulus of a set of retinal ganglion cells, coupled via effective interactions corresponding to the amacrine cells network. On these bases, I briefly discuss several potential consequences of these results at the cortical level.

1 Introduction

Let us start with a very simple experiment. Look around you... That's it, the experiment is over. A very ordinary experience, isn't it? Is it really though? Let us first point out that looking around you to see, that is, having the sense of sight, is indeed ordinary — except for those who have partially or totally lost their ability to see. We will come back to this point at the end of the paper. Now, excluding visual impairments, vision is everything but ordinary.

Think of it. A flux of photons, with frequencies in the visible spectrum range, emitted by the external world around us enters into our eyes, then "something"

happens, and we see. Thanks to constant progress in experimental and theoretical neuroscience, we understand better and better this "something", the mechanisms of vision, although our view of it is far from being complete. Especially, in these times of artificial intelligence, bio-inspired computing, computer vision, it might be helpful to understand how our brain is able to handle the complex visual information coming from the external world so rapidly and efficiently with an energy consumption of the order of a few Watts.

Certainly, the retina plays a central role in this process. It is known for long that this is definitely not a camera. The retina is smart [**?**] and it has to be. Think especially of the difference of scale between the retina and the visual cortex, in terms of size but also numbers of neurons and synapses. As everything that goes to the visual cortex comes from the retina, this little membrane, at the back of the eye, half a millimetre thick, with an area of order a cm^2 (for humans), has to some extent to filter the visual information, leaving out "irrelevant detail" and capture crucial events, and then, signal them appropriately to the brain via spike trains. As a matter of fact, the question(s) of "efficiently" encoding information by spikes has been the subject of many fascinating papers [**?**, **?**, **?**], especially in the seminal paper from Barlow [**?**] with concepts such as reducing redundancy, information compression and efficient coding. These concepts are regularly updated with recent experimental and theoretical investigations [**?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**]. We come back to this point, at the end of the paper too.

The retina has, roughly, the following structure. For more detail see e.g. [**?**] or `https://webvision.med.utah.edu/book/part-i-foundations/simple-anatomy-of-the-retina/`. It is organized in five neuronal types : Photo-receptors, rod and cones (P), horizontal cells (H cells), bipolar cells (B cells), amacrine cells (A cells), retinal ganglion cells (RG cells), to which are added glial cells (Mueller's cells). These neurons types are connected by chemical and electric synapses, in specific functional circuits or "pathways" (like the rod-cone pathway [**?**, **?**]) which are a key in the retinal capacity to convert the light coming from a visual scene into spike patterns sent to the visual cortex, through the Lateral Geniculate Nucleus (LGN), via the optic nerve made of RG cells axons. In particular, there are in the retina very specific synapses like the ribbon synapse enabling neurons to transmit light signals from photoreceptors to B cells over a dynamic range of several orders of magnitude in intensity [**?**]. Roughly, two main connectivity structures can be distinguished: feed-forward, the P-B-G path which leads from the photo transduction to the spike trains emitted by the RG cells towards the cortex. There is also a lateral connectivity through H cells, at the origin of the Center-Surround structure of the receptive fields, and the A cells whose role is still poorly understood and which are one of the main objects of study of this paper.

The structure of the retina and its behaviour are thus well studied on the experimental side. There are comparatively fewer modelling studies although important work has been done on retinal coding [**?**, **?**, **?**, **?**, **?**, **?**], biophysically de-

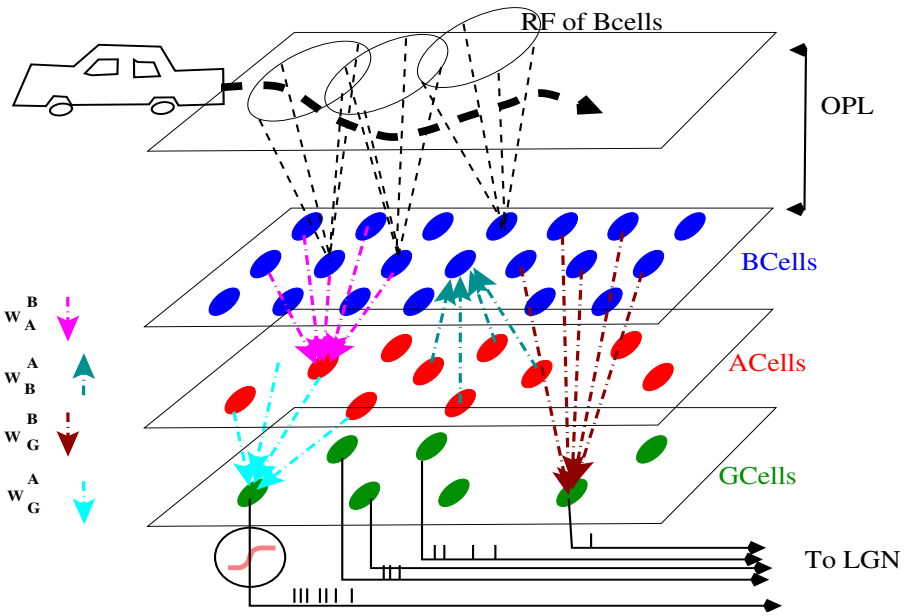

Figure 1: **Structure of the retina model introduced in section ??.** A moving object (here, presumably, a car) moves along a trajectory (dashed black line). Its image is projected by the eye optics to the upper retina layers (Photoreceptors and H cells) and stimulates them. In the model, this corresponds to the convolution of the stimulus with the Receptive Field (RF) of B cells. This provides to B cells what we call the "OPL" input to B cells. B cells (blue points) are connected to A cells (red points) via excitatory synapses (pink arrows, denoted W_A^B) and to RG cells (green points) via excitatory synapses (brown arrows, denoted W_G^B). A cells are connected to B cells via inhibitory synapses (green arrows, denoted W_B^A) and to RG cells via inhibitory synapses (cyan arrows, denoted W_G^A). The voltage of RG cells is sent through a non linearity (pink curve in the black circle) so as to produce spike trains conveyed to the LGN.

tailed models [?, ?, ?], Generalized Linear Models applied to retinal coding [?, ?, ?]. Several powerful software has been designed to model the retina at different scales such as COREM [?], Convis [?], Isetbio `https://github.com/isetbio/isetbio/wiki`. The Virtual Retina simulator, developed by A. Wohrer and P. Kornprobst [?] at INRIA was one of the first of these simulators and has given rise to subsequent simulators in our group, the platform PRANAS [?], `https://team.inria.fr/biovision/pranas-software/` and more recently Macular `https://team.inria.fr/biovision/macular-software/`. There are quite less mathematical results on how retinal structure, especially lateral A cells connectivity, shapes the spike response to spatio-temporal stimuli [?, ?, ?].

One of the goals of this paper is to elicit reflections in this direction, grounded on mathematical developments fed by the recent progress in the knowledge of retina physiology and structure. This is a humble and partial point of view, resulting from my collaboration with neurobiologists experts in the retina. The paper contains new results, essentially the mathematically tractable model of the layered retina integrating amacrine cells lateral connectivity and the mathematical framework to handle piecewise linear rectification presented in section **??**, the study of rectification effects on retinal ganglion cells receptive field (section **??**), the study of voltage and spike correlations of retinal ganglion cells (section **??**) and the discussion about mixed effect of network and stimulus on spike correlations in section **??**. It also contains already published material, essentially the framework and results dealing with Gibbs distributions and linear response (sections **??**, **??**).

The goal is to draw a common thread about the potential role of amacrine cells from retinal spatio-temporal stimuli response to spike coding. More precisely, I am addressing the following problems on mathematical grounds. In the main text I focus on neuroscience modelling perspective, whereas, in the conclusion section, I discuss potential consequences of these results out of the field of neuroscience.

Problem 1. How does the structure of the retina, in particular, amacrine lateral connectivity condition the retinal response to dynamic stimuli? The problem can be addressed at two levels.

Level 1. Single cell response to stimuli. The individual response of ganglion cells is usually expressed in terms of their receptive field. This notion is on the one hand phenomenological: it is observed that each ganglion cell responds preferentially to stimuli, localized in space, with a characteristic spatio-temporal structure. For example, a ON-Center cell preferentially responds to an increase in luminance in a circular area corresponding to the central part of the receiving field. This notion is also expressed mathematically by a kernel $\mathcal{K}_G$, i.e. a function of space and time, so that the response of a RG cell to a spatio-temporal

stimulus $S(x, y, t)$, takes the form:

$$\left[\mathcal{K}_G \overset{x,y,t}{*} \mathcal{S}\right](t) = \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{s=-\infty}^{t} \mathcal{K}_G(x-x_C, y-y_C, t-s)\, \mathcal{S}(x, y, s) dx\, dy\, ds, \tag{1}$$

where $\overset{x,y,t}{*}$ means space (x, y)-time (t) convolution. x_C, y_C are the coordinates of the RF center. The integrals are well defined since the kernel decreases fast enough to infinity, in space and time, to guarantee convergence. The upper bound in time, t, expresses causality, whereas the lower bound, $-\infty$, implicitly assumes that the stimulus has been applied in a distant past compared to t, quite longer than the characteristic times involved in RG cell response.

Equation (**??**) corresponds to a linear response. It is therefore only valid for stimuli of low amplitude in voltage. More generally, the voltage response to the stimulus is a functional of the stimulus that one can, under well posed mathematical conditions, write as a Volterra expansion [**?**], (**??**) being the lowest order (linear) term. Unfortunately, higher-order terms are essentially inaccessible experimentally and one usually constrains instead the non-linearity of the response under other modalities. Especially, taking into account that the response of a ganglion cell to a stimulus is, ultimately, a sequence of spikes, one writes the probability density of emitting a spike between t and $t + dt$ in the form $f\left(\left[\mathcal{K}_G \overset{x,y,t}{*} \mathcal{S}\right](t) + b\right)$ where f is a non-linear positive increasing function (typically, a sigmoid), b is a threshold constraining the level of activity of the RG cell in the absence of stimulation. This procedure defines an inhomogeneous Poisson process called the linear-non-linear Poisson (LNP) model [**?**, **?**, **?**] . Experimentally the kernel $\mathcal{K}_G$ is determined by Spike-Triggered Average or Spike-Triggered Correlation technique, studying the response to a white noise [**?**]. Non-linearity is then determined, typically by the Levenberg-Marquart method [**?**]. This modelling asks however the following questions:

(i) How is the kernel $\mathcal{K}_G$ of the RG cells constrained by the structure/dynamics of the upper layers of retinal cells ?

(ii) The forms (**??**) implicitly assumes that $\mathcal{K}_G$ does not depend on the stimulus. Can one write mathematical conditions that guarantee such an independence?

(iii) To which extent is the notion of Ganglion cells Receptive Field compatible with non linear effects reported in retinal neurons and synapses, such as voltage rectification or gain control ?

Level 2. Collective response to stimuli and spike statistics. RG cells do not interact directly, but amacrine connectivity induces an effective interaction between them. What is therefore the structure of the spatio-temporal correlations induced by the conjunction of the spatio-temporal stimulus and the response of the retinal network, in particular, the amacrine lateral connectivity ? A classical paradigm in neural coding is to assume that the retina decorrelates RG

cells outputs to maximize information transfer [**?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**]. It is in particular believed that A cells play a central role in this decorrelation process (see [**?**] and references therein). What can be, at the mathematical level, the conditions, on the stimulus and dynamics, that allow a network of neurons interacting with each other to produce vanishing, or at least, *weak* correlations ? When does *weak* mean *negligible* ? These questions are actually closely related to the second problem.

Problem 2. How do retina network and dynamics shape spike statistics in the response to stimuli ? More generally, considering the retina as a dynamical system forced by non-stationary, spatially inhomogeneous stimuli, what could be a general form for the (non-stationary) statistics of spike trains emitted by ganglion cells, taking into account that spike trains emitted by the retina are all that the LGN and cortex see ? One can attempt to construct a canonical form of probability distributions of the retinal spike trains taking into account that:

(i) Stimuli, thus statistics, are not stationary;

(ii) The cortex (and before, the LGN) only receive spikes, thus have no information about the biophysical processes which have generated those spikes and no information on the underlying dynamics of the retina (voltages, activation variables, conductances). All the information is contained in the spatio-temporal structure of spikes;

(iii) Spike train distributions may exhibit long time scale dependence (i.e. have a long memory).

In this paper I address these problems with the help of two models. The first, presented in section **??**, grounded on biology and e.g. the papers [**?**, **?**, **?**, **?**] mimics the Bipolar-Amacrine-Ganglion cells network and is used, in sections **??**, **??**, to make progresses in elucidating problem 1. I first show how one can obtain an explicit form for the kernel (**??**) featuring the A cells lateral connectivity. This RF explicitly depends on the BC cells-A cells network through the eigenvalues and eigenvectors of an operator I call "transport operator". I discuss some consequences of this result, especially in terms of response to propagating stimulus. This result is valid when cells act as linear integrators. However, cells are in general rectified by non linearities. I propose piecewise-linear rectifications (as used in several retina model) and I discuss how rectification acts on the RF of eq. (**??**). A striking conclusion is that, if the convolution form (**??**) is preserved, this is to the price of having a RF depending on the stimulus. A consequence of this analysis is that spike correlations may depend on the stimulus and are expected to be quite different when considering e.g. objects moving along trajectories in comparison to static images.

The second model, introduced in section **??** and analysed in section **??** attempts to propose a canonical form of probability distributions of the retinal

spike trains based on the constraints (i), (ii), (iii) above. These sections essentially presents the conclusions of works published elsewhere [**?**, **?**, **?**, **?**, **?**, **?**]. As I argue, these constraints lead to a natural notion of spike probabilities, somewhat extending the statistical physics notion of Gibbs distribution to the non stationary case. In this setting, one establishes a linear response for a network of interacting spiking cells that can mimic a set of RG cells coupled via effective interactions corresponding to the A cells network influence. This linear response theory not only gives the effect of a non stationary stimulus to first order spike statistics (firing rates) but also its effect on higher order correlations. Indeed, spike correlations are modified by a spatio-temporal stimulus and can be computed thanks to the knowledge of spontaneous correlations. The linear response formula is expressed as a convolution where the kernel can be explicitly computed for an Integrate and Fire conductance based model [**?**]. Moreover, as I argue, these spike trains distributions have close links with information geometry. Especially, they induce a natural metric in an abstract space of probabilities, with close potential links with the neuro-geometry introduced by Sarti, Citti, Petitot et al [**?**, **?**, **?**, **?**]. This is discussed in the conclusion section.

More generally, the application and discussion sections shortly proposes possible extension of this work to several domains: Retinal prostheses, section **??**; Convolutional networks, section **??**; Implications for cortical models, section **??**; Neuro-geometry, section **??**.

2 Materials and Methods

2.1 Modelling the retinal network

2.1.1 Specifics of the retina

Neurons in the retina have the same biophysics as their cortical counterparts. However, they operate under different modalities. Remarkably, with the exception of the RG cells, the retinal neurons do not emit action potentials. Their activity and interactions take therefore place through graded (continuous) membrane potentials as opposed to the sharp peak of an action potential. Furthermore, there is no long-term synaptic plasticity in the retina. Finally, the main "computational" elements in the retina are functional circuits [**?**] made of a few neurons and synapses, in large contrast with "computational" units in the visual cortex, such as cortical columns, involving thousands of neurons. A modelling consequence is that mean-field or neural masses description used in the cortex might not be relevant to study the retina.

The goal of this paper is to address mathematical questions about the dynamics and behaviour of the retina embedded in the visual system. To instantiate these questions on a firm mathematical ground we are going to consider a model of the retinal network, based on a few fundamental facts briefly exposed in the previous section:

1. The retina is a high dimensional, non autonomous and noisy dynamical system, layered and structured, with non stationary and spatially inhomogeneous entries (visual scenes).

2. Most retinal neurons are not spiking, except RG cells. Thus, retina performs analogic computing.

3. Local retinal circuits efficiently process the local visual information. These local circuits are connected together, spanning the whole retina in a regular tiling. From this perspective, it is important to consider individual neurons and synapses, in contrast, e.g., to cortical modelling where it is relevant to consider mean-field approaches averaging over populations.

Thus, the model presented below and in Fig. **??** is non stationary, with a layered retina like structure, where dynamics ruling B cells, A cells, and RG cells voltage is piecewise linear. As we discuss, the model affords additional non linearities like gain control. For RG cells, the spiking process is mimicked by a non linear firing rate so that our model enters in the class of LNP models.

2.1.2 Structure of the retina model

We assimilate the retina to a superimposition of 3 layers, each one being a flat, two dimensional square of edge length L mm where spatial coordinates are noted x, y (Fig. **??**). Each layer corresponds to a cell population (B cells, A cells, RG cells) where the density of cells is taken uniform. We note δ_p the lattice spacing in mm, and N_p the total number of cells in the layer p. Without loss of generality we assume that L, the retina's edge size, is a multiple of δ_p. We note $L_p = \frac{L}{\delta_p}$, the number of cells p per row or column so that $N_p = L_p^2$. Each cell in the population p has thus Cartesian coordinates $(x, y) = (i_x\delta_p, i_y\delta_p)$, $(i_x, i_y) \in \{1, \ldots, L_p\}^2$. To avoid multiples indices, we associate to each pair (i_x, i_y) a unique index $i = i_x + (i_y - 1) L_p$. The cell of population p, located at coordinates $(i_x\delta_p, i_y\delta_p)$ is then denoted by p_i.

One can roughly subdivide the real retina into two blocks (Fig. **??**). The first, that we name in short, for modelling purposes[1] OPL, (Outer Plexiform Layer), includes the P, H cells, B cells and the related synapses. As an "input" of this block is the flow of photons emitted by the outside world and picked up by the photo-receptors. In our model, this corresponds to a "stimulus", i.e. a function $\mathcal{S}(x, y, t)$ where x, y are (two-dimensional) space coordinates and t is the time coordinate. As we don't consider color sensitivity here $\mathcal{S}$ characterizes a black and white scene, with a control on the level of contrast $\in [0, 1]$. The

[1]Note that the terminology OPL and IPL refers actually to synaptic layers. "The outer plexiform layer has a wide external band composed of inner fibres of rods and cones and a narrower inner band consisting of synapses between photoreceptor cells and cells from the inner nuclear layer." "The inner plexiform layer consists of synaptic connections between the axons of bipolar cells and dendrites of ganglion cells" (ref `https://www.sciencedirect.com/topics/medicine-and-dentistry/`). In our model, these naming are short cuts to distinguish the network input (OPL) and the network processing (IPL).

"output" of the OPL is sent to B cells in the form of a "drive" voltage, defined in eq. (**??**) below. In the real retina, the voltage of each BCell integrates, spatially and temporally, the local visual information of the photo-receptors which are connected to it, with a lateral modulation due to the H cells. Each B cells is thus sensitive to specific local characteristics of the visual scene, defining its Receptive Field (RF). Thus, B cells, like RG cells, have a receptive field. But, as they are earlier in the vertical pathway they integrate less features. Note that the RF of distinct B cells usually overlap creating correlations between B cells voltages (see section **??**).

We label B cells (layer 1) with the index $i = 1, \ldots, N_B$ and we model the RF of B cells by a convolution kernel, $\mathcal{K}_{B_i}$, such that the voltage of BCell i is stimulus-driven by the term:

$$V_{i_{drive}}(t) = \left[\mathcal{K}_{B_i} \overset{x,y,t}{*} \mathcal{S} \right](t). \tag{2}$$

The center of the RF, located at x_i, y_i, also corresponds to the coordinates of the BCell i. A typical shape for the RF of B cells is illustrated in Fig. **??**, although the explicit form does not play a role in the subsequent mathematical developments.

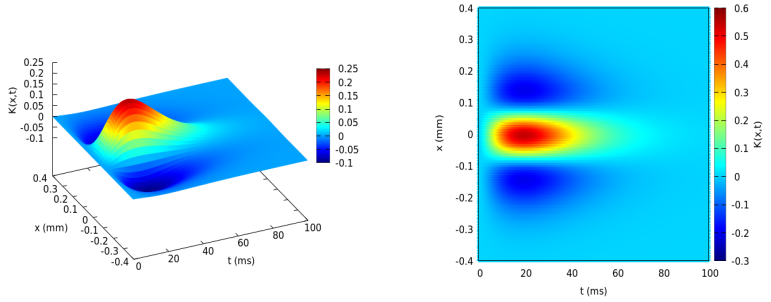

Figure 2: **Receptive Field of a ON BCell. Left.** Example of a spatio-temporal RF of B cells (ON center cell) represented in 3D (one dimension of space, x and time t). There is inhibition at the surround, physiologically due to H cells. **Right.** Spatio-temporal RF representation with a color map.

The second block, that we name in short IPL (Inner Plexiform Layer), comprises the A cells and RG cells and the afferent synapses. Its "input" is the output of the OPL, and its output, the trains of action potentials emitted by the RG cells. A cells are difficult to study experimentally because they are hardly accessible from electrophysiology measurements. There are also a large number of cell subtypes in the A cells class (around 40), of which only a small number have duly identified functions. It is however recognized that they play an essential role in the treatment of motion [**?**, **?**, **?**]. Here we address mathematically the question of the RG cell receptive field form, resulting from the pooling

of B cells, as illustrated in Fig. **??**, each with a specific RF as exemplified in Fig. **??**, and modulated by A cells lateral connectivity.

2.1.3 B cells-A cells interactions

We label A cells (second layer) with the index $j = 1 \dots N_A$. We note $W_{B_i}^{A_j}$ the synaptic weight from A cell j to B cell i and $W_{A_j}^{B_i}$ the synaptic weight from B cell i to A cell j. We set $W_{B_i}^{A_j} \leq 0$, (A cells are in general inhibitory although some excitatory A cells exist, not considered here) whereas $W_{A_j}^{B_i} \geq 0$. The synaptic weight matrices B cells to A cells and A cells to B cells are noted W_A^B, W_B^A. They are not squared in general. There also exist electric synapses (gap junctions) between B cells and A cells (e.g. in the Rod Cone pathway [**?**], see also [**?**] and https://www.ncbi.nlm.nih.gov/books/NBK549947/) but we will not consider them in first place, for simplicity. Note however, as shortly discussed in section **??**, that adding gap junctions would simply result in adding linear terms to equations (**??**), (**??**), (**??**) (when considering passive gap junctions) and modify characteristic time scales, without changing the global analysis.

The voltage of B cell i, V_{B_i}, evolves according to:

$$\frac{dV_{B_i}}{dt} = -\frac{1}{\tau_{B_i}} V_{B_i} + \sum_{j=1}^{N_A} W_{B_i}^{A_j} \mathcal{N}_A \left(V_{A_j} \right) + F_{B_i}(t). \tag{3}$$

Here, τ_{B_i} is the characteristic time scale of B cell i response (in ms). The function:

$$\mathcal{N}_A(V) = \begin{cases} V - \theta_A, & \text{if } V_A > \theta_A; \\ 0, & \text{otherwise} \end{cases}, \tag{4}$$

is a linear rectifier ensuring that the synapse $j \to i$ becomes silent when the voltage of the pre-synaptic A cell j, V_{A_j}, is lower than a threshold θ_A. This corresponds to a biophysical fact : a synapse cannot change its sign. For simplicity we consider θ_A to be the same for all A cells, although the present formalism can be extended, e.g., to several families of A cells having different thresholds. Note that linear rectifiers of type (**??**) rectify cell's voltage "from below". Rectification "from above" also exist, ensuring that the cell's voltage does not increase without bounds. A typical mechanism is gain control, where an additional variable, called the activity, increasing as voltage increases, triggers a gain function non linearly dropping down the voltage when it exceeds an upper threshold [**?**, **?**]. Under some mild assumptions gain control can also be implemented as a linear function of the activity. This is discussed in section **??** as an extension to the present model.

Finally, $F_{B_i}(t)$ is the OPL input term. To match classical retina models as developed e.g. in [**?**, **?**] it reads:

$$F_{B_i}(t) = \frac{V_{i_{drive}}}{\tau_B} + \frac{dV_{i_{drive}}}{dt} = \left[\mathcal{K}_{B_i} \overset{x,y,t}{*} \left(\frac{\mathcal{S}}{\tau_B} + \frac{d\mathcal{S}}{dt} \right) \right](t), \tag{5}$$

(where $\mathcal{K}_{B_i}(x, y, 0) = 0$). In short, $F_{B_i}(t)$ is chosen so that, in the absence of A cells interaction, $V_{B_i}(t) = V_{i_{drive}}(t)$. Note that $F_{B_i}(t)$ implements therefore a time derivative of the drive, which makes, e.g. B cells response to moving objects sensitive to changes in directions or speed.

A cells are connected to B cells with chemical synapses.The differential equation obeyed by the voltage of A cell j is:

$$\frac{dV_{A_j}}{dt} = -\frac{1}{\tau_{A_j}} V_{A_j} + \sum_{i=1}^{N_B} W_{A_j}^{B_i} \mathcal{N}_B \left(V_{B_i} \right), \tag{6}$$

where τ_{A_j} is the characteristic time scale of A cell j response, and $\mathcal{N}_B$ has the same form as (**??**), with a threshold θ_B. Note that, in contrast to B cells, A cells do not receive an OPL input.

2.1.4 RG cells.

We label RG cells (third layer) with the index $k = 1 \dots N_G$. They are connected to B cells with excitatory synaptic weights, $W_{G_k}^{B_i} \geq 0$ (e.g. glutamatergic synapses) and to A cells with inhibitory synaptic weights, $W_{G_k}^{A_j} \leq 0$ (e.g. glycinergic or GABA-ergic synapses). Their voltage, V_{G_k}, evolves according to:

$$\frac{dV_{G_k}}{dt} = -\frac{1}{\tau_G} V_{G_k} + \sum_{i=1}^{N_B} W_{G_k}^{B_i} \mathcal{N}_B(V_{B_i}) + \sum_{j=1}^{N_A} W_{G_k}^{A_j} \mathcal{N}_A(V_{A_j}). \tag{7}$$

RG cells are spiking. In the model their spiking activity (firing rate) is defined by a LNP model [**?**, **?**, **?**]. It depends on the voltage via a non linear function $\mathcal{N}_G \left(V_G \right) \equiv f \left(\frac{V_G(t) - \theta_G}{\sigma_G} \right)$, where f is typically a sigmoid. Although the detailed form of f does not matter here, it will be convenient, in the sequel, to consider:

$$\mathcal{N}_G \left(V_G \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{V_G - \theta_G}{\sigma_G}} e^{-\frac{x^2}{2}} \, dx. \tag{8}$$

The parameters θ_G (spiking threshold) and σ_G (controlling the slope of the sigmoid at $V_G = \theta_G$) corresponds, in the case where $\mathcal{N}_G$ has the form (**??**), to the probability that a Gaussian centred Ornstein-Uhlenbeck processes with mean-square deviation σ_G crosses the threshold θ_G.

2.1.5 Joint dynamics

The joint dynamics of all cells voltage is given by the dynamical system:

$$\left\{\begin{array}{lcll} \frac{dV_{B_i}}{dt} & = & -\frac{1}{\tau_B}V_{B_i} + \sum_{j=1}^{N_A} W_{B_i}^{A_j}\,\mathcal{N}_A\left(V_{A_j}\right) + F_{B_i}(t), & i = 1\ldots N_B; \\ \frac{dV_{A_j}}{dt} & = & -\frac{1}{\tau_A}V_{A_j} + \sum_{i=1}^{N_B} W_{A_j}^{B_i}\,\mathcal{N}_B\left(V_{B_i}\right), & j = 1\ldots N_A; \\ \frac{dV_{G_k}}{dt} & = & -\frac{1}{\tau_G}V_{G_k} + \sum_{i=1}^{N_B} W_{G_k}^{B_i}\mathcal{N}_B(V_{B_i}) + \sum_{j=1}^{N_A} W_{G_k}^{A_j}\mathcal{N}_A(V_{A_j}), & k = 1\ldots N_G; \end{array}\right. \tag{9}$$

whereas RG cells spikes are produced by the LNP mechanism described above.

The system of eq. (**??**) can be summarized as follows (Fig. **??**). B cells receive the visual input via the term $F_{B_i}(t)$ which depends on the stimulus and on the B cell's receptive field. They are inhibited by A cells via the synaptic weights $W_{B_i}^{A_j} < 0$. A cells are excited by B cells via the synaptic weights $W_{A_j}^{B_i} > 0$. B cells are connected to RG cells via the synaptic weights $W_{G_k}^{B_i} > 0$. A cells are connected to RG cells via the synaptic weights $W_{G_k}^{A_j} < 0$. Note that we do not impose any constraint on the connectivity here.

To study mathematically the dynamical system (**??**) we write it in a more convenient form. We use Greek indices $\alpha, \beta, \gamma = 1\ldots N \equiv N_A + N_B + N_G$, and define the state vector $\vec{\mathcal{X}}$, with entries:

$$\mathcal{X}_\alpha = \left\{\begin{array}{lll} V_{B_i}, & \alpha = i, & i = 1\ldots N_B; \\ V_{A_j}, & \alpha = N_B + j, & j = 1\ldots N_A; \\ V_{G_k}, & \alpha = N_B + N_A + k, & k = 1\ldots N_G. \end{array}\right. \tag{10}$$

We introduce $\vec{\mathcal{F}}(t)$, the non stationary input, with entries:

$$\mathcal{F}_\alpha(t) = \left\{\begin{array}{lll} F_{B_i}(t), & \alpha = i, & i = 1\ldots N_B; \\ 0, & \alpha > N_B; & \end{array}\right.$$

and $\vec{\mathcal{R}}(\vec{\mathcal{X}})$, the rectification term, with entries:

$$\mathcal{R}_\alpha(\vec{\mathcal{X}}) = \left\{\begin{array}{lll} \mathcal{N}_B\left(V_{B_i}\right), & \alpha = i, & i = 1\ldots N_B; \\ \mathcal{N}_A\left(V_{A_j}\right), & \alpha = N_B + j, & j = 1\ldots N_A; \\ 0, & \alpha = N_B + N_A + k, & k = 1\ldots N_G. \end{array}\right.$$

We use the notation $0_{n_1 n_2}$ for the $n_1 \times n_2$ matrix with zero entries. We introduce the $N \times N$ matrices:

$$\mathcal{T} = \begin{pmatrix} -\operatorname{diag}\left[\,\tau_{B_i}\,\right]_{i=1\ldots N_B} & 0_{N_B N_A} & 0_{N_B N_G} \\ 0_{N_A N_B} & -\operatorname{diag}\left[\,\tau_{A_j}\,\right]_{j=1\ldots N_A} & 0_{N_A N_G} \\ 0_{N_G N_B} & 0_{N_G N_A} & -\operatorname{diag}\left[\,\tau_{G_k}\,\right]_{k=1\ldots N_G} \end{pmatrix}, \tag{11}$$

characterizing the characteristic integration times of cells,

$$\mathcal{W} = \begin{pmatrix} 0_{N_B N_B} & W_B^A & 0_{N_B N_G} \\ W_A^B & 0_{N_A N_A} & 0_{N_A N_G} \\ W_G^B & W_G^A & 0_{N_G N_G} \end{pmatrix}, \tag{12}$$

summarizing chemical synapses interactions. Note that, to our best knowledge, there are no synapse from RG cells to RG cells, but they could be added in this formalism.

Then, the dynamical system (**??**) reads, in vector form:

$$\frac{d\vec{\mathcal{X}}}{dt} = \mathcal{T}^{-1}.\vec{\mathcal{X}} + \mathcal{W}.\vec{\mathcal{R}}(\vec{\mathcal{X}}) + \vec{\mathcal{F}}(t). \tag{13}$$

We remark that (**??**) has a specific skew-product structure: the dynamics of RG cells is driven by B cells and A cells with no feedback. This means that one can study first the coupled dynamics of B cells and A cells and then the effect on RG cells. This corresponds to a biological reality as, to our best knowledge, there is no feedback from RG cells to B cells or to A cells.

2.1.6 Piecewise linear evolution

We assume here that $\mathcal{F}_\alpha(t)$ is bounded, as well as synaptic weights. Thus, the phase space Ω of (**??**) can be taken compact. Indeed, trajectories cannot escape to infinity thanks to the rectification terms $\mathcal{N}_B, \mathcal{N}_A$, (eq. (**??**)) and thanks to the sign of synaptic weights $W_{A_j}^{B_i}, W_{B_i}^{A_j}$. More precisely, V_{B_i} cannot become arbitrary large and positive because the input term $\mathcal{F}_\alpha(t) \equiv F_{B_i}(t)$ is bounded and because $\sum_{j=1}^{N_A} W_{B_i}^{A_j} \mathcal{N}_A\left(V_{A_j}\right) \leq 0$. Assume indeed that V_{B_i} increases (due to a large enough $\mathcal{F}_\alpha > 0$ making the r.h.s. of eq. (**??**) positive). This leads to an increase of connected A cells voltages V_{A_j} (eq. (**??**)), thus to a decrease of the term $\sum_{j=1}^{N_A} W_{B_i}^{A_j} \mathcal{N}_A\left(V_{A_j}\right) \leq 0$ until the point where the r.h.s. of (**??**) becomes negative, thereby decreasing V_{B_i} and preventing it from becoming arbitrary large. This, implies as well that V_{A_j}s cannot become arbitrary large. On the opposite, if V_{B_i} (resp. V_{A_j}) becomes smaller than θ_B (resp. θ_A) it does not play any more role in the dynamics because of rectification.

Due to the specific form (**??**) of the rectification terms, the dynamical system (**??**) is piecewise linear. More precisely, we can partition the phase space Ω into sub domains $\Omega^{(n)}$, $n = 1 \dots 2^{N_B+N_A}$ defined as follow. To each cell $\alpha = 1 \dots N_B + N_A$ (B cell or A cell) we associate a "rectification label" $\eta_\alpha = 1$ if the cell α is rectified and $\eta_\alpha = 0$ otherwise. Because of the form (**??**) of the rectification, the label η_α corresponds to a partition of the voltage X_α's domain of variation into two sub domains (e.g., for a B cell, $\eta_\alpha = 1$ if $V_{B_i} < \theta_B$ and $\eta_\alpha = 0$ if $V_{B_i} \geq \theta_B$). Now, the set $\{0,1\}^{N_B+N_A}$ is made of chains $\eta = (\eta_1 \dots \eta_{N_B+N_A})$ composed of the rectification labels η_α of all B cells and A cells. To each such sequence is therefore associated a convex domain $\Gamma^{(n)}$ of $\mathbb{R}^{N_B+N_A}$ where all cells α such that $\eta_\alpha = 0$ have their voltage X_α larger

than the rectification threshold, thus, are not rectified, and all cells such that $\eta_\alpha = 1$ are rectified. To each such η is associated a unique integer (e.g. $n = \sum_{\alpha=1}^{N_B+N_A} \eta_\alpha 2^{\alpha-1}$, η is then the binary coding of n). Finally, we set $\Omega^{(n)} = \Gamma^{(n)} \times \mathbb{R}^{N_G}$, where the product with the subspace $\mathbb{R}^{N_G}$ integrates the states space of RG cells dynamics. They are slaved by B cells and A cells dynamics, but they are not rectified. In this setting, $\Omega^{(0)}$ is the subset of Ω such that neither B cells nor A cells are rectified; $\Omega^{(1)}$ the subset of the phase space where only B cell 1 is rectified; $\Omega^{(3)}$ the subset where only B cells $1, 2$ are rectified; $\Omega^{(2^{N_B})}$ the subset where only A cell 1 is rectified and so on.

It is easy to check that the sets $\Omega^{(n)}$ are disjoint and cover $\mathbb{R}^N$, thus, make a partition of the phase space. The vector $\vec{\mathcal{R}}(\vec{\mathcal{X}})$ has now the form:

$$\mathcal{R}_\alpha(\vec{\mathcal{X}}) = \begin{cases} (1-\eta_\alpha)\,(\,X_\alpha - \theta_B\,), & \alpha = 1 \dots N_B; \\ (1-\eta_\alpha)\,(\,X_\alpha - \theta_A\,), & \alpha = N_B \dots N_B + N_A; \\ 0, & \alpha = N_B + N_A + k, \qquad k = 1 \dots N_G, \end{cases}$$

and is piecewise-linear in $\vec{\mathcal{X}}$. For $\vec{\mathcal{X}} \in \Omega^{(n)}$, the transformation $\mathcal{T}^{-1}.\vec{\mathcal{X}} + \mathcal{W}.\vec{\mathcal{R}}(\vec{\mathcal{X}})$ can therefore be written $\mathcal{L}^{(n)}.\vec{\mathcal{X}} + \vec{\mathcal{C}}^{(n)}$, where $\vec{\mathcal{C}}^{(n)}$ is the vector with entries:

$$\vec{\mathcal{C}}^{(n)} = \begin{cases} -\theta_B\,(1-\eta_\alpha)\,\sum_j W_{B_\alpha}^{A_j}, & \alpha = 1 \dots N_B; \\ -\theta_A\,(1-\eta_\alpha)\,\sum_i W_{A_\alpha}^{B_i}, & \alpha = N_B \dots N_B + N_A; \\ 0, & \alpha = N_B + N_A + 1, \dots, N. \end{cases} \tag{14}$$

This is a time-constant vector, coming from the presence of a threshold in rectification (it is zero when $\theta_A, \theta_B = 0$), depending on the rectification state of cells, thus depending on the domain $\Omega^{(n)}$. Rectified cells have zero entries in $\vec{\mathcal{C}}^{(n)}$. The matrix:

$$\mathcal{L}^{(n)} = \begin{pmatrix} -\operatorname{diag}\left[\,\frac{1}{\tau_{B_i}}\,\right]_{i=1\dots N_B} & W_B^A.D_A^{(n)} & 0_{N_B N_G} \\ W_A^B.D_B^{(n)} & -\operatorname{diag}\left[\,\frac{1}{\tau_{A_j}}\,\right]_{j=1\dots N_A} & 0_{N_A N_G} \\ W_G^B.D_B^{(n)} & W_G^A.D_A^{(n)} & -\operatorname{diag}\left[\,\frac{1}{\tau_{G_k}}\,\right]_{k=1\dots N_G} \end{pmatrix}, \tag{15}$$

is called the *transport operator in the domain* $\Omega^{(n)}$. This terminology is further explained in section **??**, but, in short, $\mathcal{L}^{(n)}$ acts as a flow (or a propagator) characterizing the evolution of a trajectory within $\Omega^{(n)}$. In eq. (**??**), the matrices $D_B^{(n)} = \operatorname{diag}\,[\,1-\eta_\alpha\,]_{\alpha=1\dots N_B}$, $D_A^{(n)} = \operatorname{diag}\,[\,1-\eta_\alpha\,]_{\alpha=N_B+1\dots N_B+N_A}$ are projecting onto the subspace of non rectified cells in the domain $\Omega^{(n)}$. In other words, when the state $\vec{\mathcal{X}}$ is in $\Omega^{(n)}$, a rectified cell α gives a zero contribution to the dynamics of other cells, which corresponds to have a row and column α made of zeros in $D_A^{(n)}, D_B^{(n)}$.

The dynamical system (**??**) reads now:

$$\frac{d\vec{\mathcal{X}}}{dt} = \mathcal{L}^{(n)}.\vec{\mathcal{X}} + \vec{\mathcal{F}}^{(n)}(t), \quad \vec{\mathcal{X}} \in \Omega^{(n)}, \tag{16}$$

where we wrote $\vec{\mathcal{F}}^{(n)}(t) = \vec{\mathcal{C}}^{(n)} + \vec{\mathcal{F}}(t)$. Thanks to the decomposition of the phase space into convex sub-domains $\Omega^{(n)}$, (**??**) is now linear. This technique of phase space decomposition is classical and has been used in domains such as ergodic theory and billiards, self-organized criticality [**?**, **?**] or neurosciences [**?**, **?**, **?**, **?**]. See especially the recent paper by A. Rajakumar et al [**?**] very much in the spirit of the present model.

2.1.7 Spectra and fixed points

It is important to consider in detail the spectrum[2] of $\mathcal{L}^{(n)}$ for further studies. We note $\lambda_\beta^{(n)}, \beta = 1 \dots N$, the eigenvalues of $\mathcal{L}^{(n)}$ and its right eigenvectors are noted, $\mathcal{P}_\beta^{(n)}$. These vectors are the columns of the matrix $\mathcal{P}^{(n)}$ transforming $\mathcal{L}^{(n)}$ in diagonal form (assuming it is diagonalizable). $\left(\mathcal{P}^{(n)}\right)^{-1}$ is the inverse matrix. Its rows are the left eigenvectors of $\mathcal{L}^{(n)}$.

As $D_A^{(n)}, D_B^{(n)}$ are projections matrices, it is easy to see, from the form (**??**), that a rectified cell generates an eigenvalue $-\frac{1}{\tau_\alpha}$ and an eigenvector $\vec{e}_\alpha$, the canonical basis vector of $\mathbb{R}^N$ in the direction α. The non rectified cells span a subspace of $\mathbb{R}^N$ and the projection of $\mathcal{L}^{(n)}$ on this subspace has a spectrum depending on the connectivity matrices W_A^B, W_B^A and on other parameters like characteristic times.

The corresponding eigenvalues $\lambda_\beta^{(n)}, \beta = 1 \dots N$ can be real or complex, with a positive or a negative real part. In the case where W_A^B and W_B^A commute it is actually possible to explicitly compute the eigenvalues and the eigenvectors and obtain conditions for stability (all eigenvalues have real negative part) and real/complex eigenvalues [**?**]. If we further assume that W_A^B and W_B^A have no zero eigenvalues, the sign constraints on these matrices implies that $\mathcal{L}^{(n)}$ is invertible for all n. This is what we are going to assume from now.

It follows that, in the absence of external stimulus ($\vec{\mathcal{F}}(t) = \vec{0}$), equation (**??**) has, for each n, a unique fixed point $\vec{\mathcal{X}}^{(n)} = -\left(\mathcal{L}^{(n)}\right)^{-1}.\vec{\mathcal{C}}^{(n)}$. Note, however, that this point *may not be* in $\Omega^{(n)}$. This is a typical situation for piecewise linear dynamical systems (like Iterated Function Systems [**?**, **?**, **?**]) where dynamics can have complex attractors even if maps are linear (and contracting) into sub-domains of the phase space. The simplest non trivial case is when dynamics generates a periodic orbit, but more complex attractors (fractal sets) can be obtained. Here, it is reasonable to assume, at least, that cells at rest are not rectified. Mathematically, this means that the fixed point of $\mathcal{L}^{(0)}$, $\vec{\mathcal{X}}^* = -\left(\mathcal{L}^{(0)}\right)^{-1}.\vec{\mathcal{C}}^{(0)}$, belongs to $\Omega^{(0)}$ and this is what we are going to assume for now. This imposes a set of constraints linking synaptic weights and thresholds. A simple assumption consists of having vanishing thresholds $\theta_A = \theta_B = 0$, in which case the rest state is $\vec{0}$. We will also assume that $\vec{\mathcal{X}}^*$ is stable (eigenvalues of $\mathcal{L}^{(0)}$ have a negative real part), which imposes additional assumptions on synaptic weights and cells integration times. On biophysical

[2] Another approach consists of considering the Schur decomposition instead of the diagonalisation [**?**, **?**, **?**].

grounds it means that the rest state is stable to small perturbations, like noise. Because rectified cells produce stable eigenvalues the following holds. Taking an initial condition in any domain $\Omega^{(n)}$ spontaneous dynamics (without stimulus) eventually drives the trajectory back to $\Omega^{(0)}$ and, then, to the rest state. This is further commented below (section **??**, remark 2).

2.1.8 Extensions: gain control and gap junctions

Gap junctions. Electric synapses, e.g. between B cells and A cells, play an important role in the retina, for example in the rod-cone pathway [**?**]. We consider here passive gap junctions corresponding to electric synapses with a constant conductance (in contrast to conductances depending on variables such as light illumination, see https://webvision.med.utah.edu/book/part-iii-retinal-circuits/myriad-roles-for-gap-junctions-in-retinal-circuits/). Let us consider, for example, a gap junction between B cell i and A cell j. We note $g_{B_i}^{A_j} \geq 0$ the electric conductance from j to i (with $g_{B_i}^{A_j} = 0$ if there is no electric connection between the two cells). As gap junctions are symmetric $g_{B_i}^{A_j} = g_{A_j}^{B_i}$. We also note C_{B_i} the membrane capacitance of B cell i and C_{A_j} the membrane capacitance of A cell j and introduce the notation $G_{B_i}^{A_j} = \frac{g_{B_i}^{A_j}}{C_{B_i}}$, $G_{A_j}^{B_i} = \frac{g_{A_j}^{B_i}}{C_{A_j}}$. Remark therefore that $G_{B_i}^{A_j} = G_{A_j}^{B_i}$ if and only if B cell i and A cell j have the same capacitance. The electric synapse generates a (signed) current $-G_{B_i}^{A_j}\left(V_{B_i} - V_{A_j}\right)$ feeding B cell i and a current $-G_{A_j}^{B_i}\left(V_{A_j} - V_{B_i}\right)$ feeding A cell j. Note that, in contrast to chemical synapses, voltages are not rectified, ionic currents are simply following the gradients of electric potentials and can, therefore, go both ways.

The presence of electric synapses between B cells and A cells modifies therefore equations (**??**) as:

$$\begin{cases} \frac{dV_{B_i}}{dt} &= -\frac{1}{\tau'_{B_i}} V_{B_i} + \sum_{j=1}^{N_A} \left[W_{B_i}^{A_j} \mathcal{N}_A\left(V_{A_j}\right) + G_{B_i}^{A_j} V_{A_j} \right] + F_{B_i}(t), \quad i = 1 \dots N_B; \\ \frac{dV_{A_j}}{dt} &= -\frac{1}{\tau'_{A_j}} V_{A_j} + \sum_{i=1}^{N_B} \left[W_{A_j}^{B_i} \mathcal{N}_B\left(V_{B_i}\right) + G_{A_j}^{B_i} V_{B_i} \right], \qquad\qquad j = 1 \dots N_A; \end{cases}$$

where $\frac{1}{\tau'_{B_i}} = \frac{1}{\tau_B} + \frac{1}{C_{B_i}} \sum_{j=1}^{N_A} g_{B_i}^{A_j}$, $\frac{1}{\tau'_{A_j}} = \frac{1}{\tau_A} + \frac{1}{C_{A_j}} \sum_{i=1}^{N_B} g_{A_j}^{B_i}$ are inverse of characteristic time. Thus, gap junctions have the effect of reducing the characteristic time of cells response (increase their conductance). Gap junctions between A cells and RG cells, or between RG cells would be implemented the same way.

Gain control. This mechanism plays a prominent role in the nervous system. In short this is the property that neural systems have to adjust the non-linear transfer function relating input to output to dynamically span the varying range of incoming stimuli [**?**]. It has been reported in the retina and invoked in several motion processing features: anticipation, alert response to motion onset and motion reversal [**?**, **?**]. In particular, B cells have gain control. Here, this is a

desensitization when activated by a steady illumination [**?**], mediated by a rise in intracellular calcium Ca^{2+}, at the origin of a feedback inhibition preventing thus prolonged signalling of the ON B cell [**?**, **?**]. It can be modelled as follows [**?**, **?**, **?**]. Each B cell has a dimensionless activity variable a_{B_i} obeying the differential equation:

$$\frac{da_{B_i}}{dt} = -\frac{a_{B_i}}{\tau_a} + h_B \, \mathcal{N}_B \left(V_{B_i} \right). \tag{17}$$

The gain function is a strongly non-linear function, almost step-wise, of the form:

$$\mathcal{G}_B(a_{B_i}) = \begin{cases} 0, & \text{if } a_{B_i} \leq 0; \\ \frac{1}{1+a_{B_i}^6}, & \text{else.} \end{cases} \tag{18}$$

Gain control acts at the level of synaptic transmission from B cells to A cells, where the rectification term $\mathcal{N}_B \left(V_{B_i} \right)$ is replaced by $\mathcal{N}_B \left(V_{B_i} \right) \mathcal{G}_B(a_{B_i})$. That is the equation ruling the A cell j's voltage reads now:

$$\frac{dV_{A_j}}{dt} = -\frac{1}{\tau_A} V_{A_j} + \sum_{i=1}^{N_B} W_{A_j}^{B_i} \, \mathcal{N}_B \left(V_{B_i} \right) \mathcal{G}_B(a_{B_i}).$$

It has the following meaning. When the voltage of B cell i increases, its activity a_{B_i} increases as well, up to a range where gain control takes place. When a_{B_i} becomes too large $\mathcal{G}_B(a_{B_i})$ drops down thereby reducing the action of B cell i on A cell j. As mentioned earlier, this is a way to rectify voltages from above. Gain control has also been reported for (OFF) RG cells [**?**, **?**] and shape their firing rate. Gain control at the level of B cells and RG cells, induces retinal anticipation. When combined with A cells lateral connectivity or gap junctions connectivity it results in a wave of activity ahead of the propagating stimulus (e.g. a moving bar) for specific ranges of parameters (characteristic times of cells response, weight intensities) as studied in [**?**].

Piecewise linear system with gain control and gap junctions. Here, we want to expose how the piecewise linear formalism developed above can be applied in the case of gain control and gap junctions. Note that gap junctions actually do not pose any problem from this perspective because they add linear contributions. In the presence of gain control and gap junctions the dynamical system (**??**) becomes:

$$\begin{cases} \frac{dV_{B_i}}{dt} &= -\frac{1}{\tau'_{B_i}} V_{B_i} + \sum_{j=1}^{N_A} \left[W_{B_i}^{A_j} \, \mathcal{N}_A \left(V_{A_j} \right) + G_{B_i}^{A_j} \, V_{A_j} \right] + F_{B_i}(t), & i = 1 \dots N_B; \\ \frac{dV_{A_j}}{dt} &= -\frac{1}{\tau'_{A_j}} V_{A_j} + \sum_{i=1}^{N_B} \left[W_{A_j}^{B_i} \, \mathcal{N}_B \left(V_{B_i} \right) \mathcal{G}_B(a_{B_i}) + G_{A_j}^{B_i} \, V_{B_i} \right], & j = 1 \dots N_A; \\ \frac{dV_{G_k}}{dt} &= -\frac{1}{\tau_G} V_{G_k} + \sum_{i=1}^{N_B} W_{G_k}^{B_i} \mathcal{N}_B(V_{B_i}) \, \mathcal{G}_B(a_{B_i}) + \sum_{j=1}^{N_A} W_{G_k}^{A_j} \mathcal{N}_A(V_{A_j}), & k = 1 \dots N_G; \\ \frac{da_{B_i}}{dt} &= -\frac{a_{B_i}}{\tau_a} + h_B \, \mathcal{N}_B \left(V_{B_i} \right), & i = 1 \dots N_B; \end{cases} \tag{19}$$

We do not know about experimental evidences of gain control in A cells, that's why A cells are not gain controlled in (**??**), but the extension is straightforward. RG cells are gain controlled at the level of their firing rate (see [**?**]).

To make (**??**) a piecewise linear dynamical system, the trick is to replace the function (**??**) by a step function where $\mathcal{G}_B(a_{B_i}) = 1$ if $a_{B_i} \in [\,0, \theta_a\,]$, where θ_a is a threshold (typically $\frac{2}{3}$ coming from a linear interpolation of (**??**), see [**?**]) and $\mathcal{G}_B(a_{B_i}) = 0$ otherwise. In addition to the rectification variables η_α we introduce gain control variables $g_\alpha = 1$ if $a_{B_i} \in [\,0, \theta_a\,]$ and $g_\alpha = 0$ otherwise, $\alpha = 1 \dots N_B$. The definition of the domains $\Omega^{(n)}$ extends easily in this context by partitioning $\mathbb{R}^{N+N_B}$ into sub-domains taking the product of the voltage partition $\{\,]-\infty, \theta_B],]\theta_B, +\infty]\,\}$ with the activity partition $\{\,]-\infty, \theta_a],]\theta_a, +\infty]\,\}$. The transport operator generalizes to:

$$\mathcal{L}^{(n)} = \begin{pmatrix} -\operatorname{diag}\left[\frac{1}{\tau'_{B_i}}\right]_{i=1\dots N_B} & W_B^A.D_A^{(n)} + G_B^A & 0_{N_B N_G} & 0_{N_B N_B} \\ W_A^B.D_B^{'(n)} + G_A^B & -\operatorname{diag}\left[\frac{1}{\tau'_{A_j}}\right]_{j=1\dots N_A} & 0_{N_A N_G} & 0_{N_A N_B} \\ W_G^B.D_B^{'(n)} & W_G^A.D_A^{(n)} & -\operatorname{diag}\left[\frac{1}{\tau_{G_k}}\right]_{k=1\dots N_G} & 0_{N_G N_B} \\ h_B\,\mathcal{I}_{N_B N_B} & 0_{N_B N_A} & 0_{N_B N_G} & -\operatorname{diag}\left[\frac{1}{\tau_a}\right]_{i=1\dots N_B} \end{pmatrix}, \tag{20}$$

where $\mathcal{I}_{N_B N_B}$ the N_B-dimensional identity matrix and $D_B^{'(n)} = \operatorname{diag}\left[\,(1-\eta_\alpha)\, g_\alpha\,\right]_{\alpha=1\dots N_B}$. Thus, $D_B^{'(n)}$ has zero entries whenever a B cell is either rectified ($\eta_\alpha = 1$) or gain controlled ($g_\alpha = 0$) leading to a projection on the subspace of B cells which are neither rectified nor gain controlled. Extending the phase space with activity variables corresponds to adding N_B eigenvalues $-\frac{1}{\tau_a}$ to the spectrum. The corresponding eigenvectors are generalized eigenvectors though because the activities variables add a Jordan block to the matrix [**?**].

2.1.9 Solutions

We now consider the general situation where dynamics is in the rest state at times $t < 0$, and, from time $t = 0$ on, the stimulus $\mathcal{S}(x, y, t)$ is applied, resulting in a non stationary drive $\vec{\mathcal{F}}(t)$. In general, the stimulus is applied over a finite time. After this the system eventually returns to rest. Under this stimulation the trajectory $\left\{\,\vec{\mathcal{X}}(t)\,\right\}_{t\geq 0}$ is going to cross a sequence of domains $\Omega^{(n_k)}$, $k = 1, \dots,$ with $n_1 = 0$, entirely determined by the stimulus and the network characteristics. Call $t_-^{(n_{k+1})}$ the time where the trajectory enters the domain $\Omega^{(n_{k+1})}$ and $t_+^{(n_{k+1})}$ the time where it gets out. Note that $t_-^{(n_{k+1})} = t_+^{(n_k)}$. By direct integration of eq. (**??**), we have:

$$\vec{\mathcal{X}}(t) = e^{\mathcal{L}^{(n_{k+1})}(t - t_-^{(n_{k+1})})}.\vec{\mathcal{X}}(t_-^{(n_{k+1})}) + \int_{t_-^{(n_{k+1})}}^{t} e^{\mathcal{L}^{(n_{k+1})}(t-s)}.\vec{\mathcal{F}}^{(n_{k+1})}(s)\, ds, \quad t \in \left[\, t_-^{(n_{k+1})}, t_+^{(n_{k+1})} \,\right], \tag{21}$$

where $\vec{\mathcal{X}}(t_-^{(n_{k+1})})$, corresponding to the state of $\vec{\mathcal{X}}$ when entering $\Omega^{(n_{k+1})}$, is given by the integration of the past trajectory and can be computer explicitly.

This is:

$$\vec{\mathcal{X}}(t_-^{(n_{k+1})}) = \vec{\mathcal{X}}(t_+^{(n_k)}) = \sum_{m=0}^{k} \mathcal{H}_m^k \, \vec{\Phi}_m, \tag{22}$$

where $\mathcal{H}_m^k$ is a sequence of matrices satisfying:

$$\mathcal{H}_k^k = \mathcal{I}_N; \quad \mathcal{H}_m^k = \mathcal{H}_{k-1}^k \, \mathcal{H}_m^{k-1}; \quad \mathcal{H}_{k-1}^k = e^{\mathcal{L}^{(n_k)}(t_+^{(n_k)} - t_+^{(n_{k-1})})}, \tag{23}$$

where $\mathcal{I}_N$ is the identity matrix of dimension N. The matrix $\mathcal{H}_m^k$ transports the flow from the exit point of $\Omega^{(n_m)}$ to the exit point of $\Omega^{(n_k)}$. The vectors $\vec{\Phi}_m$ are defined by:

$$\vec{\Phi}_0 = \vec{\mathcal{X}}(0); \quad \vec{\Phi}_m = \int_{t_-^{(n_m)}}^{t_+^{(n_m)}} e^{\mathcal{L}^{(m)}(t_+^{(n_m)} - s)} . \vec{\mathcal{F}}^{(m)}(s) \, ds. \tag{24}$$

The proof of (**??**) is easily done by recurrence.

2.1.10 Remarks

Let us now make some remarks on the structure of these solutions.

1. The interpretation of (**??**) is the following. Starting from an initial condition $\vec{\mathcal{X}}(0) \in \Omega^{(n_1)}$ the dynamics (**??**) is integrated up to the possible time $t = t_-^{(n_2)} = t_+^{(n_1)}$ when $\vec{\mathcal{X}}(t)$ gets out of $\Omega^{(n_1)}$ and enters a new domain $\Omega^{(n_2)}$. This arises if, during the time evolution of the system, some cells get rectified (or gain controlled) at time t. Then, there is a drastic change in time evolution because rectified cells do not participate any more to dynamics. The value of the state vector at this time is $\vec{\mathcal{X}}(t_+^{(n_1)}) = e^{\mathcal{L}^{(n_1)}(t_+^{(n_1)} - t_-^{(n_1)})} . \vec{\mathcal{X}}(0) + \int_{t_-^{(n_1)}}^{t_+^{(n_1)}} e^{\mathcal{L}^{(n_1)}(t-s)} . \vec{\mathcal{F}}^{(n_1)}(s) \, ds$ which can be written $\vec{\mathcal{X}}(t_+^{(n_1)}) = \mathcal{H}_0^1 . \vec{\Phi}_0 + \mathcal{H}_1^1 . \vec{\Phi}_1 = \sum_{m=0}^{1} \mathcal{H}_m^1 \, \vec{\Phi}_m$ using $t_-^{(n_1)} = 0$. The system is now in the domain $\Omega^{(n_2)}$ and follows its evolution until the (possible) time $t_-^{(n_3)} = t_+^{(n_2)}$ when some new cells are rectified or some rectified cells become non rectified. The system enters a new domain $\Omega^{(n_3)}$ and so on. In general, the state at the entrance of domain $\Omega^{(n_{k+1})}$ is given by (**??**). This is a linear combination of terms $\mathcal{H}_m^k \, \vec{\Phi}_m$ where $\vec{\Phi}_m$ (eq. (**??**)) integrates the stimulus contribution from the entrance time into domain $\Omega^{(n_m)}$ up to the exit time of this domain and $\mathcal{H}_m^k$ transports the state from the exit point of $\Omega^{(n_m)}$ to the exit point of $\Omega^{(n_k)}$.

2. In the definition of $\mathcal{H}_m^k$ the operators $\mathcal{L}_{n_k}$ do not commute in general.

3. Eigenvalues of some $\mathcal{H}_m^k$ can have a positive real part leading to exponential increase along the corresponding eigen-direction. This means that some cells voltage increases exponentially in absolute value. However, when voltages become too large, voltage rectification (or gain control)

takes place, corresponding to the trajectory entering a new continuity domain. Here, unstable cells do not contribute any more to dynamics which is projected on the subspace of non rectified cells. This has the effect of transforming unstable eigenvalues into stable ones preventing the trajectories $\vec{\mathcal{X}}(t)$ to diverge. Actually, the spectrum of $\mathcal{H}_m^k$, controlling stability, resembles the Lyapunov spectrum in ergodic theory [**?**], with two main differences. First, we are simply considering product of matrices without multiplying by the adjoin so that eigenvalues can be complex. Second, we are not assuming stationarity and the existence of an invariant measure. Instead, the product $\mathcal{H}_m^k$ is constrained by the non stationary stimulus and dynamical system parameters which fixes the sequence of times n_ks.

4. Rectification induces a weak form of non linearity where e.g. the contraction/expansion in the phase space depends on the domain $\Omega^{(n_k)}$ (whereas in a differentiable non linear system it would depend on the point in the phase space). This has deep consequences on cells response, as commented in the results sections.

2.2 Spike statistics

As pointed out in the introduction, it might be helpful to propose a mathematical setting taking into account non stationarity and potentially long memory in spike trains probabilities. Such a setting exists since long but has not been applied to spike train statistics until recently. It is inherited from statistical physics on one hand [**?**] and on extensions of Markov chains to unbounded memory on the other hand [**?**] . The material briefly sketched here has been published in [**?**, **?**, **?**, **?**, **?**, **?**].

2.2.1 Mathematical setting for spike trains

Neurons variables such as membrane potential or ionic currents are described by continuous-time equations. In contrast, spikes resulting from the experimental observation are discrete events, binned with a certain time resolution δ, say of order a millisecond. We consider a network of N spiking neurons, labelled with an index $k = 1 \dots N$. We define a spike variable $\omega_k(n) = 1$ if neuron k has emitted a spike in the time interval $[n\delta, (n+1)\delta[$, and $\omega_k(n) = 0$ otherwise. We denote by $\omega(n) = [\,\omega_k(n)\,]_{k=1}^N$ the spike-state of the entire network at time n, which we call a *spiking pattern*. A *spike block* denoted by ω_m^n, $n \geq m$, is the sequence of spike patterns $\omega(m), \omega(m+1) \dots \omega(n)$. The range of a block ω_m^n is $n - m + 1$, the number of time steps from m to n. We call a spike train an infinite sequence of spikes both in the past and in the future, and, to simplify notations we note a spike train ω (instead of $\omega_{-\infty}^{+\infty}$). Of course, on operational grounds spike trains are finite, but it is mathematically more convenient to work on a space of bi-infinite spike sequences.

2.2.2 Mathematical setting for spiking probabilities

We now consider a family of transition probabilities of the form $\mathbb{P}_n\left[\,\omega(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right]$, which represent the probability, that at time n, one observes the spiking pattern $\omega(n)$ given the network spike history, extending to an infinite past. This is an extension of Markov chains where probabilities have the form $\mathbb{P}_n\left[\,\omega(n)\,\middle|\,\omega_{n-D}^{n}\,\right]$, where D is the memory depth of the Markov chain. Letting the memory be possibly infinite corresponds to situation where one cannot precisely fix the memory depth necessary to characterize the probability of a spike pattern given the past spike history. An example of a model requiring this context is presented in section **??** below. Having infinite memory imposes mathematical constraints on the memory decay that has to be sufficiently fast (typically, exponential) so that the situation is close to Markov chains. In addition to the model presented below, neural models with infinite memories have been considered by several authors such as E. Loecherbach and A. Galves [**?**]. A few remarks about this form of probability:

1. We do not assume stationarity. $\mathbb{P}_n$ may depend explicitly on time. This is actually the reason why we have an index n. A time translation invariant probability will simply be written $\mathbb{P}$.

2. For such probabilities to be well defined and useful, one need to make assumptions on their structure. Beyond technical assumptions such as measurability, summability, non nullness and continuity [**?**, **?**], the most important assumption here is that the dependence in the past (memory) decays fast enough, typically, exponentially, so that, even if this chain has infinite memory it is very close to Markov.

3. As one can associate to Markov chains an equilibrium probability (under conditions actually quite more general than detailed-balance) the system of transition probabilities $\{\mathbb{P}_n\}_{n\in\mathbb{Z}}$ also admits, under the mathematical conditions sketched in the item 2 above, an equivalent notion called "chains with complete connections" or a "chain with unbounded memory" [**?**].

4. These distributions are formally (left-sided) Gibbs distributions where the Gibbs potential is $\Phi(n,\omega) \stackrel{\text{def}}{=} \log \mathbb{P}_n\left[\,\omega(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right]$ (the non-nullness assumption imposes that $\mathbb{P}_n\left[\,\omega(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right] > 0$). This establishes a formal link to statistical physics. In particular, when the chain is stationary, expanding the potential in product of spikes events up to second order one recovers the maximum entropy models used in the literature of spike trains analysis, including the so-called Ising model [**?**, **?**, **?**, **?**]. However, the chains we consider are not necessarily stationary.

2.2.3 A model of effective interactions between RG cells

The visual cortex has no clue on which biophysical processes are taking place in the retina. All the visual information it receives is encoded in spike trains.

This leads to the idea of proposing models of spiking RG cells network where dynamics of RG cells voltage is only constrained by RG cells spikes history. Here, one assumes that RG cells dynamics is controlled by the interactions with hidden layers, for example, the B cells-A cells layers in the model (**??**), in a situation where an observer is just recording the spikes emitted by RG cells, while having no clue of the dynamics in the upper layers. These hidden layers result in providing *effective interactions* between RG cells that one can interpolate by fitting the statistics. The idea is then to construct a dynamical model where the spiking of a RG cell depends on the spike history emitted by the network, with virtual interactions that mimic hidden causal effects [**?**]. This strategy lead us to propose the model presented in the next paragraph. The advantage of this approach is that one can explicitly write the transition probabilities $\mathbb{P}_n\left[\,\omega(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right] > 0$ and infer, from this, a linear response formula telling us how statistical quantities such as firing rates, but also spike correlations are modified by a time dependent stimulus. These results are presented in the "Results" subsection **??**.

The model is inspired from the generalized Integrate and Fire model (gIF) proposed by Rudolph and Destexhe [**?**] and generalizes the Leaky-Integrate and Fire (LIF) model [**?**, **?**]. We have N neurons (say RG cells) characterized by their voltage $V_k, k = 1 \dots N$. One fixes a voltage threshold θ such that, whenever $V_k(t) = \theta$ a spike is emitted by neuron k at time t, and is reset to a reset value (typically, $V_{reset} = 0$). Below θ, the dynamics of voltage (sub-threshold dynamics) is governed by eq. (**??**) below.

In the LIF model, synaptic conductances are constant. In the gIF model, in contrast, the synaptic conductance g_{kj} between the pre-synaptic neuron j and the post-synaptic neuron k depends on spike history as:

$$g_{kj}(t,\omega) = G_{kj}\,\alpha_{kj}(t,\omega), \tag{25}$$

where:

$$\alpha_{kj}(t,\omega) = \sum_{n=-\infty}^{t} \alpha_{kj}(t-n)\,\omega_j(n). \tag{26}$$

The notation $g_{kj}(t,\omega)$ means that function g_{kj} depends on spikes occurring before time t. $G_{kj} \geq 0$ is the maximal conductance between j and k. It is zero when there is no synaptic connection between neurons j and k. In (**??**), the function $\alpha_{kj}(t)$, called α-kernel, summarizes the complex dynamical process underlying the generation of a post-synaptic potential after the emission of a pre-synaptic spike [**?**]. It has the typical form $\alpha_{kj}(t) = P(t)e^{-\frac{t}{\tau_{kj}}}H(t)$ where $P(t)$ is a polynomial in time and $H(t)$ is the Heaviside function. What matters on mathematical grounds is the exponential tail of $\alpha_{kj}(t)$ [**?**]. The function $\alpha_{kj}(t,\omega)$ depends on the spike history preceding t. It records the spikes emitted by the pre-synaptic neuron j before t, corresponding to $\omega_j(n) = 1$ and adds up a contribution $\alpha_{kj}(t-n)$ to the post synaptic conductance from pre-synaptic neuron j to post-synaptic neuron k.

Now, the gIF dynamics reads [**?**, **?**, **?**]:

$$C_k \frac{dV_k}{dt} + g_L(V_k - E_L) + \sum_j g_{kj}(t,\omega)(V_k - E_j) = S_k(t) + \sigma_B \xi_k(t), \quad \text{if } V_k(t) < \theta,$$

where g_L, E_L are respectively the leak conductance and the leak reversal potential, E_j the reversal potential characterizing the synaptic transmission between j and k. Finally, $\xi_k(t)$ is a white noise, introducing stochasticity in dynamics. Its intensity is σ_B.

Setting $W_{kj} = G_{kj}E_j$, $i_k(t,\omega) = g_L E_L + \sum_j W_{kj}\alpha_{kj}(t,\omega) + S_k(t) + \sigma_B \xi_k(t)$, $g_k(t,\omega) = g_L + \sum_{j=1}^{N} g_{kj}(t,\omega)$, one can finally write the gIF dynamics in the form:

$$C_k \frac{dV_k}{dt} + g_k(t,\omega) V_k = i_k(t,\omega), \quad \text{if } V_k(t) < \theta, \tag{27}$$

where $i_k(t,\omega)$ depends, on the network spike history via $\alpha_{kj}(t,\omega)$, on the stimulus, and contains a stochastic term. As the reversal potential E_j can be positive or negative, the synaptic weights W_{kj} define an oriented and signed graph, whose vertices are the neurons. These weights are what we call effective interactions.

What makes the gIF model very rich is that it proposes a biophysically grounded way to construct a dynamical system where the variables (here, voltages) are constrained by the only information of spike train history. The price to pay is that dynamics actually depends on the *whole spike history*, which is potentially infinite. Actually, the gIF model has an infinite memory. This is essentially because the conductance depends on the whole history, and, contrarily to voltages is not reset when the neuron fires. Nevertheless, the exponential decay in the alpha profile actually ensures the existence (and uniqueness) of transition probabilities of the form $\mathbb{P}_n\left[\omega(n) \,\middle|\, \omega_{-\infty}^{n-1}\right]$ [**?**, **?**, **?**, **?**].

Note that the integration of (**??**) does not only requires the knowledge of voltages V_k, stimulus and noise at time t. It requires, in addition, the knowledge of the spike train ω emitted by the network before t. In this sense, this is not a classical dynamical system. Nevertheless, eq. (**??**) can be explicitly integrated [**?**, **?**].

3 Results

3.1 How could lateral A cells connectivity shape the receptive field of a ganglion cell ?

The response of a RG cell to visual stimuli is shaped by the retina structure depicted in Fig. **??**. Here, with the model introduced in section **??**, we would like to characterize the respective effects of the stimulus and of the network connectivity, especially A cells, and understand under which condition can the conjugated effect of network dynamics and stimulus be represented by a convolution of the form (**??**) where the kernel $\mathcal{K}_{G_\alpha}$ is *intrinsic* to the cell, i.e. does not depend on the stimulus ?

3.1.1 Non rectified case

The answer is relatively easy when no rectification takes place, i.e. when the trajectory of (**??**) stays in the domain $\Omega^{(0)}$ (see section **??** for the definition). Indeed, in this case evolution is ruled by equation (**??**) which holds from the initial time $t = t_0$ where the stimulus starts to be applied, to the current time t. Actually, we can consider that t_0 starts far in the past and let it tend to $-\infty$. This corresponds to considering that the stimulus is applied on a time scale quite longer than the characteristic times in the problem (i.e. the inverse of the real part of eigenvalues). Then eq. (**??**) reads $\vec{\mathcal{X}}(t) = \int_{-\infty}^{t} e^{\mathcal{L}^{(0)}(t-s)}.\vec{\mathcal{F}}^{(0)}(s)\,ds$, which is $\vec{\mathcal{X}}(t) = \left[e^{\mathcal{L}^{(0)}} \overset{t}{*} \vec{\mathcal{F}}_0 \right](t)$. This equation actually makes sense only if all eigenvalues of $\mathcal{L}^{(0)}$ are stable, as we assumed above. Note also that $\vec{\mathcal{F}}^{(0)} = \vec{\mathcal{C}}^{(0)} + \vec{\mathcal{F}}$ where $\vec{\mathcal{C}}^{(0)}$ is a constant, depending on thresholds (eq. (**??**)) and whose integration in the convolution product gives $-\left(\mathcal{L}^{(0)}\right)^{-1}.\vec{\mathcal{C}}^{(0)} = \vec{\mathcal{X}}^{*}$, the base line activity of $\vec{\mathcal{X}}(t)$ without stimulus. We may ignore this constant in the sequel and focus on the time varying part of the response, $\left[e^{\mathcal{L}^{(0)}} \overset{t}{*} \vec{\mathcal{F}} \right](t)$. As $\vec{\mathcal{F}}$ is itself defined in terms of a convolution (eq. (**??**)) with the stimulus and its derivative, $\vec{\mathcal{X}}(t)$ is a convolution with the stimulus and its derivative. Here, it is useful to express $\vec{\mathcal{X}}(t)$ in components.

One can then show that [**?**]:

$$\mathcal{X}_\alpha(t) = V_{\alpha_{drive}}(t) + \mathcal{E}^{(0)}_{\alpha_{net}}(t), \quad \alpha = 1 \dots N, \tag{28}$$

where:

$$\mathcal{E}^{(0)}_{\alpha_{net}}(t) = \sum_{\beta=1}^{N} \sum_{\gamma=1}^{N_B} \mathcal{P}^{(0)}_{\alpha\beta} \left(\mathcal{P}^{(0)}_{\beta\gamma} \right)^{-1} \varpi^{(0)}_{\beta\gamma} \int_{-\infty}^{t} e^{\lambda^{(0)}_\beta (t-s)}\, V_{\gamma_{drive}}(s)\,ds, \tag{29}$$

where $\varpi^{(0)}_{\beta\gamma} = \lambda^{(0)}_\beta + \frac{1}{\tau_{B_\gamma}}$. The term $V_{\alpha_{drive}}(t)$ in eq. (**??**) is the stimulus drive and acts only on B cells (it vanishes for $\alpha > N_B$). The term (**??**) contains the network effects. The drive imposed on B cells impacts A cells via the connectivity and, thereby, have a feedback effect on B cells. In addition, the join activity of B cells and A cells drive the RG cells response ($\alpha > N_B + N_A$). In particular, this equation allows to compute explicitly the RF of a RG cell.

For this, we introduce the function $e^{(0)}_\beta(t) \equiv e^{\lambda^{(0)}_\beta\, t}\, H(t)$ so that $\int_{-\infty}^{t} e^{\lambda^{(0)}_\beta (t-s)}\, V_{\gamma_{drive}}(s)\,ds \equiv \left[e^{(0)}_\beta \overset{t}{*} V_{\gamma_{drive}} \right](t)$, which according to (**??**) is $\left[e^{(0)}_\beta \overset{t}{*} \mathcal{K}_{B_\gamma} \overset{x,y,t}{*} \mathcal{S} \right](t)$. Thus, by identification with (**??**), the kernel of RG cell $\alpha = N_B + N_A + 1 \dots N_G$ is:

$$\mathcal{K}_{G_\alpha}(x,y,t) = \sum_{\beta=1}^{N} \sum_{\gamma=1}^{N_B} \mathcal{P}^{(0)}_{\alpha\beta} \left(\mathcal{P}^{(0)}_{\beta\gamma} \right)^{-1} \varpi^{(0)}_{\beta\gamma} \left[e^{(0)}_\beta \overset{t}{*} \mathcal{K}_{B_\gamma} \right]. \tag{30}$$

This provides an explicit equation for the kernel of a RG cell, embedded in a network of B cells, A cells, RG cells with dynamics (**??**), when no rectification take place.

3.1.2 Interpretation

The kernel obtained in (**??**) is the response of the RG cell to a Dirac pulse corresponding, in experiments, to a brief light (or dark) full-field flash. It can also obtained from a white noise stimulus, corresponding, in experiments, to the so-called Spike Triggered Average (STA) [**?**, **?**, **?**]. It corresponds therefore to the functional definition of the receptive field of RG cells used in experiments. In addition, eq. (**??**), (**??**) give us the voltage of *all* cells in the network at time t under the influence of a stimulus. Interestingly, thus, these equations allow us to visualize the join evolution of B cells and A cells as well as their action of RG cells. Note that B cells and A cells are difficult to access experimentally. Given a prescribed connectivity (matrices $W_A^B, W_B^A, W_G^B, W_G^A$), eq. (**??**) provides us, therefore, a mathematical insight on the potential, hidden, dynamics of B cells and A cells leading to the experimentally observed response of RG cells. Thus, this gives us possible scenarios characterizing the potential effects of A cells networks on RG cells response. In addition, eq. (**??**) also provides the RF for B cells ($\alpha = 1 \dots N_B$) and A cells ($\alpha = N_B + 1 \dots N_B + N_A$). We observe in particular that, in a network, the RF of a B cell is therefore not only what comes from the OPL - the term $V_{\alpha_{drive}}(t)$ - it integrates as well lateral A cells connectivity. This is similar to the center-surround shaping of OPL output due to H cells, but here, we might have different effects, due to the different physiology of A cells.

3.1.3 Space-time separability

The RG cell kernel, in general, does not factorise into a product of a function of space and a function of time (separability). Even in the case where the B cells RF is separable, i.e. $\mathcal{K}_{B_\gamma}(x,y,t) = \mathcal{K}_{B_{S_\gamma}}(x,y)\,\mathcal{K}_{B_{T_\gamma}}(t)$ where $\mathcal{K}_{B_{S_\gamma}}$ is the spatial part, centred at x_γ, y_γ and $\mathcal{K}_{B_{T_\gamma}}$ the temporal part, the RG cell kernel reads:

$$\mathcal{K}_{G_\alpha}(x,y,t) = \sum_{\beta=1}^{N} \mathcal{P}^{(0)}_{\alpha\beta} \left(\sum_{\gamma=1}^{N_B} \left(\mathcal{P}^{(0)}_{\beta\gamma} \right)^{-1} \varpi^{(0)}_{\beta\gamma} \left[e^{(0)}_{\beta} \overset{t}{*} \mathcal{K}_{B_{T_\gamma}} \right] \times \mathcal{K}_{B_{S_\gamma}}(x,y) \right), \tag{31}$$

and is not separable either. Now, if B cells have the same temporal kernel $\mathcal{K}_{B_T}$, independent of γ and the same characteristic time τ_B, such that $\varpi^{(0)}_{\beta\gamma} = \lambda^{(0)}_{\beta} + \frac{1}{\tau_B}$ is independent of γ, we can write :

$$\mathcal{K}_{G_\alpha}(x,y,t) = \sum_{\beta=1}^{N} \mathcal{P}^{(0)}_{\alpha\beta}\, \varpi^{(0)}_{\beta} \left[e^{(0)}_{\beta} \overset{t}{*} \mathcal{K}_{B_T} \right] \left(\sum_{\gamma=1}^{N_B} \left(\mathcal{P}^{(0)}_{\beta\gamma} \right)^{-1} \mathcal{K}_{B_{S_\gamma}}(x,y) \right). \tag{32}$$

This kernel is not yet strictly separable as the term $\sum_{\gamma=1}^{N_B} \left(\mathcal{P}^{(0)}_{\beta\gamma} \right)^{-1} \mathcal{K}_{B_{S_\gamma}}(x,y)$ still depends on β, the eigenmode index, via $\left(\mathcal{P}^{(0)}_{\beta\gamma} \right)^{-1}$. Now, the eigenmodes

depends on connectivity. Especially, the B cell to RG cell connectivity corresponds to a pooling of B cells located in the vicinity of RG cell α. The simplest case is when there is no lateral connectivity and where each RG cell α is contacted by only B cell with index γ_α (this implies $N_B = N_G$). In this case: $\mathcal{P}^{(0)}_{\alpha\beta} = \delta_{\alpha\beta}$, $\left(\mathcal{P}^{(0)}_{\beta\gamma}\right)^{-1} = \delta_{\beta\gamma}$ so that $\mathcal{K}_{G_\alpha}(x,y,t) = \varpi^{(0)}_\alpha \left[e^{(0)}_\alpha \overset{t}{*} \mathcal{K}_{B_T}\right] \mathcal{K}_{B_{S_\alpha}}(x,y)$ is separable. More generally, pooling implies that $\mathcal{P}^{(0)}_{\alpha\beta}$ and $\left(\mathcal{P}^{(0)}_{\beta\gamma}\right)^{-1}$ are locally spread around α resulting in a spatial part
$\sum_{\gamma=1}^{N_B} \left(\mathcal{P}^{(0)}_{\beta\gamma}\right)^{-1} \mathcal{K}_{B_{S_\gamma}}(x,y)$ depending only on α.

3.1.4 Resonances

The eigenvalues of $\mathcal{L}^{(0)}$ can be complex, going by conjugated pairs. It is actually quite easy to obtain such a situation mathematically, even considering nearest neighbours interactions [**?**]. A straightforward consequence is the existence of preferred time frequencies (resonance) for a RG cell. In other words, applying periodic sequences of brief flashes with a varying frequency, one might observe a peak in the amplitude of the RG cell response, for specific frequencies. This remark could, e.g., explain the "bump" observed in experiments when the retina is submitted to the so-called "Chirp" stimulus [**?**], a stimulus composed of different phases of flashes stimulation where one varies duration, frequency, and amplitude. In the phase where the amplitude is constant but frequency is varying, some RG cells exhibit a resonance like peak (see e.g. Fig. 1 b in [**?**]). Of course, such resonances could also be explained by intrinsic cells properties, like ion channels response. The potential effect of lateral A cells connectivity would have to be tested experimentally by, e.g. inhibiting A cells synaptic transmission for RG cells exhibiting resonance peaks.

3.1.5 Stimulus induced waves

This is a general fact that networks of coupled units can produce waves. Spontaneous waves are actually reported in the developmental retina, induced, in the so-called stage II and stage III by A cells [**?**]. They are generated by non linear mechanisms and closeness to bifurcations [**?**]. This is not the type of wave we are dealing with here, though. Instead, we are referring to waves triggered by a moving stimulus, say a moving bar. The idea is that such a stimulus can induce, via A cells connectivity, a wave of connectivity which can be *ahead* of the stimulus, for a certain range of parameters (e.g. synaptic coupling intensity) compatible with physiology. Stimulus induced waves, in advance with respect to the stimulus, have been reported in the visual cortex [**?**]. They are due to lateral cortical activity and induce cortical anticipation. The mathematical analysis made in [**?**] suggests that such anticipatory waves could as well exist in the retina thanks to A cells lateral connectivity, conjugated with non linear gain control already known to induce a form of retinal anticipation [**?**, **?**].

3.1.6 Stimulus adaptation

Short term plasticity has been reported in the retina at the synapses between B cells-A cells and A cells-RG cells [**?**, **?**]. Note actually that, although most models of plasticity referring to cortical neurons, are considering spiking neurons [**?**], the physiology of short term synapse adaptation does not necessarily require spikes and is compatible with inner retinal networks dynamics. The effect of synaptic plasticity can be integrated in the model (**??**). It will result in a variations of eigenvalues and eigenvectors of the transport operator $\mathcal{L}$ with potential changes in dynamics. Although, potential and highly relevant phenomena such as bifurcations induced by plasticity would require considering a non linear version of (**??**) (at least, rectification to avoid exponential instability), we can ask about simple linear effect of plasticity on the RG cells response. A straightforward potential effect could be frequency adaptation to periodic flashes.

3.1.7 Rectification.

Let us now investigate the role of rectification. In the general case, a trajectory crosses several domains, and is characterized by eq. (**??**). Starting from the domain $\Omega^{(n_1)}$ (rest state) the state of the network submitted to a stimulus, enters a new domain $\Omega^{(n_2)}$ at time $t_+^{(n_2)}$ where some cells are rectified and so on. Can one still define a response formula of type (**??**) ? This raises several technical difficulties, first because some eigenvalues can be unstable. As we have seen above, this does not lead to an exponential explosion though precisely rectification prevents cells voltage to diverge. Mathematically, this is expressed by the exit of the trajectory from the domain with positive eigenvalue and a projection on the subspace spanned by non-rectified cells. Another difficulty also comes from the constants $\vec{\mathcal{C}}^{(n)}$ defined in (**??**) coming from the threshold in the rectification function. They can be removed by assuming that all thresholds are equal to 0. This is what we are going to do now for the sake of simplicity. One can then define domain-dependent flows $\Phi^{(n)}(\vec{\mathcal{X}},t) \equiv e^{\mathcal{L}^{(n)}\,t}\,\Theta\left(\vec{\mathcal{X}}(t)\in\Omega^{(n)}\right)$, where Θ is the indicator function so that $\left[\Phi^{(n)}(\vec{\mathcal{X}},.)\overset{t}{*}\vec{\mathcal{F}}\right](t) = \sum_{n_m=n}\int_{t_-^{(n_m)}}^{t_+^{(n_m)}} e^{\mathcal{L}^{(m)}(t_+^{(n_m)}-s)}.\vec{\mathcal{F}}(s)\,ds$ where the sum holds on indices n_m in the trajectory such that $n_m = n$. This allows us to express the recurrence formula (**??**) in terms of a convolution and thereby to express the whole trajectory in terms of a convolution with a transport operator.

However, there are several important differences with the non rectified case. First, the kernel defined this way *depends on the trajectory*. As the sequence of domains met by the trajectory (and the time where the trajectory enters in these domains) depend on the stimulus, the RF of rectifiable cells *depends now on the stimulus*. Note that the situation would actually be even worse for non linear cells. Indeed, the question hidden behind these remarks is: "to what extent the *linear response* assumption defining a RF via a convolution equation such as (**??**) is valid". We will actually come back to a similar question in section

?? for a network of spiking neurons. Linear response essentially requires the perturbation to be "weak enough", which in our case, means that cells are not rectified. The formulation in terms of a piecewise linear system allows to extend the notion of RF to rectified cells, but the price to pay is that RF now depends on the stimulus. With respect to biology, this effect would for example mean that cells identified e.g. to be ON with a STA approach, responds differently (e.g. ON-OFF) to a more sophisticated stimulus like the "chirp" stimulus [**?**].

In the rectified cases, the eigenvalues $\lambda_\beta^{(n)}, \beta = 1 \dots N$ and eigenvectors $\mathcal{P}_\beta^{(n)}$ depend on the domain, i.e. on the list of rectified cells and are different from the domain $\Omega^{(0)}$ of the rest state. They actually differ in two ways. First, rectified cells provide eigenvalues $-\frac{1}{\tau_\beta}$ and eigenvectors $\vec{e}_\beta$ so that $\mathcal{P}_{\alpha\beta}^{(n)} = \delta_{\alpha\beta}$ for these cells so that they do not contribute any more to the network response. The second effect is more intricate. Indeed, the mere fact of rectifying one cell, has, in general, the effect of *modifying the whole spectrum and eigenvectors*, with strong effects on the cells response. This can be easily understood. Consider the (not really retinal-realistic) situation where a cell is a hub in a network. Silencing it have in general dramatic effects on the global dynamics of this network.

3.1.8 Conclusion of section ??

In this section, we have given a mathematical answer to the problem 1, level 1, posed in the introduction. On the basis of a simplified model of B cells - A cells - RG cells interactions, we have produced a formalism allowing us to compute this network response to spatio-temporal stimuli. We have been able to write explicitly the RF of individual RG cells appearing in eq. (**??**) where the kernel depends explicitly on lateral connectivity. As we showed, however, the linear response formula (**??**), where the kernel is independent of the stimulus, holds when the stimuli has a weak enough amplitude so that cells are not rectified. As soon as rectification takes place the convolution form (**??**) implies, in general, that the kernel can change with the stimulus. This effect could be observed in experiments if the cell type, characterized via STA, provides a different type of response to other stimuli.

3.2 How could spatio-temporal stimuli correlations and retinal network dynamics shape the spike train correlations at the output of the retina ?

In this section, we extrapolate the previous analysis of the model (**??**) to analyse how spike trains emitted by RG cells can be correlated via the network and especially A cells connectivity. We especially want to make mathematical statements on how could A cells *decorrelate* RG cells, as claimed on the basis of experiments [**?**]. We consider first the non rectified case and then analyse how rectification can modify correlations.

3.2.1 Voltages correlations

We first compute the voltage correlations induced by a non stationary spatio-temporal stimulus in the model (**??**). Note that correlations requires some notion of probability, thus, of randomness. Moreover, it is more convenient when such a probability is stationary, while we want here to consider a non-stationary problem. This is not contradictory though. There are two simple (not incompatible) ways to address this point. First, one may consider that the dynamical system (**??**) has random initial conditions, drawn with respect to a stationary probability measure. Second, one can add to the dynamics (**??**) noise, which always present in biological systems. We can make the assumption that noise is stationary and that it is Brownian (which is a pure mathematical convenience). In biology, spike correlations are usually obtained by averaging over repeats of the same experiment where a stimuli is presented to the retinal network. This corresponds therefore to averaging over initial conditions in the presence of noise. Here, to make things simpler, we assume that initial conditions are deterministic (the network is in the rest state when the stimulus is applied) and randomness is induced by a Brownian noise.

Stimulus induced correlations in the non rectified case. Let us therefore consider a stimulus with the form $\mathcal{S}(x,y,t) = \mathcal{S}_d(x,y,t) + \sigma_S\,\xi(x,y,t)$ where $\mathcal{S}_d(x,y,t)$ is deterministic and $\xi(x,y,t)$ is a spatio-temporal white noise. σ_S controls the intensity of this noise. The spatial integration of B cells RF induces then an obvious correlation between B cells voltages. Consider indeed the term $V_{i_{drive}}(t)$ in eq. (**??**) in the presence of this stimulus. Denoting $\mathbb{E}[\,]$ the expectation with respect to the Wiener measure, we have $\mathbb{E}[\,\xi(x,y,t)\,] = 0$ and $\mathbb{E}[\,\xi(x,y,t)\,\xi(x',y',t')\,] = \delta(x-x')\,\delta(y-y')\,\delta(t-t')$. Then $\mathbb{E}[\,V_{i_{drive}}(t)\,] = \left[\mathcal{K}_{B_i} \overset{x,y,t}{*} \mathcal{S}_d\right](t)$ and the correlation between drives is :

$$\begin{aligned}&\mathbb{E}[\,(\,V_{i_{drive}}(t) - \mathbb{E}[\,V_{i_{drive}}(t)\,]\,)\,(\,V_{j_{drive}}(t') - \mathbb{E}[\,V_{j_{drive}}(t')\,]\,)\,]\\ &= \sigma_S^2 \int_{x=-\infty}^{+\infty}\int_{y=-\infty}^{+\infty}\int_{s=-\infty}^{t} \mathcal{K}_{B_i}(x-x_i, y-y_i, t-s)\,\mathcal{K}_{B_j}(x-x_j, y-y_j, t'-s)\,dx\,dy\,ds,\end{aligned} \tag{33}$$

assuming $t \leq t'$ without loss of generality. We recall that x_i, y_i are the coordinates of the center of BCell i RF. Eq. (**??**) expresses that drives are correlated due to the overlap of B cells RFs, a well known result. Especially, correlations decrease with the distance d between the two RFs center (like e^{-d^2} if RFs are Gaussian).

More generally, the term $F_{B_i}(t)$ in eq. (**??**) has mean:

$$\mathbb{E}[\,F_{B_i}(t)\,] = \left[\left(\frac{1}{\tau_B}\,\mathcal{K}_{B_i} + \frac{\partial}{\partial t}\mathcal{K}_{B_i}\right) \overset{x,y,t}{*} \mathcal{S}_d\right](t),$$

and correlation:

$$\mathcal{C}_{F_{ij}}(t,t') = \sigma_S^2 \left(\begin{array}{l} \frac{1}{\tau_B^2} \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{s=-\infty}^{t} \mathcal{K}_{B_i}(x-x_i, y-y_i, t-s)\, \mathcal{K}_{B_j}(x-x_j, y-y_j, t'-s)\, dx\, dy\, ds \\ + \frac{1}{\tau_B} \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{s=-\infty}^{t} \mathcal{K}_{B_i}(x-x_i, y-y_i, t-s)\, \frac{\partial}{\partial t}\mathcal{K}_{B_j}(x-x_j, y-y_j, t'-s)\, dx\, dy\, ds \\ + \frac{1}{\tau_B} \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{s=-\infty}^{t} \frac{\partial}{\partial t}\mathcal{K}_{B_i}(x-x_i, y-y_i, t-s)\, \mathcal{K}_{B_j}(x-x_j, y-y_j, t'-s)\, dx\, dy\, ds \\ + \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{s=-\infty}^{t} \frac{\partial}{\partial t}\mathcal{K}_{B_i}(x-x_i, y-y_i, t-s)\, \frac{\partial}{\partial t}\mathcal{K}_{B_j}(x-x_j, y-y_j, t'-s)\, dx\, dy\, ds \end{array} \right), \tag{34}$$

for $i, j = 1 \dots N_B$.

This implies that the forcing term $\vec{\mathcal{F}}$ in (**??**) has a $N \times N$ time dependent correlation matrix $\mathcal{C}_{\mathcal{F}}(t,t')$ with a $N_B \times N_B$ block corresponding to (**??**) and the rest of the matrix has zeros (A cells and RG cells have no direct stimulus drive).

Let us now consider the full dynamics (**??**), in the non rectified case: the trajectory stays in $\Omega^{(0)}$. Under the stimulus $\mathcal{S}_d(x,y,t) + \sigma_S\, \xi(x,y,t)$, $\vec{\mathcal{X}}(t)$ is a stochastic process, with mean:

$$\mathbb{E}\left[\, \vec{\mathcal{X}}(t) \,\right] = \left[\, e^{\mathcal{L}^{(0)}} \overset{t}{*} \mathbb{E}\left[\, \vec{\mathcal{F}}^{(0)} \,\right] \right], \tag{35}$$

and correlation matrix:

$$\mathcal{C}_{\vec{\mathcal{X}}}(t,t') = \int_{s=-\infty}^{t} \int_{s'=-\infty}^{t'} e^{\mathcal{L}^{(0)}(t-s)}.\mathcal{C}_{\mathcal{F}}(s,s').\, e^{\tilde{\mathcal{L}}^{(0)}(t'-s')}\, ds\, ds' \tag{36}$$

where $\tilde{\mathcal{L}}^{(0)}$ is the transpose of $\mathcal{L}^{(0)}$. This is the general form of correlations induced by and network. Note that correlations are stationary (they only depend on $t-t'$). This does not hold any more in the rectified case as discussed below.

Correlations structure and decorrelation. Equation (**??**) combines B cells RF overlap (in the matrix $\mathcal{C}_{\mathcal{F}}(s,s')$) to networks effects, A cells and/or gap junctions, via the transfer operator $\mathcal{L}^{(0)}$. One can actually better see these combined effects by projecting on the eigenvectors basis of $\mathcal{L}^{(0)}$, where $\mathcal{L}^{(0)} = \mathcal{P}^{(0)}.\Lambda^{(0)}.{\mathcal{P}^{(0)}}^{-1}$ and $\tilde{\mathcal{L}}^{(0)} = \widetilde{\mathcal{P}^{(0)}}^{-1}.\Lambda^{(0)}.\widetilde{\mathcal{P}^{(0)}}$. Denoting:

$$\Delta_{\mathcal{F}}(s,s') = {\mathcal{P}^{(0)}}^{-1}.\mathcal{C}_{\mathcal{F}}(s,s').\widetilde{\mathcal{P}^{(0)}}^{-1}, \tag{37}$$

eq. (**??**) becomes;

$$\mathcal{C}_{\vec{\mathcal{X}}}(t,t') = \int_{s=-\infty}^{t} \int_{s'=-\infty}^{t'} \mathcal{P}^{(0)}.\, e^{\Lambda^{(0)}(t-s)}\, .\Delta_{\mathcal{F}}(s,s').\, e^{\Lambda^{(0)}(t'-s')}.\, \widetilde{\mathcal{P}^{(0)}}\, ds\, ds',$$

which interprets as follows. Whereas $\mathcal{C}_{\mathcal{F}}(s,s')$ is a rank N_B matrix containing the B cells drives correlations, $\Delta_{\mathcal{F}}(s,s')$ is a full rank matrix which integrates B cells drives and network correlations (due to the product with transfer matrices

$\mathcal{P}^{(0)^{-1}}$ and $\widetilde{\mathcal{P}^{(0)}}^{-1}$). These correlations are transported in time by the diagonal matrix $e^{\Lambda^{(0)}(t-s)}$. In general, there is no way to anticipate a priori what will be the combined effect of B cells RF overlaps and network on voltages correlations. Depending on the model parameters (characteristic times, synaptic weights) it can be anything. In particular, there is no general, mathematical reason, to think that A cells would decorrelate RG cells outputs.

This mathematical consequence is in apparent contrast with the claim, found in deep experimental papers stating that "the inhibition" (mediated by A cells) "decorrelates visual feature representations in the inner retina [**?**]". What could be the origin of this discrepancy ? A first reason is that correlations in the retina are often thought in terms of the drive correlations (**??**). Reducing the overlap between B cells RFs, i.e. decreasing the magnitude of the product $\mathcal{K}_{B_i}(x-x_i, y-y_i, t-s)\,\mathcal{K}_{B_{i'}}(x-x_{i'}, y-y_{i'}, t'-s)$ in the integral (**??**) lowers the drive correlations. The idea is then that A cells lateral inhibition reduces the center part of the RF and increases the surround thereby reducing the RFs overlap. Is there a way to mathematically validate this statement in (**??**) ? Under which conditions on models parameters does it hold true ?

Let us investigate what does "decorrelation" mean in our setting. Strictly speaking it means that $\mathcal{C}_{\vec{\mathcal{X}}}(t,t')$ is diagonal, that is, that the variable change corresponding to the transfer matrix $\mathcal{P}^{(0)}$ diagonalizes the stimulus correlation matrix $\mathcal{C}_{\mathcal{F}}(s,s')$. Now, $\mathcal{C}_{\mathcal{F}}(s,s')$, as a correlation matrix, is diagonalisable by an orthogonal basis change with real eigenvalues, whereas $\mathcal{P}^{(0)}$ has to do with B cells - A cells network and it is easy to find situation where it is complex, with complex eigenvalues. So, in general, the network effects do not diagonalize $\mathcal{C}_{\vec{\mathcal{X}}}(t,t')$. Nevertheless, it is indeed possible to construct networks diagonalising $\mathcal{C}_{\mathcal{F}}(s,s')$ by using the spectral decomposition theorem. In addition, if one does not stick at strict decorrelation one can also figure out conditions on networks reducing stimuli correlations. The question is whether *real* A cells networks match these conditions. This is an interesting question for further studies. We however see below that there are, however, other potential sources of decorrelation, especially non linearities.

Non correlated drives The correlation structure, complex in the non rectified case, is actually even worse when considering rectification. In the rest of this section, we want to consider in more detail the effects of rectification on RG cells spike correlations. We want to show that they induce non stationary stimulus dependent correlations which *are not* due to the drives correlations (**??**).

For this, we are going to consider the situation where $\mathcal{C}_{\mathcal{F}}$ is δ-correlated, that is we discard drives correlations. This corresponds to setting:

$$\vec{\mathcal{F}}(t) = \vec{m}(t) + \sigma_S \xi(t), \tag{38}$$

where $\vec{m}(t)$ is deterministic.

In this situation eq. (**??**) greatly simplifies, giving a correlation matrix:

$$\mathcal{C}_{\vec{\mathcal{X}}}(t,t') = \sigma_S^2\, e^{\mathcal{L}^{(0)}(t'-t)}.\int_{-\infty}^{t} e^{\tilde{\mathcal{L}}^{(0)}(t-s)}.e^{\mathcal{L}^{(0)}(t-s)}ds, \tag{39}$$

for $t' \geq t$.

In the general case $\mathcal{L}^{(0)}$ is not symmetric and does not commute with $\tilde{\mathcal{L}}^{(0)}$. One can then compute $\mathcal{C}_{\vec{\mathcal{X}}}(t,t')$ in terms of the (common) spectrum of $\mathcal{L}^{(0)}, \tilde{\mathcal{L}}^{(0)}$ using the spectral decomposition theorem $\mathcal{L}^{(0)} = \sum_{\alpha=1}^{N} \lambda_\alpha^{(0)}\, v_\alpha^{(0)}.\tilde{w}_\alpha^{(0)}$ where $v_\alpha^{(0)}$ is the right eigenvector α of $\mathcal{L}^{(0)}$ (the α-th column of $\mathcal{P}^{(0)}$) and $\tilde{w}_\alpha^{(0)}$ is the left eigenvector α of $\mathcal{L}^{(0)}$ (the α-th row of $\left(\mathcal{P}^{(0)}\right)^{-1}$). In general, right (left) eigenvectors are not mutually orthogonal but $\tilde{w}_\alpha^{(0)}.v_\beta^{(0)} = \delta_{\alpha\beta}$ so that $v_\alpha^{(0)}.\tilde{w}_\alpha^{(0)}$ is the projector on eigendirection α. From this, one obtains the correlation matrix:

$$\mathcal{C}_{\vec{\mathcal{X}}}(t,t') = -\sigma_S^2 \sum_{\alpha=1}^{N} e^{\lambda_\alpha^{(0)}(t'-t)}\, v_\alpha^{(0)}.\tilde{w}_\alpha^{(0)} \sum_{\beta=1}^{N} \frac{v_\beta^{(0)}.\tilde{w}_\beta^{(0)}}{\lambda_\alpha^{(0)} + \lambda_\beta^{(0)}}, \tag{40}$$

where eigenvalues are real or complex conjugate and are assumed to be stable (negative real part). Note that eigenvalues and projectors combine so that, in fine, the correlation matrix is real.

We will keep this general form for further discussions on the rectified case, but here, it is insightful to consider the case where $\mathcal{L}^{(0)}$ is symmetric. Here, it is diagonalizable in a orthogonal basis, with $\left(\mathcal{P}^{(0)}\right)^{-1} = \tilde{\mathcal{P}}^{(0)}$ and with real eigenvalues $\lambda_\beta \equiv -s_\beta, \beta = 1 \ldots N$. where s_β is real, positive. Then, (**??**) reduces, in form of components, to:

$$\mathcal{C}_{\alpha_2,\alpha_1}(t'-t) = \frac{\sigma_S^2}{2} \sum_{\beta=1}^{N} \frac{P_{\alpha_2\beta}\, P_{\alpha_1\beta}}{s_\beta}\, e^{-s_\beta(t'-t)}. \tag{41}$$

It is useful to express, from (**??**), the variance of cell α_{i_1}'s voltage (independent of time due to stationarity):

$$\sigma_{\alpha_1}^2 = \frac{\sigma_S^2}{2} \sum_{\beta=1}^{N} \frac{P_{\alpha_2\beta}\, P_{\alpha_1\beta}}{s_\beta}. \tag{42}$$

These computations provide the network correlations between cells voltage in the absence of drive correlations.

3.2.2 Spike correlations

We now compute spike correlations of RG cells induced by network correlations (**??**). We assume a spiking probability of the form (**??**). The probability that RG cell $\alpha_1 (> N_B + N_A)$ spikes at time t_1 is induced by the voltage probability

$\mathbb{P}$ and is given by $\nu_{\alpha_1}(t_1) \equiv \mathbb{E}\left[f\left(\frac{V_G(t)-\theta_G}{\sigma_G} \right) \right]$ where the expectation is taken with respect to $\mathbb{P}$. Taking the form (**??**) for f this is:

$$\nu_{\alpha_1}(t_1) = f\left(\frac{m_{\alpha_1}(t_1) - \theta_G}{\sqrt{\sigma_G^2 + \sigma_{\alpha_1}^2}} \right), \tag{43}$$

where m_{α_1} is the entry α_1 of the deterministic drive term in (**??**). As pointed out above, two sources of noise add up here: the implicit noise, with variance σ_G^2 appearing in the LNP formulation (**??**), which is intrinsic to the cell, and the network induced noise, explicit in the term $\sigma_{\alpha_1}^2$.

Likewise, the probability that RG cell $\alpha_1 (> N_B + N_A)$ spikes at time t_1 and RG cell $\alpha_2 (> N_B + N_A)$ spikes at time t_2 is:

$$\nu_{\alpha_1\alpha_2}(t_1,t_2) = \int f\left(\frac{\sqrt{\mu_1}\cos(\phi)\, y_1 - \sqrt{\mu_2}\sin(\phi)\, y_2 + m_{\alpha_1}(t_1) - \theta_G}{\sigma_G} \right) f\left(\frac{\sqrt{\mu_1}\sin(\phi)\, y_1 + \sqrt{\mu_2}\cos(\phi)\, y_2 + m_{\alpha_2}(t_2) - \theta_G}{\sigma_G} \right) DY, \tag{44}$$

where the integral holds on $\mathbb{R}^2$ and where $DY = \frac{1}{2\pi} e^{-\frac{y_1^2+y_2^2}{2}}\, dy_1\, dy_2$. Here, μ_1, μ_2 are the eigenvalues of the pairwise correlation matrix $\mathcal{C} = \begin{pmatrix} \sigma_{\alpha_1}^2 & \mathcal{C}_{\alpha_{i1}\alpha_{i2}}(t_1 - t_2) \\ \mathcal{C}_{\alpha_{i2}\alpha_{i1}}(t_2 - t_1) & \sigma_{\alpha_2}^2 \end{pmatrix}$ which is diagonalizable in an orthogonal basis with an orthogonal transformation, a rotation with angle ϕ determined by the coefficients of $\mathcal{C}$.

3.2.3 Decorrelation induced by non linearities

It is evident that the double integral (**??**) factorizes only in the case where $\mathcal{C}$ is diagonal ($\phi = 0, \mu_1 = \sigma_{\alpha_1}, \mu_2 = \sigma_{\alpha_2}$), and it reduces to $\nu_{\alpha_1\alpha_2}(t_1,t_2) = \nu_{\alpha_1}(t_1)\nu_{\alpha_2}(t_2)$. Thus, spikes of RG cell α_1 at time t_1 and of RG cell α_2 at time t_2 are decorrelated if and only if the correlation matrix (**??**) is diagonal. This matrix is diagonal only when there is no A cells. Otherwise, A cells have the effect to correlate voltages and thereby spikes. We already discussed above the possible effect of A cells in decorrelating the B cell drive term. Here, as we have removed this effect we are in position to discuss other potential effects inducing RG cells spikes decorrelation.

First, note that if the correlations we compute are non vanishing they can nevertheless be weak. The weakness of pairwise correlations in the retina has actually be reported by many authors [**?**, **?**]. It is known since Lancaster, 1957 [**?**] that the passage of two correlated Gaussian variables through a subsequent non linearity always reduces the correlation of the two signals, regardless of the shape of the non-linearity. Thus, in our case, the non linear function of the LNP model reduces the decorrelation.

Now, the LNP non linearity is not the only source of decorellation. Rectification also plays a crucial role. What happens, indeed, in the rectified case ? Mathematically, one can use equations (**??**) to compute the correlation matrices (**??**) (or even (**??**)), but the main, quite intricate problem is now that the

entrance and exit time of domains $t_-^{(n_k)}$, $t_+^{(n_k)}$ appearing in (**??**) are *themselves* random. This is again a consequence of the stimulus dependence of these times. The computation of the voltage correlations in this case being, for the moment, out of reach, I am going to give some straightforward although insightful remarks.

The non rectified case corresponds to a trajectory staying in the domain $\Omega^{(0)}$ (forgetting about conditions on noise ensuring that this holds for an infinite time). Now, the computation of voltage correlation is essentially the same if the trajectory stays in the domain $\Omega^{(n)}$. The only difference is that eigenvalues and projectors have a superscript (n) instead of (0). This difference is essential though, because rectification induces a projection on the space of non rectified cells. The contribution to rectified cells to voltage correlations with other cells vanishes thereby transforming the voltage correlations matrix. By permutation of rows and columns, one can convert this matrix in a form containing a diagonal block (correlations rectified cells $\leftrightarrow$ rectified cells) and a block characterizing the correlations non rectified cells $\leftrightarrow$ all cells. This reduces the model dimensionality and the global correlations. This effect, composed with the LN non linearity can reduce correlations even more.

The last important remark here is that rectification implies that RG cells correlations are *stimulus dependent even if we have removed the drives correlations* because the exit times of continuity domains are stimulus dependent. In addition, the obtained correlations are non stationary. This effect might not be noticeable with full field stimuli or white noise, which weakly solicit the lateral A cell connectivity, but it could be more prominent when studying spatio-temporal stimuli, in particular moving trajectories or non stationary stimuli, which constitute most of real visual scenes.

3.2.4 Conclusion of section ??

In this section, we have mathematically investigated the structure of correlations induced by the model (**??**), Fig. **??**. Our conclusion is essentially that the stimulus generates RG cells spike correlations modulated, on one hand by the drive correlations, and, on the other hand, by the B cells-A cells networks. More precisely, by the eigenvalues-eigenmodes of the transport operator. In addition, rectification and non linearities further impact correlations. This fact was reported by Pitkow and Meister in their paper "Decorrelation and efficient coding by retinal ganglion cells" [**?**] where they insist on the prominent role of non-linearities: "Most of the decorrelation was accomplished not by the receptive fields, but by non-linear processing in the retina". From these remarks they conclude about information transmission by the retinal network: "At very high thresholds, the information transmission is poor. Notably, transmission also drops at low thresholds. Thus, the choice of threshold involves a trade-off between rarely using reliable symbols, such as high spike counts, or frequently using unreliable symbols, such as low spike counts". Thus, non linearities plays a role in retinal coding making the spike rate of RG cells as sparse as possible, so that these cells are silent most of the time and fire at high rate only

when salient features of the stimulus make it necessary. This effect should be even more prominent for moving objects which is clearly an example of a stimulus with salient features and strong spatio-temporal correlations induced by its trajectory, especially if this trajectory shows sharp changes. This could be mathematically analysed in the present setting although to the price of consequent technical efforts.

Let us also remark that rectification makes the stochastic process of voltages non Gaussian, because the times of entering and exiting domains are now random variables too. As a consequence, spike statistics involves higher order correlations. Although it has to be further investigated on experimental grounds, this would lead to important consequences in terms of coding. As pointed out, again, by Pitkow and Meister [**?**], "for highly non-Gaussian signals, such as neural spike trains and natural images, correlation may be only weakly related to redundancy."·

Sticking at the model we may ask the following questions. Assume that we submit the model to different type of stimuli: the "classical" ones such as white noise, "Chirp" stimulus, natural images; but also more elaborated ones such as moving objects with different type of trajectories, or "natural movies" including motion and "surprise". For example, a bird crossing the visual scene, with, on the background, a forest of trees in the wind. It is known that the retina is able to filter the "noisy" motion of tree leaves while signalling the bird, thanks to dedicated circuits involving A cells [**?**, **?**]. Such circuits can be easily implemented in the model (**??**) [**?**]. What will be the structure of its spike trains, depending on the different type of stimuli ? How can one "efficiently" decode the stimulus from the mere knowledge of those spike trains ? How efficient is a decoding scheme based on independent, decorrelated RG cells ? In contrast, would cooperative network effects make the code more precise affording faster responses to motion [**?**] ?

Although we are not going to answer these questions here (there is still a long way to it), we give, in the next sections, several insightful mathematical results in this direction.

3.3 Computing the mixed effect of network and stimulus on spike correlations

3.3.1 Context

Let us now consider the retina from the point of view of its output. We sit on the optic nerve and measure the spikes sent to the LGN and cortex via the optic nerve. We have no access to the biophysical machinery taking place in the retina and generating those spikes, but we know that the spike trains contains information about the external world stimuli that we want to extract. We can measure as many quantities as we want such as firing rate, or higher correlations. More generally, we are seeking the (time dependent) joint probability of spikes adopting the approach described in section **??**, Methods.

In this context, assume a retina "at rest" i.e. receiving no stimulus or stationary stimuli like noise. We can describe the spike trains emitted by this retina by a stationary transition probability $\mathbb{P}$, associated with a stationary probability $\mu^{(sp)}$ (for "spontaneous"). In general, this probability has spike correlations of order 2 and higher. Assume now that, from time t_0 on a stimulus (say a moving object) is getting through the visual field of this retina. As exposed in section **??** one expects the spike correlations (at any order) to be modified by this stimulation. Typically, a moving object carries spatio-temporal correlations in its trajectory which will superimposed upon the network correlations, resulting in a mixed effect where non linearities can also play a role. Can we predict, for a given stimulus, how correlations will be modified ?

Let us give an example. Consider a linear chain of neurons, as depicted in Fig. **??** top. Each neuron (black points), is connected to its neighbours with an excitatory connection (red arrows) and to its second nearest neighbours with an inhibitory connection (blue arrows). The model here is a classical leaky integrate and fire model in the presence of noise, where parameters have been tuned to have a spontaneous asynchronous activity as depicted in Fig. **??** bottom, left. See [**?**] for more detail. Consider a moving stimulus $\mathcal{S}(x,t)$ propagating from left to right (cyan, bell shaped curve) Fig. **??** top. $\mathcal{S}(x,t)$ acts as an input current of the form $\mathcal{S}(x,t) = f(x-vt)$ where v is the propagation speed and f, typically, a Gaussian. This stimulus is going to modify the spike patterns, as seen in Fig. **??** bottom, left, where one sees clearly nearest neighbours excitation and second nearest neighbours inhibition. The remarkable fact is that the stimulus not only modifies the firing rates of neurons, but also *their correlations*. The question is: can we compute this effect ?

This question has been solved in the paper [**?**] for the gIF model (**??**). Here, we briefly state the main result (see the paper for technical details). Consider a function $f(t,\omega)$ (observable) depending on time and spike history up to time t. Let $\mu^{(sp)}$ be the join probability distribution of spikes in spontaneous activity (no stimulus), and μ the join probability distribution of spikes in the presence of a spatio-temporal stimulus $S(x,t)$. We note $\delta\mu\,[\,f\,]\,(t) = \mu\,[\,f\,]\,(t) - \mu^{(sp)}\,[\,f\,]$, where $\mu\,[\,f\,]\,(t)$ is the average of f, at time t, in the presence of the stimulus and $\mu^{(sp)}\,[\,f\,]$ the average of f in spontaneous activity (which does not depend of time because spontaneous dynamics is stationary). $\delta\mu\,[\,f(t)\,]$ characterizes how much the time dependent mean of $f(t,\omega)$ under stimulation departs from the spontaneous mean at time t. In the simplest case $\delta\mu\,[\,f(t)\,]$ characterizes the variation in the firing rate of neuron k, if $f(t,\omega) = \omega_k(t)$, or the variation in the correlation between neuron k_1 at time t_1 and neuron k_2 at time $t_1 + t$ if $f(t,\omega) = \big(\omega_{k_1}(t_1) - \mu^{(sp)}\,[\,\omega_{k_1}\,]\big)\,\big(\omega_{k_2}(t_1+t) - \mu^{(sp)}\,[\,\omega_{k_2}\,]\big)$, and so on.

One can show that, when the stimulus amplitude is weak enough, $\delta\mu\,[\,f(t)\,]$ is given by a linear response formula of the form:

$$\delta\mu\,[\,f(t)\,] \;=\; [\,K_f * S\,]\,(\,t\,) \tag{45}$$

That is, by the convolution of the stimulus with a specific kernel, K_f, *depending on the observable f and on the spontaneous distribution $\mu^{(sp)}$*. We do not give

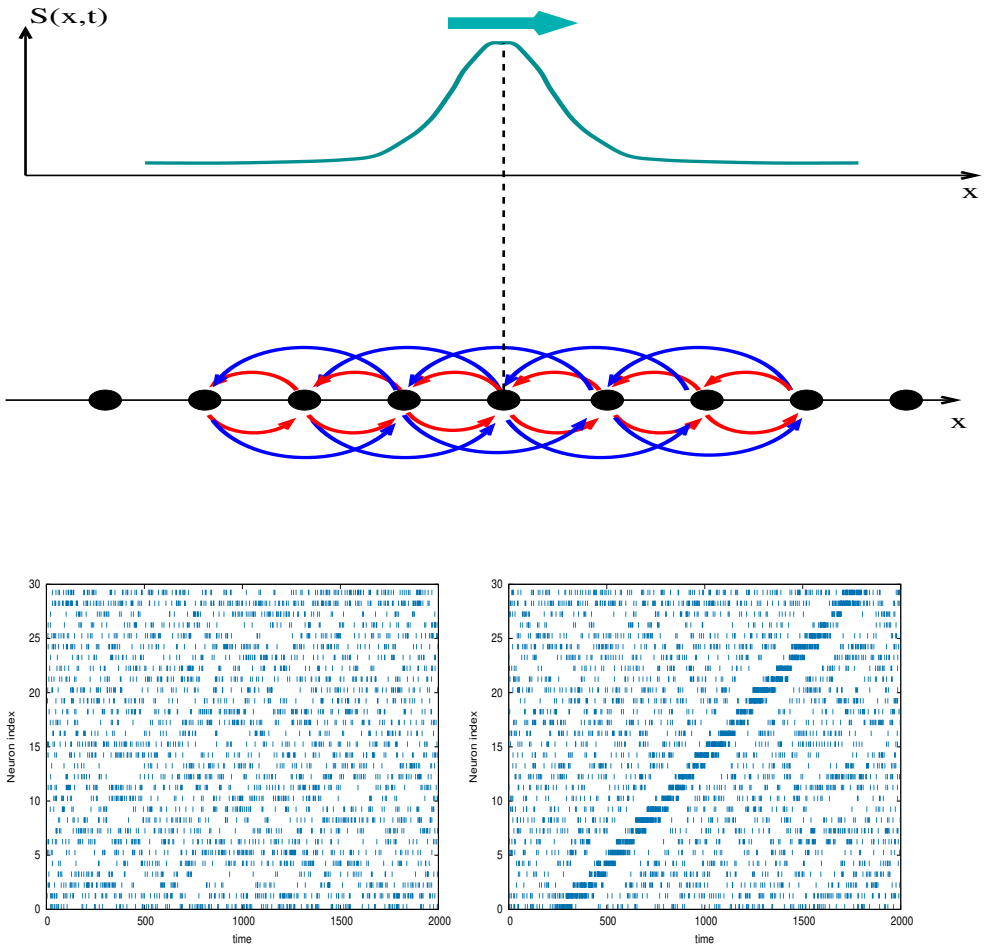

Figure 3: **Top. Network of spiking neurons sensing a stimulus** (adapted from [**?**]). Each neuron, represented as a black point, is connected to its neighbours with an excitatory connection (red arrows) and to its second nearest neighbours with a inhibitory connection (blue arrows). In addition, each neuron is able to sense external stimuli $S(x,t)$ (cyan, bell shaped curve). **Bottom. Left.** Spontaneous spiking activity. **Bottom. Right.** Spiking activity in the presence of the moving stimulus.

the expression of this kernel here, for simplicity, but the reader can refer to the paper [**?**].

3.3.2 Consequences

Convolution. Similarly to (**??**) (RG cells response to stimuli) or (**??**) (B cells response to stimuli), we have here again a linear response where the effect of a stimulus on a system is expressed by a convolution. We are however in a completely different perspective. Indeed, while we were considering formerly voltage response of individuals cells (shaped by network effects), we are now working on a more abstract level, where we attempt to measure the effect of a stimulus on *statistics*. This is of course due to the difference in what is accessible by experiments, what the observer is able to deal with in his observations, here spikes. Thus, the mathematical machinery allowing to extract the response requires to define spike statistics in a non stationary setting, where the influence of the stimulus can be inferred.

Kernel. The kernel K_f can be explicitly computed in the gIF model. It depends on several features. First, on network characteristics (especially the effective interaction W_{kj}, and more generally, the parameters shaping the model dynamics). It also depends on the observable f. However, the main content of this result is that the kernel K_f is actually determined by spike correlations in *spontaneous activity*. In other word, it is possible to anticipate the response to a non stationary stimulus from the knowledge of the spontaneous activity. Although this result is expected from Kubo theory in non equilibrium statistical physics [**?**, **?**] or from Volterra-Wiener expansions [**?**], it has interesting consequences when dealing with neural dynamics, and more specifically here, with retina outputs. First, it provides a consistent treatment of the expected perturbation of higher-order correlations, beyond the known linear perturbation of firing rates and instantaneous pairwise correlations; in particular, it extends to time-dependent correlations. In addition it reveals how the stimulus-response and dynamics are entangled in a complex manner. For example, the response of a neuron k to a stimulus applied on neuron i does not only depends on the synaptic weight W_{ki} but, in general, on all synaptic weights, because the dynamics create complex causality loops which build up the response of neuron k [**?**, **?**, **?**]. The linear response formula is written in terms of the parameters of a spiking neuronal network model and the spike history of the network. In the presence of stimuli, the whole architecture of synaptic connectivity, history and the dynamical properties of the networks are playing a role in the spatio-temporal correlations structure.

Linear response and higher order corrections. The derived formula provides a good agreement with simulations in the gIF model under time dependent stimuli (typically, a moving object). It requires however that the stimulus amplitude is weak enough. That is, higher corrections are weaker than the leading

order. For larger amplitude stimuli one would to compute higher order correlations. This can be done using the same formalism [**?**] although it might not be the best approach. Indeed, this method requires to measure spontaneous correlations which are difficult to obtain experimentally for orders higher than 2. This is actually one of the reason why LNP-like models exist. The expected non linearity in the response is handled by a static non linear function. Exploring what could be the best non linear correction to the linear response in such models is definitely an interesting mathematical challenge.

3.4 Conclusion

Beyond naive RF description. This linear response theory actually shows how the neuronal network substrate and stimulus response are entangled. Indeed, in contrast to naive RF representation where the convolution kernel is assumed to depend only on the *cell*, here, mathematics show that it depends as well on the *observable*. The explicit form of the kernel is also tightly constrained by the neurons connections. Finally, a convolution implies an integration over histories, requiring thereby to consider spikes probabilities with memory, instead of "instantaneous" spikes probabilities (not or weakly depending on the past). Of course, one may always argue that, on experimental grounds, long tail memory are just impossible to measure so "instantaneous" [**?**] or first order Markov [**?**] models are largely sufficient. But what means "sufficient" ? This is a difficult question which requires sophisticated methods to determine the "best performing" memory depth from data [**?**, **?**, **?**]. Actually, numerical computations of the kernel use, of course, Markovian approximations [**?**] although with a memory depth that can be controlled.

Link with the retina model. Can we relate the formalism developed here with the retinal model presented in the section **??** ? As RG cells voltage are Gaussian it is in principle possible to compute transition probabilities using the transport operator formalism. However, even in the non rectified case, the computation promises to be a formidable task, unless one adds some additional constraints. For example, a big advantage of Integrate and Fire models is that a spiking neuron loses memory after spiking, a property which is not implemented in LNP like models.

Information geometry. There is a close link between Gibbs distributions and information geometry. This theory, developed by Shun'ichi Amari and his collaborators (see [**?**] and references therein) on the basis of early work from Rao [**?**] establishes a geometric theory of information where probabilities are considered as points on Riemannian manifolds. A prominent family of probability measures is called the exponential family. It contains the Gibbs distributions in the standard statistical physics sense, i.e. probabilities having the form $\frac{e^{-\beta H}}{Z}$ where the energy H does not depend on time. In this case, the metric is given by the Hessian of $\log(Z)$, the free energy, and is tightly linked to Fisher information

on one hand and to linear response on the other hand. The linear response is actually a correlation function, from the fluctuation dissipation theorem. Thus, correlation functions induce a natural geometry for Gibbs distributions providing strong insights on how these distribution are modified by smooth, local, transformations of their parameters (like learning [**?**]) or under a stimulation of weak amplitude. In this last case, the stimulus action corresponds to a perturbation in the tangent space of the manifold [**?**, **?**]. Although information geometry has not been extended, to our best knowledge, to the type of Gibbs distribution we study here (they are non stationary) the mathematical formalism is similar. This essentially tells us that the structure of spatio-temporal correlations observed in spike trains reveals an hidden geometrical structure which, somewhat, shapes the response of the retina, and, henceforth of cortex, to stimuli. We come back to this point in the conclusion section.

4 Applications

The OPL-B cells-A cells processing is based on graded potentials departing from the classical paradigm of binary spike processing. Mathematically, this has strong consequences in terms of response to a spatio-temporal stimulus: existence of eigenmodes, potentially modulated by non linear effects, inducing properties such as activity waves ahead of the stimulus (anticipation), resonances, correlations modified by the stimulus. In this section, and although this paper is essentially theoretical I would like to shortly propose possible applications of these results, outside the field of neuroscience.

4.1 Retinal prostheses

Retinal pathologies such as Age Macular Degeneration or Retinitis Pigmentosa, are due to the degeneration of photo-receptors [**?**]. In addition, they induce morphological and structural changes in the retina with significant pathogenic effects: inflammation, change in connectivity, the appearance of large-scale spontaneous electrical oscillations, and, of course, attenuation of response to visual stimuli [**?**, **?**, **?**, **?**]. In this process of degeneration, however, the RG cells are the last to be deficient, maintaining, therefore, a link between the retina and the brain, provided they are suitably stimulated. The strategy of retinal prostheses is to stimulate the retina electrically by an array of electrodes. Stimulation of an electrode generates, in the visual cortex, a phosphene, the perception of a light spot. By stimulating the electrodes one induces in the cortex an image "pixelised" by the phosphenes, with resolution limited, on the one hand, by the number of electrodes, and, on the other hand, by the size of the phosphenes, which can be enlarged by diffusion and non linear effects [**?**]. Technological solutions, taking into account the physiological limitation on the electrical power that can be injected in an electrode, improve resolution [**?**]. However, there are still obstacles which cannot be resolved by purely technological solutions (hardware). In addition, a valid stimulation strategy at a given period of the

pathology may not be later because the retina degeneration evolves in time.

Stimulation strategies use processor pre-processing to calculate, from a given image (captured by a camera) the pattern of stimulation of the prosthesis, by mimicking the calculation that a healthy retina would make, or by incorporating corrections taking into account the pathology [**?**]. These algorithms might be improved using what we know about the retinal structure, especially A cells lateral connectivity, where a model like (**??**) can be easily implemented with a relatively low energy consumption cost. The idea would be to improve electrodes stimulation sequences so as to allow an implanted patient to perceive in real time a moving object. The model (**??**) with A cells lateral connectivity and gain control is known to produce a wave of activity ahead of a stimulus, performing a form of anticipation [**?**]. This could be used to compensate for the processing times imposed by the equipment, in the same way that the visual system knows how to compensate for the delays induced by photo-transduction [**?**]. The ideal would also be to have adaptive algorithms, i.e., depending on parameters adjustable according to the patient and the course of his pathology.

4.2 Convolutional networks

Several recent studies attempt to understand how retinal response to stimuli is related to circuit processes using convolutional neural network models [**?**] to grasp the structure of retinal prediction [**?**]. Reciprocally, these networks can be used to design deep-learning models to encode dynamic visual scenes with important potential outcomes in the domain of computer vision. Especially, a recent work by Zheng et al, [**?**], shows the important role played by recurrence in encoding complex natural scenes. To my best knowledge (which is quite scarce in this field) there is no mathematical analysis of the dynamics of these models, especially the dependence in parameters and robustness of the training schemes. The present study could bring some insights in this perspective. Even if the model (**??**) is different from what these researchers were using, the techniques of piecewise linear phase decomposition and eigenmodes study could be insightful to better understand the dynamical evolution of these convolutional networks and the role played by rectification.

5 Discussion

In this paper we have addressed mathematically the potential effect of A cells lateral connectivity on retinal response to spatio-temporal stimuli. We have seen how, mathematically, the retina structure and the collective dynamics of retinal cells organized in local circuits spanning the whole retina might constrain this response. In particular, the structure of correlations is expected to depend on the stimulus, as soon as non linear effects are involved. This goes beyond the expected effect of stimulus correlations induced by RF overlap.

These properties are established on the basis of theoretical results which are based on incomplete modelling of the retina, and specific assumptions. Their

validation would require experiments, some of which may require time and others are not yet accessible, for example, simultaneously measuring retina and cortex. As a matter of fact, one may argue that the models presented here are far too simplistic compared to the real retina(s) having a large number of B cells, A cells, RG cells type, making complex circuits [**?**] and whose characteristics depend, in addition, on species, age or pathologies. However, the idea behind mathematical modelling is precisely to try and infer some generic mechanism underlying the real object under study, here the retina. This is the simplicity of the structure which makes it generic. The question is: "Would the addition of more elaborated retinal features make the response to stimuli simpler ?"

In the next section I discuss some further implications of this work leading to some new questions.

5.1 Cortical response

If a dynamical stimulus, combined with the retinal network and non linearities produces non negligible dynamical spatio-temporal correlations, what could be the consequences at the cortical level[3] ? There is a physiological transformation, called retinotopy, which maps smoothly the retina topology to the cortical V1 topology. In models, it is usually considered to be the identity map, although it is not. This is a non linear transformation, depending, in addition on the species [**?**, **?**, **?**]. Nevertheless, what matters here is that this mapping is smooth and invertible. Therefore, retinotopy transports, in a smooth and invertible way the spatio-temporal retinal correlations to the visual cortex. This leads to a question: "How to define a cortical model taking into account spatio-temporal spike correlations ?"

Cortical models are usually based on mean-field approximations where one features firing rates evolution, but not spike correlations. This is the case of the Wilson-Cowan model [**?**, **?**, **?**] or neural field models [**?**, **?**, **?**]. I know about 2 mean-field approaches taking care of spike correlations.

The first approach is the one initiated by S. El Boustani et A. Destexhe [**?**] using a Markovian approach to write down mean field equations of second order (i.e. including pairwise spatial correlations) and a non static thalamic entry, that can feature the retinal-LGN input. This model can be used to construct a retino-cortical model [**?**], although the mathematical consequences of having correlated retinal entries have not been explored yet.

The second approach is based on the so-called Ott-Antonsen Ansatz [**?**] and has been used by Montbrio, Pazo, Roxin to propose an exact mean field approach with second order statistics [**?**]. Since their paper there has been a lot of activity in developing this model, especially in connection with cortical imaging, with impressive results [**?**, **?**, **?**, **?**]. It is a promising track.

All these approaches could certainly provide powerful numerical and mathematical tools to better understand how spatio-temporal retinal correlations could be processed. In particular, having a retino-(LGN)-cortical model allows

[3]For simplicity, I am going to consider the LGN as a simple relay

to do a task which is currently impossible experimentally: measure simultaneously retina and cortex.

5.2 Retinal correlations and neurogeometry.

We have also seen that retinal correlations and Gibbs distributions naturally define a metric on a Riemannian manifold where probabilities are points on this manifold. In particular, the application of a weak amplitude stimulus corresponds to a perturbation along the tangent space of this manifold. What is the image of this metric under the retinotopic transformation ? Let us make this question a bit more precise.

The visual system has evolved to map as efficiently as possible retinal output to cortical structures. The shaping of the visual system during development is actually a highly dynamical process involving retinal waves and synaptic plasticity [**?**]. These processes provide the visual system a structure allowing it to respond in a fast and efficient way to the stimuli coming from the external world, via the retina. In particular, the capacity of the visual cortex to respond to spike trains with spatio-temporal correlations induced by natural stimuli should be somewhat imprinted in the cortical connectivity.

Visual perception is actually highly geometrically structured and shaped by the structure of cortical connectivity. This leads researchers to introduce a link between the geometry of cortex and the geometry of vision in the concept of neurogeometry (or neuromathematics) where the functional architecture of V1 is considered as a Lie group of symmetry with a Riemannian geometry (see [**?**, **?**, **?**, **?**] and reference therein). In this approach cortical columns are point-like processors detecting visual features where functional connectivity is represented in terms of geodesics. To my best knowledge, neurogeometry essentially deals with V1 and static percepts, although extensions to motor cortex [**?**] and motion areas [**?**] have been done. Now, a natural question is: "Is there a relation between the cortical metric of neurogeometry and the metric induced by spatio-temporal spike correlations observed by the retina ?".

Let us address the problem the other way round: Project the cortical metric back to the retina via the inverse retinotopy map, what do we find ? Is there a physiological correspondence with the retina structure and especially lateral connectivity ? What could be the consequences on spike trains statistics and on the way retina processes visual stimuli ? What does cortical metric tell us about retinal spikes correlations ? Dealing with neural coding of vision, the simplest assumption consists of assuming that RG cells are independent encoders and that cortex makes the job of restoring the spatio-temporal correlations existing in the visual scene (e.g. in the trajectory of a moving object). The alternative proposition, where spatio-temporal correlations imprinted in the RGCs spike trains are deciphered by the cortex makes the question of stimuli decoding by the cortex more challenging, but opens up far more possibilities. Answering to these questions could be based, as a first step, on important results existing in the literature. Especially, recent works asking the extent to which retinal connectivity and dynamics affect higher order features later derived from its

outputs (e.g. orientation, spatial frequency speed etc) in V1 via LGN [**?**, **?**, **?**].

Acknowledgements

I would like to warmly acknowledge all the neuroscientist collaborators from which I learned about this beautiful object, the retina, and, more generally, about vision: Frédéric Chavane, Gerrit Hilgen, Olivier Marre, Adrian Palacios, Serge Picaud and Evelyne Sernagor. I thanks the reviewers for their detailed review and helpful criticism which helped to significantly improve the paper.

References

References

[1] T. Gollisch and M. Meister. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron*, 65(2):150–164, January 2010.

[2] Fred Attneave. Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193, May 1954.

[3] Oliver G. Selfridge. Pandaemonium: A paradigm for learning. In *Proceedings of the Symposium on the mechanisation of thought processes*, London: HMSO, 1959.

[4] C.Y. Lee. Representation of switching functions by binary decision programs. *Bell Systems Technical Journal*, (38):985–999, 1959.

[5] H.B. Barlow. Possible principles underlying the transformation of sensory messages. *Sensory communication*, pages 217–234, 1961.

[6] Joseph J. Atick and A. Norman Redlich. Towards a Theory of Early Visual Processing. *Neural Computation*, 2(3):308–320, 1990.

[7] Bruno A. Olshausen and David J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23):3311–3325, 1997.

[8] William E. Vinje and Jack L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–1276, 2000.

[9] E.P. Simoncelli and B.A. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1):1193–1216, 2001.

[10] Eero P Simoncelli. Vision and the statistics of the visual environment. In *Current Opinion in Neurobiology*, pages 144–149, 2003.

[11] Li Zhaoping. Theoretical understanding of the early visual processes by data compression and data selection. *Network: Computation in Neural Systems*, 17(4):301–334, 2006. PMID: 17283516.

[12] Xaq Pitkow and Markus Meister. Decorrelation and efficient coding by retinal ganglion cells. *Nature neuroscience*, 15(4):628–635, 2012.

[13] L Zhaoping. *Understanding Vision: Theory, Models, and Data*, chapter The efficient coding principle. Oxford University Press, 2014.

[14] Sophie Denève and Christian K Machens. Efficient codes and balanced networks. *Nature neuroscience*, 19(3):375–382, 2016.

[15] K. Franke, P. Berens, T. Schubert, M. Bethge, T. Euler, and T. Baden. Inhibition decorrelates visual feature representations in the inner retina. *Nature*, 542:439–444, 2017.

[16] J. Besharse and D. Bok. *The Retina and its Disorders*. Elsevier Science, 2011.

[17] Ralph Nelson and Helga Kolb. On and off pathways in the vertebrate retina and visual system. *The Visual Neurosciences*, 1:260–278, 2004.

[18] Rava Azeredo da Silveira and Botond Roska. Cell types, circuits, computation. *Current Opinion in Neurobiology*, 21(5):664–671, 2011. Networks, circuits and computation.

[19] N.W. Oesch and J.S. Diamond. Ribbon synapses compute temporal contrast and encode luminance in retinal rod bipolar cells. *Nat Neurosci.*, 23(14(12)):1555–61, 2011.

[20] M. Meister and M.J. Berry. The neural code of the retina. *Neuron*, 22:435–450, March 1999.

[21] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes: Exploring the Neural Code*. Bradford Books, 1997.

[22] E. Schneidman, M.J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[23] Stephanie E Palmer, Olivier Marre, II Berry, J Michael, and William Bialek. Predictive information in a sensory population. *arXiv:1307.0225*, 2013.

[24] G. Tkacik, O. Marre, T. Mora, D. Amodei, M. Berry, and W. Bialek. The simplest maximum entropy model for collective behavior in a neural network. *J Stat Mech*, page P03011, 2013.

[25] Stephanie E. Palmer, Olivier Marre, Michael J. Berry, and William Bialek. Predictive information in a sensory population. *Proceedings of the National Academy of Sciences*, 112(22):6908–6913, 2015.

[26] A.S. Ecker, P. Berens, G.A. Keliris, M. Bethge, N.K. Logothetis, and A.S. Tolias. Decorrelated neuronal firing in cortical microcircuits. *Science*, 327(5965):584, January 2010.

[27] Tom Baden, Philipp Berens, Matthias Bethge, and Thomas Euler. Spikes in mammalian bipolar cells support temporal layering of the inner retina. *Current Biology*, 23(1):48 – 52, 2013.

[28] D. Karvouniari, L. Gil, O. Marre, S. Picaud, and B. Cessac. A biophysical model explains the oscillatory behaviour of immature starburst amacrine cells. *Scientific Reports*, 9:1859, 2019.

[29] J.W. Pillow, L. Paninski, V.J. Uzzell, E.P. Simoncelli, and E.J. Chichilnisky. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *Journal of Neuroscience*, 25(47):11003–11013, 2005.

[30] Kiersten Ruda, Joel Zylberberg, and Greg Field. Ignoring correlated activity causes a failure of retinal population codes. *Nature communications*, 11:4605, 2020.

[31] Sudarshan Sekhar, Poornima Ramesh, Giacomo Bassetto, Eberhart Zrenner, Jakob H. Macke, and Daniel L. Rathbun. Characterizing retinal ganglion cell responses to electrical stimulation using generalized linear models. *Frontiers in Neuroscience*, 14:378, 2020.

[32] P. Martínez-Cañada, C. Morillas, B. Pino, E. Ros, and F. Pelayo. A computational framework for realistic retina modeling. *International Journal of Neural Systems*, 26(07), 2016.

[33] Jacob Huth, Timothée Masquelier, and Angelo Arleo. Convis: A toolbox to fit and simulate filter-based models of early visual processing. *Frontiers in Neuroinformatics*, 12:9, 2018.

[34] Adrien Wohrer and Pierre Kornprobst. Virtual Retina : A biological retina model and simulator, with contrast gain control. *Journal of Computational Neuroscience*, 26(2):219, 2009. DOI 10.1007/s10827-008-0108-4.

[35] Bruno Cessac, Pierre Kornprobst, Selim Kraria, Hassan Nasser, Daniela Pamplona, Geoffrey Portelli, and Thierry Viéville. PRANAS: A new platform for retinal analysis and simulation. *Frontiers in Neuroinformatics*, 11:49, 2017.

[36] W.R. Taylor and R.G. Smith. The role of starburst amacrine cells in visual signal processing. *Visual Neuroscience*, 29(1):73–81, 2012.

[37] J. B. Demb and J. H. Singer. Intrinsic properties and functional circuitry of the aii amacrine cell. *Visual Neuroscience*, 29(1):51–60, 2012.

[38] Joseph Pottackal, Joshua H. Singer, and Jonathan B. Demb. Computational and molecular properties of starburst amacrine cell synapses differ with postsynaptic cell type. *Frontiers in Cellular Neuroscience*, 15:244, 2021.

[39] E. J. Chichilnisky. A simple white noise analysis of neuronal light responses. *Network: Comput. Neural Syst.*, 12:199–213, 2001.

[40] E P Simoncelli, L Paninski, J Pillow, and O Schwartz. Characterization of neural responses with stochastic stimuli. In M Gazzaniga, editor, *The Cognitive Neurosciences, III*, chapter 23, pages 327–338. MIT Press, 2004.

[41] Odelia Schwartz, Jonathan W. Pillow, Nicole C. Rust, and Eero P. Simoncelli. Spike-triggered neural characterization. *Journal of Vision*, 6(4):13–13, 2006.

[42] H. Motulsky and A. Christopoulos. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting.* Oxford University Press, Oxford, 2004.

[43] M.J. Berry, I.H. Brivanlou, T.A. Jordan, and M. Meister. Anticipation of moving stimuli by the retina. *Nature*, 398(6725):334—338, 1999.

[44] T. Hosoya, S. A. Baccus, and M. Meister. Dynamic predictive coding by the retina. *Nature*, 436:71–77, 2005.

[45] Eric Y Chen, Olivier Marre, Clark Fisher, Greg Schwartz, Joshua Levy, Rava Azeredo da Silviera, and Michael Berry. Alert response to motion onset in the retina. *Journal of Neuroscience*, 33(1):120–132, 2013.

[46] Selma Souihel and Bruno Cessac. On the potential role of lateral connectivity in retinal anticipation. *J. Math. Neuro. to appear*, December 2020.

[47] Bruno Cessac. Statistics of spike trains in conductance-based neural networks: Rigorous results. *The Journal of Mathematical Neuroscience*, 1(8):1–42, 2011.

[48] Rodrigo Cofré and Bruno Cessac. Dynamics and spike trains statistics in conductance-based integrate-and-fire neural networks with chemical and electric synapses. *Chaos, Solitons & Fractals*, 50(13):3, 2013.

[49] Bruno Cessac and Rodrigo Cofré. Spike train statistics and Gibbs distributions. *Journal of Physiology-Paris*, 107(5):360–368, November 2013. Special issue: Neural Coding and Natural Image Statistics.

[50] Rodrigo Cofré and Bruno Cessac. Exact computation of the maximum-entropy potential of spiking neural-network models. *Phys. Rev. E*, 89(052117), 2014.

[51] Rodrigo Cofré, Cesar Maldonado, and Bruno Cessac. Thermodynamic formalism in neuronal dynamics and spike train statistics. *Entropy*, 22(1330), 2020.

[52] Bruno Cessac, Ignacio Ampuero, and Rodrigo Cofré. Linear response of general observables in spiking neuronal network models. *Entropy*, 23(2), 2021.

[53] Alessandro Sarti and Giovanna Citti. The constitution of visual perceptual units in the functional architecture of v1. *Journal of computational neuroscience*, pages 1–16, 2014.

[54] Giovanna Citti and Alessandro Sarti, editors. *Neuromathematics of Vision*. Springer, 2014.

[55] Jean Petitot. *Elements of neurogeometry*. Springer, 2017.

[56] Giovanna Citti and Alessandro Sarti. Neurogeometry of perception: Isotropic and anisotropic aspects. *Axiomathes*, 2019.

[57] S. Baccus and M. Meister. Fast and slow contrast adaptation in retinal circuitry. *Neuron*, 36(5):909–919, 2002.

[58] Jamie Johnston and Leon Lagnado. General features of the retinal connectome determine the computation of motion anticipation. *Elife*, 2015.

[59] S. Trenholm and G.B. Awatramani. Chapter 9 - dynamic properties of electrically coupled retinal networks. In Jian Jing, editor, *Network Functions and Plasticity*, pages 183–208. Academic Press, 2017.

[60] Philippe Blanchard, Bruno Cessac, and Tyll Krueger. A dynamical system approach to soc models of zhang's type. *J. Stat. Phys.*, 88:307–318, 1997.

[61] Ph. Blanchard, Bruno Cessac, and T. Krueger. What can one learn about self-organized criticality from dynamical system theory ? *Journal of Statistical Physics*, 98:375–404, 2000.

[62] Bruno Cessac. A discrete time neural network model with spiking neurons. rigorous results on the spontaneous dynamics. *J. Math. Biol.*, 56(3):311–345, 2008.

[63] Bruno Cessac and Thierry Viéville. On dynamics of integrate-and-fire neural networks with adaptive conductances. *Frontiers in neuroscience*, 2(2), July 2008.

[64] Bruno Cessac. A discrete time neural network model with spiking neurons ii. dynamics with noise. *Journal of Mathematical Biology*, 62(6):863–900, 2011.

[65] S. Coombes, Y. M. Lai, M. Şayli, and R. Thul. Networks of piecewise linear neural mass models. *European Journal of Applied Mathematics*, 29(5):869–890, 2018.

[66] Alfred Rajakumar, John Rinzel, and Zhe S. Chen. Stimulus-Driven and Spontaneous Dynamics in Excitatory-Inhibitory Recurrent Neural Networks for Sequence Representation. *Neural Computation*, 33(10):2603–2645, 2021.

[67] Mark S. Goldman. Memory without feedback in a neural network. *Neuron*, 61(4):621–634, 2009.

[68] B.K. Murphy and K.D. Miller. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron*, 61(4):635–648, 2009.

[69] K. J. Falconer. *The Geometry of Fractal Sets*. Cambridge University Press, USA, 1985.

[70] M.F. Barnsley and H. Rising. *Fractals Everywhere*. Elsevier Science, 1993.

[71] Kenneth John Falconer. Techniques in fractal geometry, 1997. John Wiley & Sons, Ltd., Chichester.

[72] Rebecca A. Mease, Michael Famulare, Julijana Gjorgjieva, William J. Moody, and Adrienne L. Fairhall. Emergence of adaptive computation by single neurons in the developing cortex. *Journal of Neuroscience*, 33(30):12154–12170, 2013.

[73] Yuguo Yu and Tai Sing Lee. Adaptive contrast gain control and information maximization. *Neurocomputing*, 65-66:111 – 116, 2005. Computational Neuroscience: Trends in Research 2005.

[74] J. Snellman, T. Kaur, Y. Shen, and S. Nawy. Regulation of on bipolar cell activity. *Progress in retinal and eye research*, 27(4):450–63, 2008.

[75] Lai-Sang Young. Mathematical theory of lyapunov exponents. *Journal of Physics A: Mathematical and Theoretical*, 46(25):254001, 2013.

[76] Hans-Otto Georgii. *Gibbs Measures and Phase Transitions*. De Gruyter, 1988.

[77] O. Onicescu and G. Mihoc. Sur les chaînes statistiques. *C. R. Acad. Sci. Paris*, 200:511–512, 1935.

[78] A. Galves and E. Löcherbach. Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets. *J Stat Phys*, 151:896–921, 2013.

[79] Roberto Fernandez and Grégory Maillard. Chains with complete connections : General theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.*, 118(3-4):555–588, 2005.

[80] A. LeNy. Introduction to (generalized) gibbs measures. *Ensaios Matematicos*, 15:1–126, 2008.

[81] J. Shlens, G.D. Field, J.L. Gauthier, M.I. Grivich, D. Petrusca, A. Sher, A.M. Litke, and E.J. Chichilnisky. The structure of multi-neuron firing patterns in primate retina. *Journal of Neuroscience*, 26(32):8254, 2006.

[82] 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. *Phys. Rev. E*, 98:012402, 2018.

[83] Simona Cocco, Stanislas Leibler, and Rémi Monasson. Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *PNAS*, 106(33):14058–14062, 2009.

[84] M. Rudolph and A. Destexhe. Analytical integrate and fire neuron models with conductance-based dynamics for event driven simulation strategies. *Neural Computation*, 18:2146–2210, 2006.

[85] W. Gerstner and W. M. Kistler. Mathematical formulations of hebbian learning. *Biological Cybernetics*, 87:404–415, 2002.

[86] P. Dayan and L.F. Abbott. *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

[87] Alain Destexhe, Zachary F. Mainen, and Terrence J. Sejnowski. *Methods in Neuronal Modeling*, chapter Kinetic models of synaptic transmission, pages 1–25. The MIT Press, 1998.

[88] T Baden, P Berens, K Franke, M Román Rosón, M Bethge, and T Euler. The functional diversity of retinal ganglion cells in the mouse. *Nature*, 2016.

[89] E. Sernagor and M.H. Hennig. Chapter 49 - retinal waves: Underlying cellular mechanisms and theoretical considerations. In John L.R. Rubenstein and Pasko Rakic, editors, *Cellular Migration and Formation of Neuronal Connections*, pages 909 – 920. Academic Press, Oxford, 2013.

[90] Giacomo Benvenuti, Sandrine Chemla, Guillaume Masson Arjan Boonman, and Frédéric Chavane. Anticipation of an approaching bar by neuronal populations in awake monkey v1. *Journal of Vision*, 2015.

[91] David Kastner, Yusuf Ozuysal, Georgia Panagiotakos, and Stephen Baccus. Adaptation of inhibition mediates retinal sensitization. *Current Biology*, 29, 2019.

[92] Matthias Hennig. Theoretical models of synaptic short term plasticity. *Frontiers in Computational Neuroscience*, 7:154, 2013.

[93] H. O. Lancaster. Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, 44(1-2):289–292, 1957.

[94] Ryogo Kubo. Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957.

[95] R Kubo. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29(1):255–284, 1966.

[96] B. Cessac and J.A. Sepulchre. Stable resonances and signal propagation in a chaotic network of coupled units. *Phys. Rev. E*, 70(056111), 2004.

[97] B. Cessac and J.A. Sepulchre. Transmitting a signal by amplitude modulation in a chaotic network. *Chaos*, 16(013104), 2006.

[98] David Ruelle. Nonequilibrium statistical mechanics near equilibrium: computing higher-order terms. *Nonlinearity*, 11(1):5–18, 1998.

[99] O. Marre, S. El Boustani, Y. Frégnac, and A. Destexhe. Prediction of spatiotemporal patterns of neural activity from pairwise correlations. *Phys. rev. Let.*, 102:138101, 2009.

[100] Hassan Nasser, Olivier Marre, Michael Berry, and Bruno Cessac. Spatio temporal gibbs distribution analysis of spike trains using monte carlo method. In *AREADNE 2012 Research in Encoding And Decoding of Neural Ensembles*, 2012.

[101] Hassan Nasser and Bruno Cessac. Parameters estimation for spatiotemporal maximum entropy distributions: application to neural spike trains. *Entropy*, 16(4):2244–2277, 2014.

[102] Shun-Ichi Amari and Hiroshi Nagaoka. *Methods of Information Geometry*, volume 191. Oxford, 2000.

[103] C.R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.

[104] Shun-Ichi Amari. Information geometry in optimization, machine learning and statistical inference. *Frontiers of Electrical and Electronic Engineering in China*, 5:241–260, 2010.

[105] D. Ruelle. Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics. *J. Statist. Phys.*, 95:393–468, 1999.

[106] B. Cessac. *Recent Trends in Chaotic, Nonlinear and Complex Dynamics, in honour of Prof. Miguel A.F. Sanjuán on his 60th Birthday.*, chapter The retina as a dynamical system. World Scientific, 2020.

[107] Corie Lok. Curing blindness: Vision quest. *Nature*, 513:160, 2014.

[108] B.W. Jones, M. Kondo, H. Terasaki, Y. Lin, M. McCall, and R.E. Marc. Retinal remodeling. *Jpn J Ophthalmol*, 56(4):289–306, 2012.

[109] R. E. Marc, B. W. Jones, C. B. Watt, and E. Strettoi. Neural remodeling in retinal degeneration. *Progress in retinal and eye research*, 22(5):607–655, 2003.

[110] J.M. Barrett, P. Degenaar, and E. Sernagor. Blockade of pathological retinal ganglion cell hyperactivity improves optogenetically evoked light responses in rd1 mice. *Frontiers in cellular neuroscience*, 9:330, 2015.

[111] J. Barrett, G. Hilgen, and E. Sernagor. Dampening spontaneous activity improves the light sensitivity and spatial acuity of optogenetic retinal prosthetic responses. *Sci Rep*, 6:33565, 2016.

[112] S. Roux, F. Matonti, F. Dupont, L. Hoffart, S. Takerkart, S. Picaud, P. Pham, and F. Chavane. Probing the functional impact of sub-retinal prosthesis. *eLife*, 5(e12687), 2016.

[113] Pascale Pham, Sébastien Roux, Frédéric Matonti, Florent Dupont, Vincent Agache, and Frédéric Chavane. Post-implantation impedance spectroscopy of subretinal micro-electrode arrays, OCT imaging and numerical simulation: towards a more precise neuroprosthesis monitoring tool. *Journal of Neural Engineering*, 10(4):046002, 2013.

[114] W. Al-Atabany, B. McGovern, K. Mehran, R. Berlinguer-Palmini, and P. Degenaar. A Processing Platform for Optoelectronic/Optogenetic Retinal Prosthesis. *IEEE Transactions on Biomedical Engineering*, 60(3):781–791, March 2013.

[115] Niru Maheswaranathan, Lane McIntosh, David B. Kastner, Joshua B. Melander, Luke Brezovec, Aran Nayebi, Julia Wang, Surya Ganguli, and Stephen A. Baccus. Deep learning models reveal internal structure and diverse computations in the retina under natural scenes. *bioRxiv*, page 340943, 2018.

[116] 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 Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in*

Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, pages 8535–8545, 2019.

[117] Yajing Zheng, Shanshan Jia, Zhaofei Yu, Jian K. Liu, and Tiejun Huang. Unraveling neural coding of dynamic natural visual scenes via convolutional recurrent neural networks. *Patterns*, 2(10):100350, 2021.

[118] C. Gias, N. Hewson-Stoate, M. Jones, D. Johnston, J.E. Mayhew, and P.J. Coffey. Retinotopy within rat primary visual cortex using optical imaging. *NeuroImage*, 24(1):200–206, 2005.

[119] Mark M. Schira, Christopher W. Tyler, Branka Spehar, and Michael Breakspear. Modeling magnification and anisotropy in the primate foveal confluence. *PLOS Computational Biology*, 6(1):1–10, 2010.

[120] Inbal Ayzenshtat, Ariel Gilad, Guy Zurawel, and Hamutal Slovin. Population response to natural images in the primary visual cortex encodes local stimulus attributes and perceptual processing. *The Journal of Neuroscience*, 32:13971 – 13986, 2012.

[121] S. Amari. Characteristics of randomly connected threshold element networks and neural systems. *Proc. IEEE*, 59:35–47, 1971.

[122] H.R. Wilson and J.D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12:1–24, 1972.

[123] H.R. Wilson and J.D. Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biological Cybernetics*, 13(2):55–80, September 1973.

[124] P.C. Bressloff. Traveling fronts and wave propagation failure in an inhomogeneous neural network. *Physica D: Nonlinear Phenomena*, 155(1-2):83–100, 2001.

[125] P.C. Bressloff, J.D. Cowan, M. Golubitsky, P.J. Thomas, and M.C. Wiener. Geometric visual hallucinations, euclidean symmetry and the functional architecture of striate cortex. *Phil. Trans. R. Soc. Lond. B*, 306(1407):299–330, March 2001.

[126] P.C. Bressloff. Stochastic neural field theory and the system-size expansion. *SIAM J. Appl. Math*, 70:1488–1521, 2009.

[127] S. ElBoustani and A. Destexhe. A master equation formalism for macroscopic modeling of asynchronous irregular activity states. *Neural computation*, 21(1):46–100, 2009.

[128] Bruno Cessac, Selma Souihel, Matteo Di Volo, Frédéric Chavane, Alain Destexhe, Sandrine Chemla, and Olivier Marre. Anticipation in the retina and the primary visual cortex : towards an integrated retino-cortical model for motion processing. In *Workshop on visuo motor integration*, Paris, France, June 2019.

[129] Edward Ott and Thomas M. Antonsen. Low dimensional behavior of large systems of globally coupled oscillators. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18(3):037113, 2008.

[130] E. Montbrió, D. Pazó, and Alex Roxin. Macroscopic description for networks of spiking neurons. *Physical Review X*, 5:021028, 2015.

[131] C. Bick, M. Goodfellow, C.R. Laing, and Erik A. Martens. Understanding the dynamics of biological and neural oscillator networks through exact mean-field reductions: a review. *J. Math. Neurosc.*, 10(9), 2020.

[132] Hongjie Bi, Marco Segneri, Matteo di Volo, and Alessandro Torcini. Coexistence of fast and slow gamma oscillations in one population of inhibitory spiking neurons. *Phys. Rev. Research*, 2:013042, 2020.

[133] Victor Buendía, Pablo Villegas, Raffaella Burioni, and Miguel A. Muñoz. Hybrid-type synchronization transitions: Where incipient oscillations, scale-free avalanches, and bistability live together. *Phys. Rev. Research*, 3:023224, 2021.

[134] Matteo Di Volo, Marco Segneri, Denis Goldobin, Antonio Politi, and Alessandro Torcini. Coherent oscillations in balanced neural networks driven by endogenous fluctuations, 2021.

[135] Caterina Mazzetti. *A mathematical model of the motor cortex*. PhD thesis, Bologna, 2017.

[136] Davide Barbieri, Giovanna Citti, Giacomo Cocci, and Alessandro Sarti. A cortical-inspired geometry for contour perception and motion integration. *Journal of Mathematical Imaging and Vision*, 49(3):511–529, 2014.

[137] A. Romagnoni, J. Ribot, D. Bennequin, and J. Touboul. Parsimony, exhaustivity and balanced detection in neocortex. *PLOS Computational Biology*, 11(11):1–17, 2015.

[138] J. Rankin and F. Chavane. Neural field model to reconcile structure with function in primary visual cortex. *PLOS Computational Biology*, 13(10):1–30, 2017.

[139] Rachel Nicks, Abigail Cocks, Daniele Avitabile, Alan Johnston, and Stephen Coombes. Understanding sensory induced hallucinations: From neural fields to amplitude equations. *SIAM Journal on Applied Dynamical Systems*, 20(4):1683–1714, 2021.