UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

Luisa Fernanda Ramirez Ochoa

VERTEBRATE NEURONAL NETWORKS: Efficient coding and compressibility of interactions in the retina and the hippocampus

BELO HORIZONTE
2022

Luisa Fernanda Ramirez Ochoa

VERTEBRATE NEURONAL NETWORKS: Efficient coding and compressibility of interactions in the retina and the hippocampus

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Ronald Dickman

BELO HORIZONTE

2022

Dados Internacionais de Catalogação na Publicação (CIP)

R173v Ramirez Ochoa, Luisa Fernanda.
Vertebrate neuronal networks: efficient coding and compressibility of interactions in the retina and the hippocampus / Luisa Fernanda Ramirez Ochoa. – 2022.
107f. : il.

Orientador: Ronald Dickman.
Tese (doutorado) – Universidade Federal de Minas Gerais, Departamento de Física.
Bibliografia: f. 82-93.

1. Retina. 2. Hipocampo. I. Título. II. Dickman, Ronald. III. Universidade Federal de Minas Gerais, Departamento de Física.

CDU – 611.843 (043)

Ficha catalográfica elaborada por Romário Martins Ribeiro – CRB6 3595
Biblioteca Professor Manoel Lopes de Siqueira – Departamento de Física - UFMG

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

ATA DE DEFESA DE TESE

ATA DA SESSÃO DE ARGUIÇÃO DA 406ª TESE DO PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA, DEFENDIDA POR LUISA FERNANDA RAMIREZ OCHOA orientada pelo professor Ronald Dickman, para obtenção do grau de **DOUTORA EM CIÊNCIAS, área de concentração Física**. Às 11 horas de primeiro de dezembro de dois mil e vinte e dois reuniu-se, por videoconferência, a Comissão Examinadora, composta pelos professores **Ronald Dickman** (Orientador - Departamento de Física/UFMG), **Lucas Lages Wardil** (Departamento de Física/UFMG), **Mauro Copelli Lopes da Silva** (Departamento de Física/UFPE), **William Bialek** (Princeton University) e **Stephanie Palmer** (University of Chicago) para dar cumprimento ao Artigo 37 do Regimento Geral da UFMG, submetendo a Mestre **LUISA FERNANDA RAMIREZ OCHOA** à arguição de seu trabalho de Tese de Doutorado, que recebeu o título de **"Vertebrate neuronal networks: Efficient coding and compressibility of interactions in the retina and the hippocampus"**. A candidata fez uma exposição oral de seu trabalho durante aproximadamente 50 minutos. Após esta, os membros da comissão prosseguiram com a sua arguição, e apresentaram seus pareceres individuais sobre o trabalho, concluindo pela aprovação da candidata.

Belo Horizonte, 01 de dezembro de 2022.

Prof. Ronald Dickman
Orientador do estudante
Departamento de Física/UFMG

Prof. William Bialek
Princeton University

Prof. Lucas Lages Wardil
Departamento de Física/UFMG

Prof. Stephanie Palmer
University of Chicago

Prof. Mauro Copelli Lopes da Silva
Departamento de Física/UFPE

Candidata: Luisa Fernanda Ramirez Ochoa

Documento assinado eletronicamente por **Stephanie Elizabeth Palmer**, **Usuária Externa**, em 01/12/2022, às 14:01, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Ronald Dickman**, **Professor do Magistério Superior**, em 01/12/2022, às 14:12, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Luisa Fernanda Ramirez Ochoa**, **Usuária Externa**, em 02/12/2022, às 06:24, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Mauro Copelli Lopes da Silva**, **Usuário Externo**, em 02/12/2022, às 16:56, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Lucas Lages Wardil**, **Professor do Magistério Superior**, em 12/01/2023, às 10:32, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **William Bialek**, **Usuário Externo**, em 08/02/2023, às 00:45, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1936468** e o código CRC **8FAB35E0**.

Acknowledgements

I begin by acknowledging my parents, Angela and Fernando, and my brother Jaime for their love since always. To my mother, thanks for being an example of discipline ans perseverance, to my father, thanks for being an example of modesty and goodness, and to my brother, thanks for being example of trust and loyalty. I thank you for the strength and tenacity during this period at long distance, thanks for being always in the airport with a balloon, thanks for the giant fruit salads and thanks for the chats and challenges about algorithms. Without you, my PhD would have not been possible.

I thank my advisor Ronald Dickman for the discussions and lessons during my scientific path. His understanding of physics and art always was and continues to be, a motivation to keep learning. I thank William Bialek for all the scientific discussions and for allowing me to be part of the Princeton biophysics community. I also thank Tom Baden for his outstanding work in making science accessible at all levels, allowing me to develop my work on his group's scientific contributions.

I want to give special thanks to Ricardo Martinez-Garcia, my life partner, for his love and support that have been key pieces to go through the ups and downs of this stage. Thank you for your infinite ideas that filled my days of joy, and thanks for the temper when listening and helping me to solve my multiple tempests in teapots. I thank to my friends Jenifer Salguero, Jaen Barbosa, Rafael Rojas, and Raul Rincón for many years of unconditional friendship. To Jennifer and Jaen, thanks to the invaluable gift of making me feel always at home. To Rafa and Raúl, my friends with enviable hair, thanks for being part of this brazilian experience, full of cachaza, music and long conversations. I thank my colleagues and friends Leonardo Calazans, Marianne Bauer, Xiaowen Chen, Kamesh Krishnamurthy, Kevin Liu, and all other members of my research group for the always exciting academic discussions. I thank all cultural exchanges with everybody I met in Brazil and the United States. Thanks for the afternoons of parties, samba and Forró.

I acknowledge the UFMG and especially the professors Paulo Sergio, Mario Mazzoni, Carlos Monken, and Sebastiao Pádua, for their work in teaching and mentoring me in the core courses of my PhD program. Finally, I thank Capes and Cnpq for the financial support that made this work possible. I thank ICTP-SAIFR and Princeton University for the financial support that made my international research experience possible. I also thank other institutions and universities that supported my participation in both national and international schools, workshops, and conferences.

Agradecimientos

Comienzo dando mis más profundos agradecimientos a mis padres Angela y Fernando y a mi hermano Jaime, por el amor brindado desde siempre. A mi mamá, gracias por ser ejemplo de disciplina y perseverancia, a mi papá, gracias por ser ejemplo de humildad y bondad y a mi hermano, gracias por ejemplo de confianza y lealtad. Les agradezco la fuerza y tenacidad en este largo tiempo a distancia, gracias por esperarme siempre con un globito en el aeropuerto, gracias por las gigantes ensaladas de fruta y gracias por las charlas y desafíos sobre algorítmos. Sin ustedes, mi doctorado no habría sido posible.

Agradezco a mi orientador Ronald Dickman por las discusiones y enseñanzas a lo largo de mi periodo de formación. Su comprensión de la física y el arte siempre fue, y sigue siendo, una motivación para continuar aprendiendo. Agradezco a William Bialek por todas las discusiones científicas y por haberme dado la oportunidad de pertenecer a la comunidad de biofísica de Princeton. También agradezco a Tom Baden por su gran labor de hacer la ciencia accesible a todos los niveles, permitiendome desarrollar parte de mi trabajo como una continuación a sus contribuciones científicas

Quiero agradecer en especial a Ricardo Martinez-Garcia, mi compañero de vida, por el amor y apoyo que han sido piezas claves para superar los altibajos de esta etapa. Gracias por tus infinitas ocurrencias que han llenado mis días de alegría y carcajadas, y gracias por el temple que siempre has tenido para escuchar y ayudarme a resolver mis múltiples tormentas en vasos de agua. Les agradezco a mis amigos Jenifer Salguero, Jaen Barbosa, Rafael Rojas y Raúl rincón por años de amistad incondicional. A Jenifer y Jaen, gracias por el invaluable regalo de hacerme sentir siempre en casa. A Rafa y a Raúl, mis amigos de cabello envidiable, gracias por hacer parte de esta experiencia brasileña, llena de cachaza, música y eternas discusiones. Les agradezco a mis amigos y colegas Leonardo Calazans, Marianne Bauer, Xiaowen Chen, Kamesh Krishnamurthy, Kevin Liu y otros miembros de mi grupo de investigación for las siempre estimulantes discusiones académicas. Agradezco a todos aquellos que pasaron por mi vida en Brasil y Estados Unidos, permitiendo un intercambio cultural. Gracias por las tardes de reuniones, samba y forró.

Agradezco a la UFMG y especialmente a los profesores Paulo Sergio, Mario Mazzoni, Carlos Monken y Sebastiao Pádua, por su labor como mentores en los cursos de formación básica. Finalmente, agradezco a las instituciones de fomento Capes y Cnpq por el apoyo financiero que hizo posible este trabajo. Agradezco al ICTP-SAIFR y a la universidad de Princeton por el apoyo financiero que hizo posible mi periodo de investigación internacional. Agradezco a otras instituciones y entidades que apoyaron financieramente mi participación en escuelas, workshops y conferencias a nivel nacional e internacional.

Abstract

Our success in understanding the retina is partially due to its layered structure, that facilitates the study of circuit motifs and neuronal function. We can understand retinal information processing in three broad stages: I. encoding of light stimuli via electrical signals; II. signal processing by retinal circuits; and III. generation of the retinal code. In this thesis, I focus in the study of the first and third stages from an information-theory perspective.

On the first stage, we investigate color coding in zebrafish retinal circuits based on recent experimental findings showing evidence of efficient coding. We propose a theoretical framework to study the encoding performance of different types of outer retinal networks, contrasting the role of excitation and inhibition. More specifically, we use a neuronal population model with chromatic stimulation to study the dynamical properties of such networks. Our findings suggest that inhibition plays a key role in encoding color information reliably, which is not guaranteed in networks with strong excitatory inter-cone couplings. Similarly, we find that networks optimized to encode aquatic spectral information are similar to that observed in zebrafish, providing more general understanding of zebrafish-like retinal circuits of color coding. These results provide quantitative evidence that the zebrafish retina is adapted to efficiently encode information from the environment, enhancing this animal's color vision capabilities.

Studies in other species show that animals can adopt different strategies to improve color vision. For instance, in birds and turtles oil droplets serve as a filter to provide a plethora of distinguishable colors. Oil droplets and adapted retinal circuits have been investigated separately. Nevertheless, studies on the combination of both remain unknown. We implement a light transmission model of droplets to investigate the encoding performance of zebrafish-like retinal circuits exhibiting efficient coding. Our findings suggest that introducing droplets in a circuit for chromatic efficient coding creates a trade-off between coding efficiency and color-space gamut. That is, droplets decrease the network encoding performance while increasing the number of distinguishable colors.

Regarding the third stage of processing, we focus on the theoretical study of neuronal interactions in the ganglion layer and their compressibility as a path to building simpler models of neuronal activity. Conventional models of neuronal activity introduce assumptions about neural interactions inspired in condensed-matter systems. But these models fail when the number of neurons increases, leading to an exponential explosion in the number of parameters. Here, we implement information theory and renormalization group ideas to explore efficient descriptions of neuronal activity. More specifically, we apply the compression-bottleneck formalism to a population of ganglion cells in the salamander retina. We find that compression leads to a vast simplification in the description of neuronal

activity, outperforming conventional pairwise-interaction models. As a generalization, we implement this approach in a population of hippocampus neurons, yielding broadly similar results, suggesting that compressibility is a general feature of spiking neuronal networks.

Keywords: Neuronal code; Visual system; Retina; Efficient coding.

Resumo

O nosso sucesso no entendimento da retina se deve em parte à estrutura de camadas que facilita o estudo de diferentes tipos de circuitos e funções neuronais. O processamento de informação na retina pode ser entendido em três etapas: I. Codificação de estímulos visuais em sinais elétricos. II. Processamento desses sinais por circuitos na retina. III. Geração do código retinal. O meu trabalho de tese está focado no estudo da primeira e última etapa desde uma perspectiva da teoria da informação.

Na primeira etapa, investigamos a codificação de cor pelos circuitos retinais do zebrafish, baseados em resultados experimentais recentes que mostram evidencia de codificação eficiente. Nós propormos um modelo teórico para estudar codificação em diferentes tipos de redes retinais nas camadas externas, contrastando o papel da excitação e inibição. Os nossos resultados sugerem que a inibição joga um rol importante na codificação reliable de cor, o que não é garantido em redes com fortes interações excitatórias entre cones. De forma similar, os nossos resultados mostram que as redes otimizadas para codificar informação espectral em ambientes aquáticos são similares a aquelas observadas no zebrafish, validando os resultados experimentais e provendo um entendimento mais geral de circuitos retinais para codificação de cor com estruturas similares a aquelas do zebrafish. Estes resultados sugerem que a retina do zebrafish poderia ter se adaptado para fazer uma codificação eficiente da informação do ambiente, melhorando as capacidades de visão colorida.

Estudos em outras espécies mostram que os animais podem adoptar diferentes estratégias para melhorar a visão colorida. Por exemplo, em pássaros e tartarugas, a existência de gotas oleosas serve como um filtro para expandir a quantidade de cores distinguíveis. Gotas oleosas e circuitos retinais adaptados tem sido investigados de forma separada. Porém estudos sobre a combinação dos dois continuam desconhecidos. Nós implementamos um modelo de transmissão de luz das gotas para investigar a codificação de circuitos similares a aqueles do zebrafish que mostram uma codificação eficiente. Nossos resultados sugerem que a introdução de gotas oleosas naqueles circuitos gera um trade-off entre codificação eficiente e espaço de cor. Isto é, gotas oleosas diminuem a qualidade de codificação da rede enquanto expandem o numero de cores potencialmente distinguíveis pelo animal.

Relacionado à última etapa de processamento retinal, o nosso trabalho se foca no estudo teórico da atividade neuronal na camada ganglionar e a compressibilidade como um caminho para construir modelos simples de atividade neuronal. Modelos convencionais de atividade neuronal fazem suposições sobre interações neuronais inspiradas em sistemas de matéria condensada. Mas, esses modelos falham quando o número de neurônios aumenta, levando a um aumento exponencial no número de parâmetros. Neste trabalho implementamos ideias

de teoria de informação de grupo de renormalização para procurar descrições eficientes da atividade neuronal. Mais especificamente, aplicamos o formalismo de compression-bottleneck numa população de células ganglionares na retina da salamandra. Os nossos resultados mostram que a compressão leva a uma grande simplificação da descrição da atividade neuronal, ultrapassando os resultados de modelos convencionais tais como o de interação entre pares. Como uma generalização, implementamos o nosso modelo de compressão numa população de neurônios do hipocampo, achando de forma geral os mesmos resultados. Isto leva a concluir que a compressão é uma característica geral de redes neuronais de spikes.

Palavras-chave: Código neuronal; Sistema visual; Retina; Codificação eficiente.

Contents

Part I

Introduction

From an anthropomorphic perspective, the visual system represents one of the most important sensory mechanisms in nature. Early efforts to grasp the principles behind vision started with psychophysical experiments and anatomical studies of the retina, leading to a greater understanding of perception and the neuronal basis for visual stimuli processing. Similarly, functional studies of ganglion cells (GCs) in the inner retina, allowed a detailed characterization of the neuronal processing of visual features, such as spatial or temporal patterns [1–6]. Remarkably, the characterization of different types of retinal neurons (approximately 100 in vertebrates) [7, 8], as well as the identification of specific circuit motifs, such as the ON-OFF pathways [9–12], has led to a solid retinal model of coding and processing of specific visual stimuli. On one hand, such a model has shown high performance at predicting neuronal responses to visual stimuli [13, 14], demonstrating our good understanding of retinal information processing. On the other hand, it raises questions, from an evolutionary perspective, as to why retinas have evolved such specific features. The efficient coding theory has been quite successful at explaining, from an information theory perspective, some of these features [15–21]. However, we still lack answers to fundamental questions that seem to be consistent across species, such as the existence of about a hundred of different types of retinal neurons.

The success in our understanding of the retina is partially due to its organized and layered structure that allows a systematic study of the circuit motifs and neuronal diversity. Broadly, we can divide the retina information processing into three stages: I. Codification of visual information into electrical signals; II. signal processing by retinal circuits, and III. generation of the retinal code. The first is done by photoreceptors (cones and rods) and horizontal cells (HCs), the second by HCs, bipolar cells (BCs), and amacrine cells (As) and the last by GCs. This thesis is focused in the study of the first and last stages, both from an information theory perspective.

For the first stage, I focused in the study of photoreceptor codification from an efficient coding perspective. Specifically, in the circuits in charge of chromatic codification, that is, circuits that differentiate spectral features in the stimulus. Several studies have linked the spatial distribution of photoreceptors in some species with the spatial distribution of spectral information in their typical environments, supporting ideas of coding efficiency [16, 22]. In this work, I investigate the encoding of such spectral information by retinal circuits in the outer layer, showing that typical neuronal architectures can also be explained by coding efficiency ideas.

For the last stage, I focused in the theoretical study of neuronal interactions and their plausible compressibility as a path to building simpler models of neuronal activity. In the last decade, we have seen a remarkable advance in experimental techniques that allow us to observe the simultaneous activity of thousands of neurons, even across different brain regions [23–26]. Nevertheless, we still lack a solid theoretical framework to understand such

huge amounts of data. Conventional models of neuronal activity, inspired in condensed matter systems, make assumptions on neuronal interactions. Nevertheless, these models fail when the number of neurons dramatically increases, leading to an exponential explosion in the number of parameters. Here, we implement ideas of information theory to explore efficient approaches to describe neuronal activity accurately. More specifically, we use the compression bottleneck formalism as a means to investigate compressibility in a population of GCs in the salamander retina. As a generalization, we implement this approach in a population of hippocampus neurons, yielding broadly similar results, suggesting that compressibility is a general feature of spiking neuronal networks.

Part II

1 Efficient coding in the outer retina

[1]Visual stimuli are discriminated by specific features such as contrast, color or motion. In retinas, neurons make specific synaptic connections in a layered structure, forming specialized circuits in charge of processing each of these features [28–30]. For many species, color represents a crucial cue to thrive in nature [31–33]. For instance, for animals living in dense forests, color allows the quick discrimination of mature fruits in a complex background (see Fig.1a and ref. [34]). For other species, such as butterflies, color seems to play a key role in crucial behaviors, such as mating [31, 35].

The study of color processing in the retina has a long tradition, leading to a deep understanding of the biological mechanisms behind it. Cones are the main photoreceptor neurons in charge of codifying this feature (for more details on the biological structure of the retina see Appendix B). Depending on the species, the retina can have from one to twelve different type of photoreceptors, characterized by their spectral response functions (see Fig. 1b and ref. [35]).

Discriminating colors requires at least two photoreceptor types with different spectral responses. The reason is that a single photoreceptor is unable to differentiate color from intensity. As an example, consider the situation shown in Fig. 1c. The black curve represents the sensitivity function of a photoreceptor and the green and orange bars represent two stimuli, S_1 and S_2, with different intensities. The photoreceptor responses, R_1 and R_2, to these stimuli (white circles) is indistinguishable since it is proportional to the sensitivity function multiplied by the stimuli spectral function.

Visual systems with more than one photoreceptor type can discriminate colors by using principles of color-opponency [36,37], in which photoreceptor responses are subtracted, eliminating information of the stimulus intensity. Several experiments have provided evidence of the biological implementation of such opponency in retinal ganglion cells and downstream neuronal circuits via receptive fields [38, 39]. More recent experiments [40, 41] have shown evidence of early color opponency in the photoreceptor layer of certain species. Here, we focus on zebrafish, a model organism of visual neuroscience. First, we propose a theoretical model to investigate the neuronal dynamics of typical retinal architectures in charge of chromatic encoding as a function of their synaptic couplings. Second, we analyze our findings in the context of coding efficiency, by introducing the statistics of chromatic information available in zebrafish environments. Finally, we compare zebrafish photoreceptors with other possible combinations to test how efficient is this species in

[1] **Remark:** Some results, text and figures are from the paper *Data-driven models of optimal chromatic coding in the outer retina*, written in collaboration with Ronald Dickman. The article has available in the preprint server bioRxiv [27].

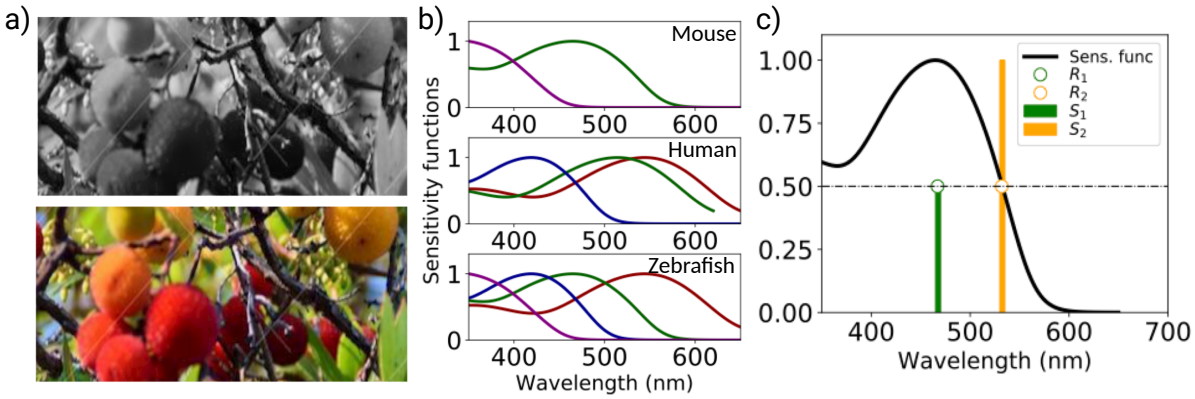

Figure 1 –

a) Image of fruits in a forest background. Grey scale (above) and colored (below). b) Spectral sensitivity curves of mouse (above), human (middle) and zebrafish (below). c) Photoreceptor responses, R_1, R_2, to the two stimuli, S_1, S_2 with intensities 0.5 and 1 and characteristic wavelengths 470nm and 530nm respectively.

codifying chromatic information.

2 Efficient chromatic coding in retinal networks: zebrafish

Coding efficiency theory postulates that sensory circuits evolve and organize to optimize the encoding and transmission of external information. The formalization of these ideas started with Barlow's hypothesis on redundancy reduction [42], where the definition of redundancy was equal to the one introduced by Shannon [43, 44], that is, the amount of channel capacity, C, that remains unused when transmitting certain message with entropy H in a certain time period T. The formalization of this idea for noiseless binary channels – neurons– suggested that an efficient strategy to minimize such redundancy and optimize the encoding process was to reduce spiking activity, such that it is mostly used to codify rare events in nature. Further efforts to model more realistic scenarios lead to the formalization of these ideas for noisy channels, allowing important predictions on neuronal features such as receptive fields [20, 45].

As mentioned, these works were based in the definition of redundancy by Shannon. Nevertheless, some experimental and theoretical works evidenced the limitations of this definition applied to the brain [46, 47]. For instance, Simoncelli and others [48] investigated the statistical structure of naturalistic images, evidencing a large characteristic redundancy which explains why humans, for instance, are good at visual-prediction tasks[1] [49]. Similarly, work by Atick, Field, Bialek and others [49–53] demonstrated that to capture these redundancies typically found in nature is necessary to use joint probabilities between symbols –neurons–, accounting for higher order correlations. In a revised work [54], Barlow discusses and integrates all these works, redefining the ideas of efficient coding in a Bayesian probabilistic framework. He proposes that the goal of coding efficiency is still to identify these *high-order* redundancies, encoding them efficiently to make more accurate and efficient predictions about the world.

Although coding efficiency theory is still in progress, its prediction power in sensory systems is notable [14, 18, 19, 28, 55]. For our purposes, we will adapt these ideas hypothesizing that neuronal circuits evolve and adapt to efficiently extract redundancies from their external world while preserving all relevant bits of information. For example, some works [16, 22] have shown that the spatial distribution of the spectrum in naturalistic images typical of species such as mice is largely structured, exhibiting a gradient from ultraviolet to green in the vertical axis (see Fig. 2a). Remarkably, experiments in the outer layer of the mouse retina, evidence that the spatial distribution of photoreceptors

[1] An example of one of these tasks is guessing the missing pixels of an incomplete image. Experiments [49] show that humans have a good performance if using naturalistic images, otherwise their prediction power quickly decreases.

(UV- and green-sensitive) also follows such gradient, suggesting that retinas have evolved to optimize the encoding of this natural spectral redundancy, as predicted by coding efficiency theory [16, 22]. Other species, such as zebrafish, exhibit similar heterogeneity in their photoreceptor spatial distribution, suggesting similar matches to their specific environmental conditions [16, 22].

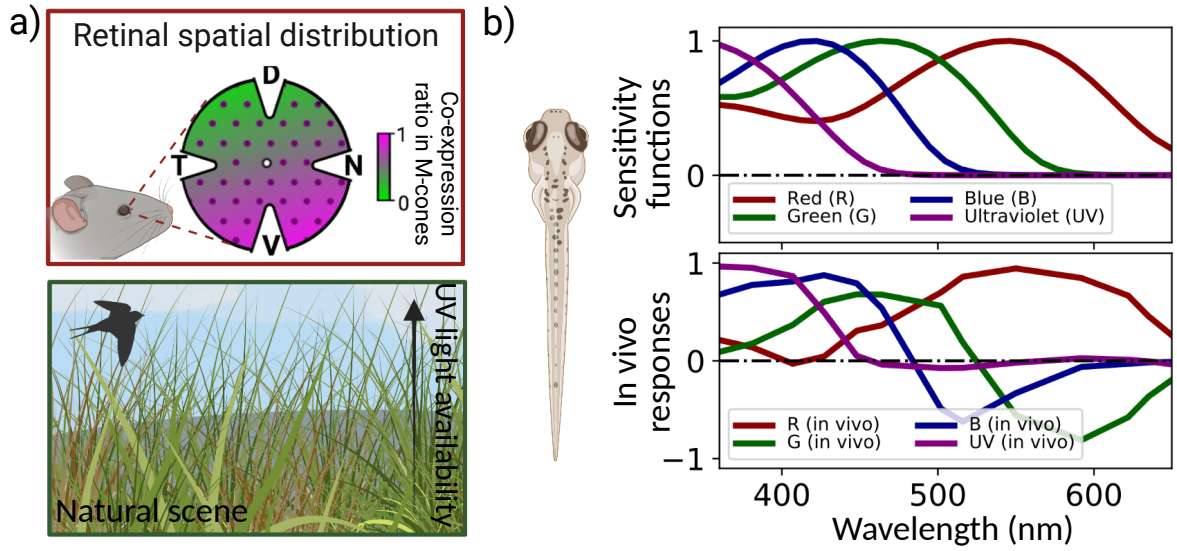

Figure 2 –

a) Above: spatial distribution of co-expression ratio of M-cones in the mouse retina; V: ventral, D: dorsal, T: temporal and N: nasal. Below: Sketch of a typical natural scene from the mouse view field. UV light in upper regions is more abundant than in lower regions. b) Zebrafish photoreceptor responses. Above: sensitivity curves of red-, green-, blue- and ultraviolet-sensitive cones. Below:*in vivo* cone responses (see ref. [40])

These experiments show evidence of efficient coding in the organization of retinal circuits. From a functional perspective, the question is whether information in the stimuli is efficiently encoded by retinal circuits. Recent experiments in zebrafish larva [40] show the *in vivo* activity of photoreceptors in response to chromatic stimuli. The upper plot of Fig. 2b shows the typical sensitivity curves of the four zebrafish photoreceptor types, ultraviolet (UV), blue (B), green (G) and red (R) in isolation. The lower plot of Fig. 2b shows the *in vivo* activity of the same photoreceptors, evidencing that at this early stage, chromatic photoreceptor responses are already opponent, that is, inhibitory in some ranges of the spectrum and excitatory otherwise.

In this chapter, we investigate typical outer retinal circuits enabling the above-mentioned opponent signals. Specifically, we investigate the role of excitatory vs inhibitory feedback mechanisms in a broad set of network architectures. Subsequently, we investigate efficient coding in zebrafish by comparing these opponent photoreceptor responses with the statistics of natural chromatic stimuli. Specifically, we use natural spectral correlations by implementing a principal component analysis on hyperspectral data from zebrafish natural

images. Finally, we search for efficient retinal networks that match typical chromatic statistics of the zebrafish environment to quantify the efficiency of the zebrafish visual system. In this work, we use the experimental data of zebrafish photoreceptors from [56], and the hyperspectral data of zebrafish naturalistic images from [57, 58].

2.1 Neuronal population model of the outer retinal layer

Broadly, outer retinal circuits for color discrimination tasks have N types of photoreceptors and M types of horizontal cells (HCs), as sketched in fig. 3a. Photoreceptors connect to HCs, leading to an inhibitory feedback mechanism that modifies the originally independent photoreceptor responses. Networks with excitatory and inhibitory feedback have additional cone-cone connections, as indicated by the dashed connection lines of Fig. 3a. We denote networks with dominant inhibitory feedback and weak inter-cone connections as Type I; networks with strong excitatory inter-cone connections are denoted as Type II. In zebrafish, experimental observations show that there are $N = 4$ cone photoreceptor types and $M = 1$ HC type that contribute to color discrimination circuits. As shown in Fig. 2b, the original independent signals correspond to the sensitivity curves (above), while the modified responses correspond to the *in vivo* observations, mediated by one type of HC (below).

In retinas, photoreceptor types and their retinal connectivity determine the species-specific color space [35, 59]. For instance, human retinas have three types of cone photoreceptors, sensitive to long-, middle-, and short-wavelength stimuli, that shape a trichromatic visual system. Broadly, the number of cones in a vertebrate retina can be of the order of millions, which can be further categorized into N populations; one for each cone type. We can study the population dynamics of each population i via the average membrane potential, $h_i(t)$. That is,

$$\tau_i \frac{\partial h_i}{\partial t} = -h_i + h_0 \, I_G(t), \tag{2.1}$$

where we have defined the resting potential at zero. The variable h_0 on the right hand side (r.h.s.) corresponds to the membrane potential change when applying a unit of current, that is $h_0 = R_i \, u_0$; with R_i the membrane resistance and $u_0 = 1\text{mA}$ the current. The variable I_G corresponds to the dimensionless intensity of both internal and external currents in the circuit. In the outer retina, external currents come from light stimuli while internal currents come from synaptic connections with the other populations, that is,

$$I_G = I_S + \sum_j^N I_{ij}^{(E)} + \sum_j^M I_{ij}^{(I)}. \tag{2.2}$$

The second term on the r.h.s. corresponds to synaptic connections between cone-cone populations, which in the vertebrate retina are normally excitatory. The last term on

the r.h.s. corresponds to the cone-HC synaptic connections, which are normally inhibitory. Depending on the nature of the pre-synaptic population, excitatory or inhibitory, the post-synaptic population might exhibit a different response, $F_E[u]$ or $F_I[h]$ respectively, so that we can write Eq. 2.2 as,

$$I_G = I_S + \sum_j^N w_{ij} F_E[u] + \sum_j^M w_{ij} F_I[h], \tag{2.3}$$

where the parameters w_{ij} correspond to the synaptic connection strengths. We multiply both sides of Eq.2.1 by $(h_0)^{-1}$, such that we get,

$$\frac{\tau_i}{h_0}\frac{\partial h_i}{\partial t} = -\frac{h_i}{h_0} + I_G(t). \tag{2.4}$$

To model the response functions, we must consider photoreceptor responses as a function of the input intensity. Experimental works show that cone responses saturate at strong stimulus intensity [60]. To model this, we use the function,

$$F[h] = \tanh(\gamma h + \beta), \tag{2.5}$$

where γ and β are free parameters. Fig. 3b shows this function for three different sets of parameters. In the subsequent sections, we set $\gamma = 1$ and $\beta = 0$; we show later that changes in these values do not affect our conclusions.

Regarding the first term on the right hand of Eq. (2.2), we know that one way of characterizing visual stimuli is via the spectral density, $S(\lambda)$, which contains all relevant chromatic information. To calculate the isolated cone response intensity, $I_s^{(i)}$, to a stimulus with spectral density $S(\lambda)$, we integrate the product of $\Theta_i(\lambda)$, the corresponding sensitivity function, and $S(\lambda)$, over the spectrum:

$$I_s^{(i)} = \tanh\left(\int \Theta_i(\lambda)\, S(\lambda)\, d\lambda\right). \tag{2.6}$$

Determining responses to specific wavelength intervals requires stimuli with a narrow spectral distribution, conveniently represented by Gaussian distributions centered at a characteristic wavelength λ_0, and having a small standard deviation (see Fig. 3c), that is,

$$S(\lambda) = \alpha \exp\left(\frac{-(\lambda - \lambda_0)^2}{2\sigma^2}\right), \tag{2.7}$$

with α representing the normalized luminosity and σ the standard deviation. Fig. 3d shows the current curve of three photoreceptors as a function of different color pulses, with $\sigma = 1$nm and $\alpha = 0.5$. To evidence the current dependence on the stimuli free parameters,

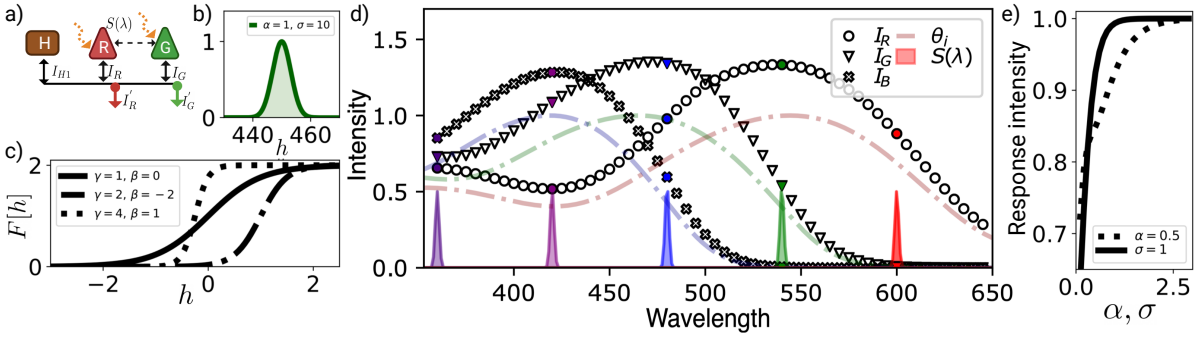

Figure 3 –

a) Sketch of an outer retinal network with N types of cone photoreceptors and M types of HCs. $S(\lambda)$ represents the external stimuli, while I_i and $I_i^{'}$ represent the independent and modified cone responses respectively. b) Example of a green gaussian-like stimulus with $\sigma = 10$, $\alpha = 1$ and $\lambda_0 = 450$nm. c) Response function $F[h] = \tanh(\gamma h + \beta)$ with three different sets of adjust parameters. d) Independent responses of photoreceptors to narrow gaussian stimuli. Dashed curves are red, green and blue zebrafish sensitivity curves and colored distributions correspond to five Gaussian stimuli normalized to exhibit a maximum intensity of $\alpha = 0.5$, with variance $\sigma = 1$nm. Markers correspond to the independent responses, described by Eq. 2.6, of the three photoreceptors. Colored markers show the response to the five plotted stimuli. e) Cone response as a function of the pulse standard deviation, σ (dashed line), and intensity, α (solid line).

we used a light pulse centered at $\lambda_0 = 500$nm, and calculated the current intensity $I_s^{(i)}$ as a function of σ and α. Fig 3e shows the current behavior for both σ (dashed line) and α (solid line). In the case of sigma, we observe saturation for $\sigma > 2$nm, while in the case of α, we observe saturation at unity. Considering these results as well as our goal of modeling responses to specific wavelengths, we use narrow standard deviation $\sigma = 1$nm. Regarding the intensity parameter, we fix the value $\alpha = 0.5$, such that we avoid any saturating behavior and we keep the same light intensity over all different stimuli pulses.

Replacing Eqs. (2.2) and (2.6) into Eq. (2.4), we get the population equations,

$$\tau_i \frac{\partial h_i}{\partial t} = -h_i + \tanh\left(\int \Theta_i(\lambda)\, S(\lambda)\, d\lambda\right) + \sum_j^N w_{ij} F_E[h] + \sum_j^M w_{ij} F_I[h], \text{ for } i = 1, \ldots, N, \tag{2.8}$$

where we have set h_0 to unity without loss of generality. For simplicity, in the subsequent text we refer to the variables h_i and I_0 as the membrane potential and the current respectively. But, the reader must remember that these quantities are dimensionless as follows from Eq. 2.8 and 2.4.

2.2 Dynamics of outer retinal networks

In this work, we focus in dichromatic, trichromatic and tetrachromatic networks, which are the most abundant across species. Nevertheless, it is worth to mention that this network model can be applied to more complex networks with a larger number of photoreceptors and HCs.

2.2.1 Dichromatic networks

As previously described, photoreceptors in dichromatic networks can be grouped into two main populations characterized by their sensitivity function. Similarly, horizontal cells can be grouped into one or more populations depending on their functional properties. We focus in the dichromatic network sketched in Fig 3a, composed of two excitatory populations, R and G, for red and green photoreceptors, and a single horizontal-cell population, H. Following Eq. 2.8, we get,

$$\tau_E \frac{\partial h_R}{\partial t} = -h_R + R_R \tanh\left(\int \Theta_R(\lambda)\, S(\lambda)\, d\lambda\right) + w_{RG} F_E[h_G] + w_{RH} F_I[h_H] \quad (2.9)$$

$$\tau_E \frac{\partial h_G}{\partial t} = -h_G + R_G \tanh\left(\int \Theta_G(\lambda)\, S(\lambda)\, d\lambda\right) + w_{GR} F_E[h_R] + w_{GH} F_I[h_H] \quad (2.10)$$

$$\tau_I \frac{\partial h_H}{\partial t} = -h_H + w_{HR} F_E[h_R] + w_{HG} F_E[h_G]., \quad (2.11)$$

where R_i is the membrane resistance, which we set to unity [61]. Since HCs do not interact directly with photons, the current terms contributing to the dynamics of the membrane potential, h_H, correspond to synaptic connections with cone populations. The last two terms on the r.h.s. of Eqs. (2.9) and (2.10) correspond to excitatory and inhibitory connections, respectively, with coupling parameters, w_{ij}, which are positive for excitatory currents and negative otherwise; the first (second) subscript denotes the postsynaptic (presynaptic) population.

The membrane time constants on the left hand side of Eqs. (2.9)-(2.11), τ_E for excitatory neurons and τ_I for inhibitory neurons, introduce two time scales that are related to neuron responses and the latency of the feedback mechanism. Some experimental works, reviewed in ref. [62], show the existence of two fast feedback mechanisms from HCs; ephaptic and proton-mediated feedback, highlighting the suitability of HC for tasks involving fast adjustment of cone responses. We take this into consideration by assuming that the membrane time constant of inhibitory neurons is much smaller than that of excitatory neurons, i.e., $\tau_I \ll \tau_E$, which allows us to simplify Eqs. (2.9)-(2.11) as,

$$\begin{aligned}
\frac{\partial h_R}{\partial t} &= -h_R + I_R + w_{RH} F_I\left[\, w_{HR}\, F_E[h_R] + w_{HG}\, F_E[h_G]\,\right] + w_{RG}\, F_E[h_G] \\
\frac{\partial h_G}{\partial t} &= -h_G + I_G + w_{GH} F_I\left[\, w_{HR}\, F_E[h_R] + w_{HG}\, F_E[h_G]\,\right] + w_{GR}\, F_E[h_R] \\
h_H &= w_{HR}\, F_E[h_R] + w_{HG}\, F_E[h_G] \qquad (2.12)
\end{aligned}$$

where we have set $\tau_E = 1$.

As previously mentioned, the experimental observations of zebrafish outer retinal layers lead us to ask whether networks of Type I (weak inter-cone connections) provide an advantage in terms of chromatic information encoding and transmission. As noted above, inter-cone gap-junctions and chemical synapses have been observed in some vertebrates; the question is whether they are involved in chromatic discrimination. The set of equations (2.12) provides a simplified framework to study the time evolution of population membrane potentials in dichromatic networks as a function of the synaptic strengths, allowing us to compare Type I networks, with only inhibitory feedback, and Type II networks, with strong inter-cone connections. Comparing the dynamics of the two types of network should provide insights into encoding performance.

A Type I network is characterized by weak inter-cone connections, that is, $w_{RG} = w_{GR} \approx 0$ in Eq. (2.12). This regime allows an analytic solution of the stationary state as a function of the four remaining coupling parameters, and all the chromatic stimuli illustrated in Fig. 3d, that is,

$$h_r^* = \frac{w_{RH}\, h_G^* + w_{GH}\, I_R - w_{RH}\, I_G}{w_{GH}}. \qquad (2.13)$$

From Eq. (2.13), we conclude that for any set of parameters, there is a unique stable fixed point in the two dimensional domain $\{h_r \times h_g, \in \mathbb{R}^2\}$. Figure 4a shows a typical phase portrait. We determine the stability of the fixed points via the Jacobian matrix, that is,

$$\begin{aligned}
J &= \begin{pmatrix} -1 + w_{RH}\, w_{HR} \operatorname{sech}^2 h_R\, \operatorname{sech}^2 D & \operatorname{sech}^2 h_G (w_{RH}\, w_{HG}\, \operatorname{sech}^2 D + w_{RG}) \\ \operatorname{sech}^2 h_R (w_{GH}\, w_{HR}\, \operatorname{sech}^2 D + w_{GR}) & -1 + w_{GH}\, w_{HG} \operatorname{sech}^2 h_G\, \operatorname{sech}^2 D. \end{pmatrix} \\
&= \begin{pmatrix} J_{11}^{(1)} + J_{11}^{(2)} & J_{12}^{(1)} + J_{12}^{(2)} \\ J_{21}^{(1)} + J_{21}^{(2)} & J_{22}^{(1)} + J_{22}^{(2)} \end{pmatrix}, \qquad (2.14)
\end{aligned}$$

with $D = \tanh(w_{HR}\, F_E[h_R] + w_{HG}\, F_E[h_G]) + 1$.

The trace of this matrix is negative for all coupling parameters. On the other hand, the determinant sign depends on the coupling parameter strengths. If Det < 0 the fixed

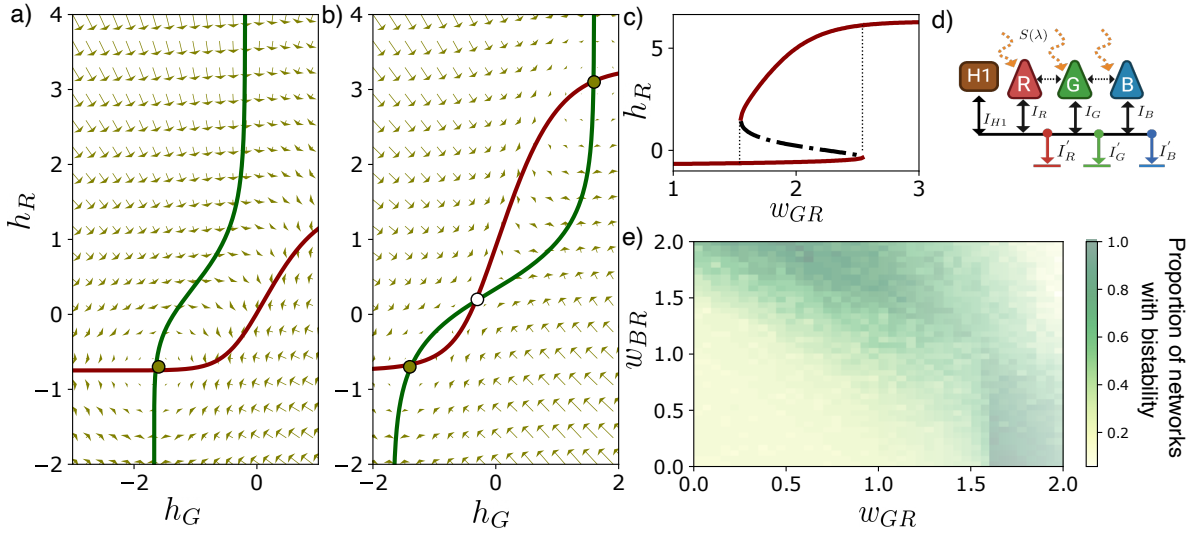

Figure 4 –

(a, b) Phase portraits of Eq. 2.12 for a dichromatic network with red and green photoreceptors, and with the parameters $w_{HR} = 1.5, w_{RH} = -1.7$, $w_{HG} = 0.9, w_{GH} = -1.1, S(\lambda) = \mathcal{N}(\lambda = 380, 1)$, and a) $w_{GR} = w_{RG} = 0$ corresponding to Type I, and b) $w_{GR} = w_{RG} = 1.8$, corresponding to a Type II network. c) Bifurcation diagram of h_R for the inter-cone coupling parameters $w_{GR} = w_{RG} = [1, 3.5]$. d) Trichromatic sketch of a fully connected network with an external stimulus, $S(\lambda)$. R: red cones, G: green cones, B:blue cones, H: horizontal cells. Solid black arrows represent inhibitory synaptic conections on cones from horizontal cells. Dashed black arrows represent excitatory synaptic connections between cones. e) Proportion of networks exhibiting multistability. Color intensity represents the normalized sum over the discrete interval $w_{RG} = [0.1, \ldots, 2.0]$, for a fixed combination of the excitatory parameters, w_{GR} and w_{BR}. All points correspond to the same combination of inhibitory parameters used before.

point is classified as a saddle point. Otherwise, it is a stable sink. The existence of a single fixed point implies that the stationary response to a stimulus $S(\lambda)$ is unique and independent of the initial state. Our results show that regardless of variations in the coupling strengths among neuronal populations, networks with only inhibitory feedback exhibit a unique response to a given chromatic stimulus, hence reliable encoding of chromatic stimuli. In the outermost retinal layers, this is desirable to avoid ambiguity in chromatic encoding and transmission.

We now ask whether such behavior persists in a Type II network. We calculate the fixed points of Eqs. (2.12) by determining the intersection of the nullclines minimizing the cost function,

$$\begin{aligned}\mathcal{L}^2 \;=\; & \left(-h_R + I_R + w_{RH} F_I\left[\, w_{HR}\, F_E[h_R] + w_{HG}\, F_E[h_G]\,\right] + w_{RG}\, F_E[h_G]\right)^2 \\ & + \left(-h_G + I_G + w_{GH} F_I\left[\, w_{HR}\, F_E[h_R] + w_{HG}\, F_E[h_G]\,\right] + w_{GR}\, F_E[h_R]\right)^2\end{aligned}$$

We calculate the solutions for all coupling parameters in the discrete space $w_{ij} \in [0.1, 0.2, \ldots, 10.0)$ for excitatory and $w_{ij} \in (-10.0, \ldots, -0.1, 0)$ for inhibitory parameters. This parameter range is sufficient to investigate the regimes of strong and medium excitatory connections, $w_{RH}/w_{RG} \ll 1$ and $w_{RG}/w_{RH} \approx 1$ respectively. As shown in Figures 4b-c, we find that, in contrast to Type I networks, Type II networks with strong inter-cone couplings (compared with HC-couplings), can exhibit three fixed points, two of them stable and one saddle node. Under this scenario, a photoreceptor may exhibit multiple responses to the same stimulus, which, as previously mentioned, could lead to misleading or ambiguous chromatic signals. Networks with other photoreceptor combinations, such as red-blue and green-blue, have similar equations and solutions. Fig. 5 shows that such combinations lead to similar phase portraits and dynamical properties, reinforcing our general conclusion.

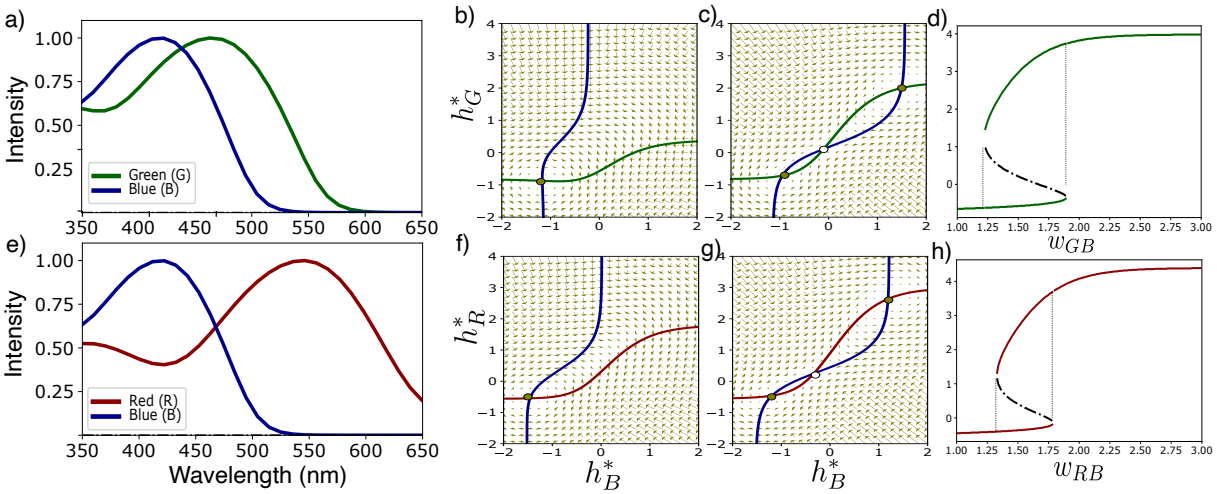

Figure 5 –

a) Green and blue zebrafish photoreceptor sensitivity curves. (b, c) Phase portraits of Eq. (2.12) for a dichromatic network with green and blue photoreceptors, and with the parameters a) $w_{HG} = 0.9, w_{GH} = -1.1$, $w_{HB} = 1.5$, $w_{BH} = -1.5, w_{BG} = 0$ and $S(\lambda) = \mathcal{N}(\lambda = 412, 1)$, corresponding to a Type I network, and b) $w_{HG} = 0.9, w_{GH} = -1.1$, $w_{HB} = 1.5$, $w_{BH} = -1.5, w_{BG} = 1.6$ and $S(\lambda) = \mathcal{N}(\lambda = 412, 1)$, corresponding to a Type II network. d) Bifurcation diagram as a function of the coupling parameter $w_{GB} = w_{BG}$. e) Red and blue zebrafish sensitivity curves. (f, g) Phase portraits of Eq. (2.12) for a dichromatic network with red and blue photoreceptors, with parameters a) $w_{HR} = 0.9$, $w_{RH} = -1.1$, $w_{HB} = 1.5$, $w_{BH} = -1.7$, $w_{RB} = 0$ and $S(\lambda) = \mathcal{N}(\lambda = 390, 1)$, corresponding to a Type I network, and b) $w_{HR} = 0.9$, $w_{RH} = -1.1$, $w_{HB} = 1.5$, $w_{BH} = -1.7$, $w_{RB} = 1.7$ and $S(\lambda) = \mathcal{N}(\lambda = 390, 1)$, corresponding to a Type II network. h) Bifurcation diagram as a function of the coupling parameter $w_{RB} = w_{BR}$.

2.2.2 Trichromatic networks

The preceding study of networks with two populations of photoreceptors sensitive to different spectral ranges, allows extrapolation of our results to other species with dichro-

matic retinas. Generalizing these results to species with more complex visual systems, however, is not straightforward. We advance in this direction by investigating trichromatic systems, expanding considerably the diversity of color-vision systems. The trichromatic networks studied here include long- (R), middle- (G) and short-wavelength (B) photoreceptors, and a single horizontal cell population (H) that provides inhibitory feedback to cone populations (see Fig.5d). Assuming that, as in the dichromatic network, $\tau_I \ll \tau_E$, the equations of motion are,

$$\begin{aligned} \tau_E \frac{\partial h_i}{\partial t} &= -h_i + I_s^i + w_{iH} F_I \left[\sum_j w_{Hj} F_E[h_j] \right] + \sum_j w_{ij} F_E[h_j], \quad i,j = \{R,G,B\}, \\ F_E[h] &= F_I[h] = \tanh h + 1 \end{aligned} \tag{2.15}$$

We first investigate whether trichromatic networks with dominant inhibitory feedback (Type I) are also advantageous for chromatic encoding, as found in dichromatic networks. The nullcline equations of Type I networks reduce to,

$$\begin{aligned} h_R &= \frac{w_{RH} h_G + (w_{GH}\, I_R - w_{RH}\, I_G)}{w_{GH}} \\ h_G &= \frac{w_{GH} h_B + (w_{BH}\, I_G - w_{GH}\, I_B)}{w_{BH}}. \end{aligned}$$

Similarly to the dichromatic case, type I networks exhibit a unique response to the same chromatic stimulus regardless of the coupling parameters strength. To investigate Type-II networks, we characterize the fixed points of Eq. (2.15) for a discrete set of coupling combinations and for all the chromatic stimuli considered previously. In contrast to dichromatic networks, such fixed points, if any, are embedded in a three-dimensional phase portrait. To locate the fixed points, we minimize a cost function similarly as in the dichromatic case, $\mathcal{L}$, that is zero if and only if $\dot{h}_i = 0$ for all three photoreceptor populations in Eq. 2.15, that is,

$$\mathcal{L} = \sum_i \left(-h_i + I_i + w_{iH} F_I \left[\sum_j w_{Hj} F_E[h_j] \right] + \sum_j w_{ij} F_E[h_j] \right)^2. \tag{2.16}$$

We use a gradient-descent algorithm to find the global minima of $\mathcal{L}$ for each parameter combination. It is easy to verify that, depending on the parameter combination, the network can exhibit one, two or three fixed points. Similarly to the dichromatic case, multiple fixed points are common in Type II networks with strong inter-cone connections. To see this in detail, note that Eq. 2.16 has six excitatory parameters corresponding to the couplings between red, green and blue cone populations. Considering a symmetric interaction between populations, only three free parameters, $w_{RG} = w_{GR}$, $w_{RB} = w_{BR}$

and $w_{GB} = w_{BG}$, remain. For each combination of inhibitory couplings previously studied, we calculate the number of fixed points for all combinations of these three remaining excitatory couplings, finding that the stronger the excitatory parameters, the larger the likelihood of bistability. We summarize this result in the intensity plot of Fig. 5e. For each fixed combination of the parameters w_{GB} and w_{RB}, we count the number of networks exhibiting multiple stable fixed points when varying the discrete values $w_{RG} = [0.1, 0.2 \dots 2]$. We find that networks with at least one strong excitatory inter-cone coupling tend to exhibit multiple responses to the same chromatic stimulus. This behaviour is similar for all Gaussian stimuli studied.

We conclude that adding a third type of photoreceptor to the outermost retinal networks does not change our general conclusion regarding population interactions. In fact, these results support the hypothesis that networks with predominantly inhibitory feedback provide an advantage for reliable and unambiguous chromatic encoding.

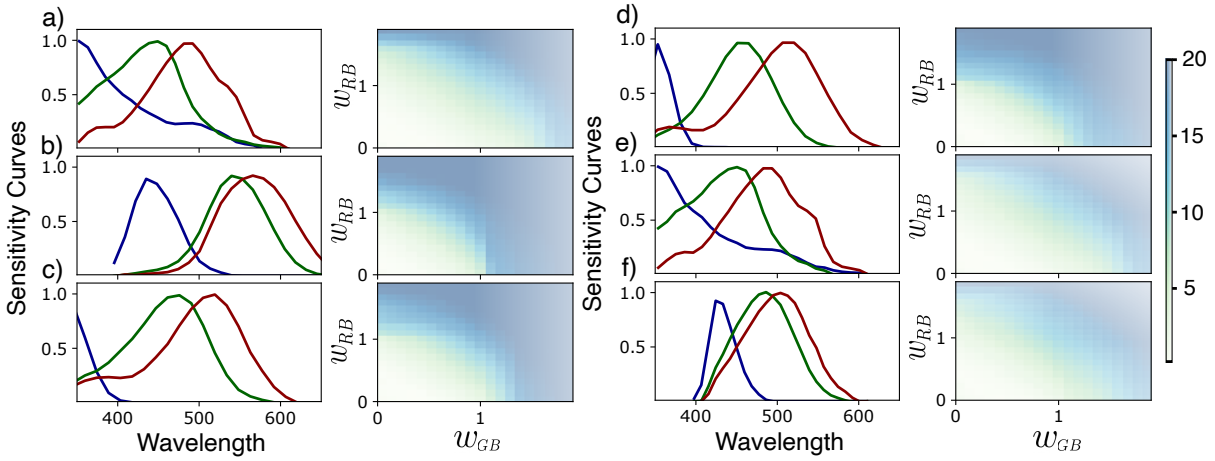

Figure 6 –

Sensitivity curves (left) and number of networks with multiple fixed points when varying $w_{RG} = [0, 0.1, \dots, 2.0)$, for the fixed parameters $W_{rh} = W_{hg} = 0.1$, $W_{hr} = W_{bh} = 0.2$ and $W_{gh} = W_{hb} = 0.3$, for: (a) Honeybee. (b) Human. (c) Spider. (d) Damselfish. (e) Giant clam. (f) Triggerfish. Data on sensitivity curves were extracted from ref [35].

As previously discussed, different species might exhibit combinations of opsins with other spectral sensitivity ranges. To test whether our results depend on the specific choice of opsins, we study six different trichromatic species, assuming the same network architecture. For type I networks, the analytic solution remains the same as in Eq. (2.16). Changes across species will be reflected in the stimuli current, I_S, which does not change the number of fixed points. For type II networks, the values of these external currents contribute to determine the bistability. Nevertheless, in all cases we observed networks exhibiting a bifurcation as a function of the excitatory inter-cone synaptic strength (see Fig. 6).

2.3 Optimal architectures of retinal circuits for chromatic encoding

In the previous section, we focused on the synaptic strengths between populations that yield stable responses. In this section, we use both zebrafish *in vivo* photoreceptor activities [40] and hyperspectral data –spectral measurements from ≈ 300nm to 700nm in steps of 1nm per pixel– from zebrafish environments [63] to investigate the architecture of outer retinal networks from the viewpoint of efficient coding and transmission of chromatic information.

The lower plot of fig. 2b shows the *in vivo* spectral responses of zebrafish photoreceptors. Remarkably, interactions with horizontal cells cause green and blue neurons to exhibit opponent responses to chromatic stimuli, which as discussed above, suggests early optimization of chromatic information transmission. Adopting this optimization hypothesis, we expect: (1) Reduced redundancy of network responses in spectral space [37]; (2) efficient coding of the environmentally available chromatic information. As a proxy for optimal chromatic encoding, we use a principal component analysis of the hyperspectral data of aquatic naturalistic images typical of zebrafish environments [63].

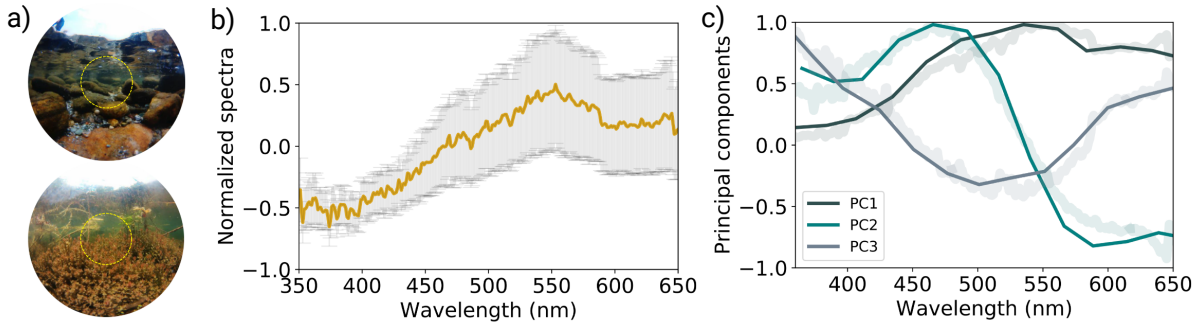

Figure 7 –

a) Examples of natural images from zebrafish environment. Each image contains 1000 sample points of the hyperspectra. b) Z-normalized hyperspectral data from a total of 31 images, with 1000 sample points each. c) First three principal components (PC1, PC2 and PC3) of the hyperspectral curve. Explained variance:
$\{PC1\} = 0.88$, $\{PC1, PC2\} = 0.93$, $\{PC1, PC2, PC3\} = 0.97$.

Fig. 7a shows two examples of typical natural images from zebrafish environments. Each image contains a total of 1000 sample points with the corresponding hyperspectral measure. The total data set contains 31 images, for a total of 31000 sample points. Fig. 7b shows the average spectrum curve. We calculated the principal components of the hyperspectal data by diagonalizing the covariance matrix between spectral intervals over all sampling points. We find that the first three principal components (PC1, PC2 and PC3) explain ≈ 0.97 of the hyperspectral data variance, in accordance with the analysis of ref. [40]. Figure 7c shows these three principal components (PC1-PC3), the first without

zero-crossings, the second with one zero-crossing, and the third with two. Comparing the *in vivo* responses (lower plot of Fig. 2b with these principal component curves, we observe that the red and green cone responses match the first two principal components qualitatively, supporting the hypothesis of efficient coding of chromatic stimuli in zebrafish outermost retinal layers.

We begin our analysis by studying the dichromatic network described by Eq. (2.12). We adjust the coupling parameters, w_{ij}, such that the spectral responses of the photoreceptor populations, h_R and h_G, match the first two principal components, which together explain more than 91% of the hyperspectral data variance. (We use the L-BFGS-B minimization algorithm to obtain the coupling parameters yielding the best match). Figure 8c shows the solutions of Eq. 2.12 for a network with only inhibitory feedback (in red and green), with both a no zero-crossing and a single zero-crossing curve, as expected for color-opponent signals [37, 64]. The inset shows the probability densities of the coupling parameter absolute value $|w_{ij}|$ of Eq. 2.12 over different basins of attraction. Similarly, we adjust the parameters of a network with both inhibitory feedback and excitatory inter-cone connections; in all cases, the optimal excitatory couplings are weak or negligible, leading to results similar to those found for the inhibitory network.

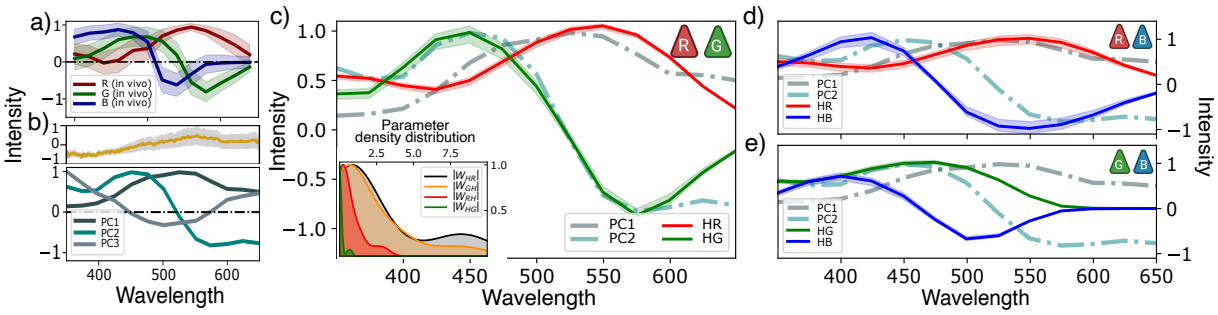

Figure 8 –

a) *in vivo* responses of red, green and blue zebrafish opsins [58]. b) Upper panel: Hyperspectral data from 31 images of aquatic natural images typical of the zebrafish larvae environment [58]. Lower panel: First three principal components obtained from the hyperspectral data. (c, d, e) Stationary solutions of Eq. (2.12) for the optimal coupling parameters of a dichromatic network with a) red and green b) red and blue c)green and blue photoreceptors. Inset a) shows the distributions of the optimal parameters (absolute value, $|w_{ij}|$) over 20 repetitions of the gradient descent algorithm starting from different initial values; inhibitory parameters w_{iH} are negative by definition; we used a KDE method to infer the curves. Red, green and blue curves correspond to the stationary solutions of the membrane potentials h_r, h_g and h_b, respectively. Dashed curves correspond to the principal component curves in b.

As expected from our previous dynamical analysis, we find a unique stationary

fixed point for all networks that exhibit the expected opponent chromatic responses. This means that the membrane-potential dynamics is the same (at least for $\tau_I \ll \tau_E$) for any initial condition, allowing reliable information encoding of chromatic stimuli. We also studied other opsin combinations, leading to the results shown in figures 8(d, e). We observe that having only blue and red photoreceptors provides a qualitatively good fit to the first principal component, but not to the second. Moreover, dichromatic networks with only blue and green photoreceptors cannot fit either of the first two principal components. Contrasting the results for these two-photoreceptor combinations, we conclude that dichromatic networks with long -and middle-range photoreceptors have the best performance when codifying chromatic information typical of the zebrafish environment.

Next, we analyze trichromatic networks, restricting the parameter space to reproduce the expected opponent responses, as we have done for dichromatic networks. Following the same procedure as before, we fit the membrane potential response of each photoreceptor population, h_R, h_G and h_B, to the first three principal components of the naturalistic images. As shown in Fig. 8a, the optimized network yields a poor fit to the third principal component, and the network responses do not match qualitatively the expected efficient responses. Instead, we find a single zero-crossing in the blue photoreceptor response curve. We note, though, that *in vivo* recordings of photoreceptor responses do not match this third principal component either, as shown in Fig.8a.

For trichromatic systems, type I and type II networks are unable to reproduce all three principal components, leading as to ask whether us to ask whether an expanded network, e.g., with a second parallel inhibitory feedback or with a fourth type of photoreceptor, is capable of realizing this task. We begin by including a second type of horizontal cell, H_2, integrating responses from only two of the three cone populations (see Fig. 9b), that is,

$$\begin{aligned}
\tau_E \frac{\partial h_i}{\partial t} &= -h_i + I_s^i + w_{iH_1} F_I\left[h_{H_1}\right] + w_{iH_2} F_I\left[h_{H_2}\right] + \sum_j w_{ij} F_E[h_j], \quad i,j = \{R,G,B\}, \\
h_{H_1} &= w_{H_1R}\, F_E[h_R] + w_{H_1G}\, F_E[h_G] + w_{H_1B}\, F_E[h_B] \\
h_{H_2} &= w_{H_2R}\, F_E[h_R] + w_{H_2G}\, F_E[h_G]. \qquad (2.17)
\end{aligned}$$

Repeating the previous analyses, we find that networks with two parallel inhibitory feedback mechanisms do not eliminate the errors in fitting of the third principal component. To study whether a fourth photoreceptor type would improve the network response, we include an ultraviolet photoreceptor in our formalism (see Fig. 9a), that is,

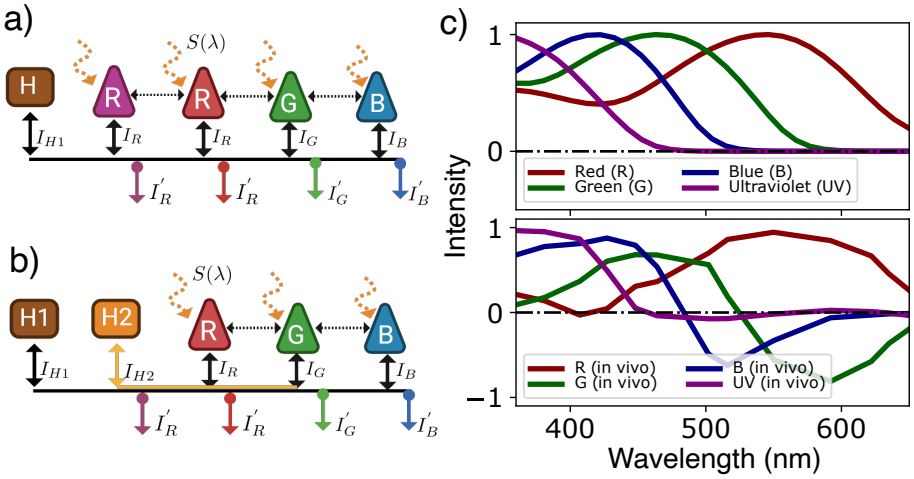

Figure 9 –

Tetrachromatic network: a) sketch of a network with an external stimulus, $S(\lambda)$. R: red cones, G: green cones, B:blue cones, U: uv cones, H: horizontal cells. Solid black arrows represent inhibitory synaptic conections between cones and horizontal cells. Dashed black arrows represent excitatory synaptic connections between cones. b) Sensitivity functions of independent red, green, blue and uv zebrafish opsins. d)*in vivo* responses of red, green, blue and uv zebrafish opsins. c) Stationary solutions for the optimal coupling parameters of the tetrachromatic network sketched in a). Red, green and blue curves correspond to the stationary solutions of the membrane potentials h_r, h_g and h_b, respectively. Dashed curves correspond to the principal component curves of Fig. 3b. d) Trichromatic network sketch of a fully connected network with an external stimulus, $S(\lambda)$, and two inhibitory feedback mechanisms. R: red cones, G: green cones, B:blue cones, H1: first population of horizontal cells, H2: second population of horizontal cells. Solid black arrows represent inhibitory synaptic conections between cones and horizontal cells. Dashed black arrows represent excitatory synaptic connections between cones.

$$\begin{aligned} \tau_E \frac{\partial h_i}{\partial t} &= -h_i + I_s^i + w_{iH} F_I\left[h_H\right] + \sum_j w_{ij} F_E[h_j], \quad i,j = \{R,G,B,U\}, \\ h_H &= w_{HR}\, F_E[h_R] + w_{HG}\, F_E[h_G] + w_{HB}\, F_E[h_B]. \end{aligned} \tag{2.18}$$

As shown in Fig. 9b, the responses of both the functional and isolated UV cones are similar, suggesting the absence of significant feedback from other populations to UV cones. We include a new population (U) that feeds the HC population (H), but does not receive excitatory/inhibitory feedback. The equation of motion remains the same as Eq. 2.15, but now with $i,j \in \{R,G,B,U\}$, and $W_{Ui} = W_{UH} = 0$. As shown in Fig. 9c, the gradient-descent algorithm to fit the coupling parameters yields a poor fit to the third principal component, as in the trichromatic network. We also included unilateral excitatory connections W_{Ui}, finding a maximum improvement of $< 2\%$ in the blue photoreceptor response. This suggests that ultraviolet cones do not play a key role in the chromatic opponent process of red, blue and green zebrafish photoreceptors. We conclude that adding

such a UV cone population has little effect on the functional responses of the other cones, in agreement with the experimental analysis of zebrafish by Yoshimatsu et al. [40].

From these results we conclude that in zebrafish retinas, trichromatic networks are incompatible with obtaining the first three principal components of environmental chromatic data at the first synaptic connection. This motivates us to add a second synaptic connection, or network layer, to verify whether such a network can be optimized to reproduce the expected results. Figure 10b shows a sketch of such a network, with a second type of horizontal cell, H2*, integrating the first layer responses, and providing feedback only to population B, which is the one unable to fit the third principal component. The blue cone population response, h_B^*, at the second layer is described by equation,

$$\frac{\partial h_B^*}{\partial t} = -h_B^* + h_B + w_{H_2^*R}\, h_R + w_{H_2^*G}\, h_G. \tag{2.19}$$

with $w_{H_2^*R}, w_{H_2^*G} < 0$. Figure 10c shows the responses of the optimal network after learning the parameters of equations 2.15 and 2.19. The improved agreement between response curves and PCs suggests that obtaining the three optimal chromatic channels requires a second synaptic contact (or network layer), including a second type of inhibitory neurons. Although we modeled this population as a second type of horizontal neuron, other neuronal populations in inner retinal layers, such as bipolar cells, might also realize this task. We also stress that the first three principal components explain more than 97% of the hyperspectral data variance, but that the third PC alone explains less than 6%. Consequently, despite the fact trichromatic networks with a unique inhibitory feedback mechanism do not reproduce the expected third principal component, the other two channels capture more than 91.3% of the hyperspectral data variance, which is still a good performance.

In summary, adopting ideas from information theory and chromatic opponency, we have shown that after the first synaptic connection, dichromatic networks can yield efficient responses when red cones are available. Otherwise, responses of photoreceptor populations are suboptimal. Generalizing to trichromatic networks, we find that, after the first synaptic connection, photoreceptors are unable to exhibit all three opponent responses. Only by including a second synaptic connection, or network layer, with a fourth neuronal population, can photoreceptor responses match the PCs.

2.4 Efficient opsin combinations in zebrafish

Our study identified networks that optimize the encoding of chromatic information available in the zebrafish environment, using as a template of photoreceptor responses the set of zebrafish sensitivity curves. Now we ask whether the zebrafish opsins are efficient at fitting the given chromatic information, or if instead, there are other opsin combinations

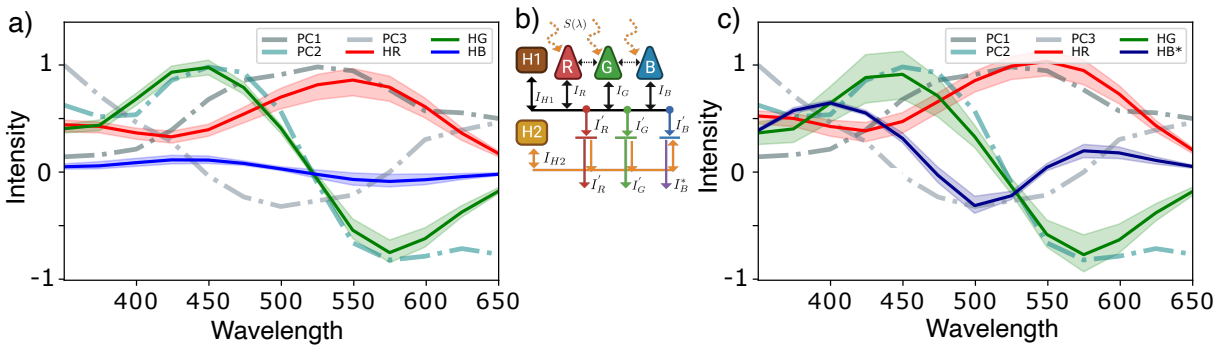

Figure 10 –

a) Stationary solutions of Eq. (2.15) for the optimal coupling parameters of the trichromatic network sketched in the inset. b) Sketch of a fully connected trichromatic network with two types of horizontal cells providing two successive inhibitory feedback mechanisms. c) Stationary solutions of Eq. (2.15, 2.19) for the optimal coupling parameters of the trichromatic network sketched in the inset. Red, green and blue curves correspond to the stationary solutions of the membrane potentials h_r, h_g and h_b, respectively. Dashed curves correspond to the principal component curves of Fig. 8b.

leading to more precise chromatic encoding. Altough optimization is not the only feature that determines visual system properties, it has been shown that some species adapt aspects, such as the sensitivity functions, to fit the environmental conditions [65, 66].

To answer this question, we contrast the performance of networks composed of different opsin combinations obtained by varying the wavelength of maximum sensitivity of zebrafish cones, while maintaining the shape of the distribution fixed (we use a discrete step interval of $\Delta_\lambda \approx 12$nm). For all opsin combinations, we calculate the optimal two-layer network, described by Eqs. (2.15) and (2.19), fitting the first three principal components of the zebrafish environment, as in the previous sections. To quantify the network performance, we define the cost function,

$$C_o^2 = (h_R - PC_1)^2 + (h_G - PC_2)^2 + (h_B^* - PC_3)^2 \,. \tag{2.20}$$

Using the same invertal Δ_λ for blue and green opsins, we calculate the cost function for all possible combinations within the interval $(350, 650)$nm, and search for an optimal set of opsins. Figure 11a shows the intensity plot of the cost function for all green and blue opsin combinations given a fixed red sensitivity curve. We find the global minimum is reached at $\Theta_R^*(\lambda) = \Theta_R(\lambda + 2\Delta_\lambda)$, $\Theta_G^*(\lambda) = \Theta_G(\lambda - \Delta_\lambda)$ and $\Theta_B^*(\lambda) = \Theta_B(\lambda - 3\Delta_\lambda)$, as illustrated in Fig. 11b,c.

We conclude that compared with zebrafish, the more efficient set of cones improves performance by approximately 13%, suggesting that retinal networks with zebrafish

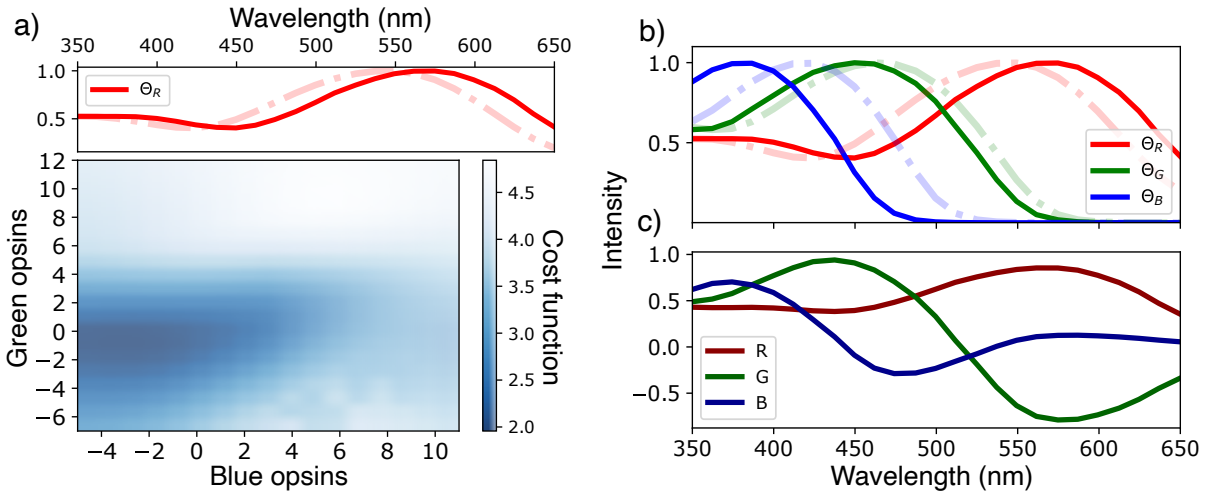

Figure 11 –

a) Intensity plot of the cost function (Eq. 2.20) for different opsin curves combinations, darker colors represent smaller values. Red opsin was first fixed to optimize all possible combinations; the optimal curve is shown in the upper plot. X and y labels in the intensity plot represent the number of shifts (in steps of 5nm) of blue and green curves respectively, with the sign indicating the shift direction. b) Comparison between experimentally observed sensitivity functions (dashed lines) and the optimal fitting curves (solid lines). Arrows indicate the shift direction yielding the most efficient responses. c) Network responses for this optimal set of opins.

cones are quite efficient to encode chromatic information. Extrapolating these results to other species with similar environmental chromatic conditions requires knowledge of the independent cone responses, or sensitivity curves to properly compare the networks. Studies involving species with different environmental conditions would naturally require analysis of the corresponding environmental hyperspectral data.

2.5 Discussion

Identifying the retinal circuits specialized in chromatic discrimination leads to an understanding of how organisms extract spectral information from their environment [67, 68]. *in vivo* experiments on outer retinal circuits, though, are only feasible in a restricted number of species, limiting a broader study. In this situation, theoretical models are useful to predict general features that can be tested in such model organisms and extrapolated to others beyond experimental access. Our study attempts to identify features of zebrafish-inspired outer retinal networks to understand more broadly the biological fundamentals of chromatic encoding and transmission in visual systems. As mentioned earlier, we focused our work on networks that mediate chromatic information via HCs. Other species, such as butterflies [69], with different network architectures will be studied in future work.

In the first part of our study, we found that in outer retinal networks with fast inhibitory feedback, inter-cone excitatory connections can lead to bistability, so that network responses to a given chromatic stimulus could be ambiguous. Ideally, one expects photoreceptors to codify as much relevant visual information as possible from the environment and transmit it reliably to downstream circuits. More complex tasks, such as perception or recognition, for which such bistability might be desirable, are supposed to take place later in the visual system. Our results suggest that, in retinas, inter-cone gap junctions, if present, are likely not involved in chromatic encoding. Some experimental studies on macaques and other vertebrate species [70–72] have shown evidence of both cone-cone and cone-rod gap junctions in the foveal region. The role of such connections, however is not fully understood. In ground squirrel retinas, for instance, such connections seem to play an important role in increasing the signal-to-noise ratio [71]; similar behaviour has been found in macaque retinas [73]. Noise reduction is highly relevant in low-luminance environments. At high luminance, where color discrimination is possible, the effects of gap junctions seem to be negligible, supporting the hypothesis that gap junctions are involved in achromatic tasks. Some other works [74] have shown evidence of excitatory feedback in individual synapses between HCs and cones, which might lead to other effective connections, not considered in our model.

In addition to studying different types of feedback, we investigated network architectures leading to efficient coding of the available chromatic information. Determining whether outer retinal circuits are optimized to codify available chromatic information allows one to gauge the relevance of color discrimination in a given species. Such findings might afford insights into other areas, such as ecology and animal behavior. As suggested by Yosimatsu et al. [40], zebrafish seem to codify of the chromatic information efficiently, qualitatively matching the first two PCs of the hyperspectral data. With our model, we find similar results. Specifically, we found that zebrafish-inspired outer retinal networks with a single HC layer can reproduce only the first two PCs of the zebrafish hypterspectral data, explaining 91% of the variance. Capturing more than 97% of the variance requires fitting all three PCs. We find that a more complex network, with two interneuron layers is necessary. We would expect retinas to optimize the trade-off between information transmission and metabolic cost, such that information transmission is guaranteed at the lowest energy expenditure. Nevertheless, even with an optimal architecture, more complex networks are necessary to transmit specific information, as suggested by our results. Whether it is advantageous for an organism to invest energy in adding an additional retinal layer to improve chromatic discrimination depends on species-specific interactions with its habitat. These results are general for networks with inhibitory interneuronal layers and similar environmental conditions.

Our study of network architecture was performed using zebrafish opsin curves as a reference. Subsequently, we investigated whether other opsin combinations might lead to further optimization of the network, that is, whether it would be possible to improve zebrafish performance at codifying environmental chromatic information. We find that optimizing opsin combinations leads to an improvement of approximately 13% in zebrafish performance. As mentioned earlier, in some species, such as bees, photoreceptor sensitivity functions are highly tuned to the environmental chromatic spectrum, leading to an efficient coding of color information [65, 66]. Our results on zebrafish show that they are quite efficient at chromatic encoding, but that there are other photoreceptor combinations yielding modest improvements. Other relevant features, such as movement or edge detection, might favor retinal circuits that optimize other general aspects of visual stimuli, by penalizing slightly chromatic information. As before, these results hold for zebrafish-inspired retinal circuits under similar environmental conditions.

3 Trade-off between coding efficiency and color space in outer retinal circuits with oil droplets

Retinal circuits of color discrimination are quite diverse [67]. Across species, a variety of biological mechanisms and strategies allow animals to discriminate relevant spectral information in their environments. One such mechanism is the colored oil droplet organelle thought to improve color vision by expanding the space of distinguisable colors [75, 76]. These organelles are typical for instance in birds, whose retinal basis provides them with over-average chromatic-discrimination abilities [77–79]. Other species, that also rely on chromatic information to thrive, such as the teleost zebrafish [80], do not develop oil droplets. Instead, as described in the previous chapter, their retinal circuits are apparently adapted to efficiently encode spectral information of their environment [40, 81]. Studies of these strategies show how each of them separately leads species to optimize their color vision capabilities. Nevertheless, the interplay between them is unclear since functional retinal circuits for color vision have yet to be analyzed in species with droplets.

Colored oil droplets are organelles filled with carotenoids, located within the inner segment of cones, right before the corresponding visual pigment [35, 75]. Different cone types are related to a specific oil droplet [75, 77]. For instance, in birds we find the cone types SMW1, MMW and LMW with the corresponding colored droplets C-type, Y-type and R-type [77]; UV cones have transparent droplets without known spectral effects [79]. Such organelles act as optical filters with a full-absorbance limit characterized by the cut-wavelength λ_{cut}; all wavelengths below λ_{cut} are absorbed by the oil droplet before reaching the cone visual pigment. As a result, the sensitivity function of the different cone types shrinks and diminishes, reducing their spectral range of response as well as their photon-capture probability. On one hand, the response overlap among cone types decreases, reducing cone-signal redundancy and expanding the potential color space [82, 83]. On the other hand, the attandant reduction in of the photon-capture probability worsens the signal-to-noise ratio, limiting chromatic discrimination in dim-light environments [84, 85]. Some works [86] have also suggested that clear and transparent droplets might serve as micro-lenses to enhance light focusing. However, regarding color vision in photopic conditions, droplets are considered as optical filters whose transmittance can be theoretically modeled using electrodynamics [87, 88].

Experimental and theoretical work (see previous chapter) in species without droplets, such as zebrafish [40, 81], has documented the functional circuits of color in the outermost

retinal layers, revealing the network architecture that allows an efficient coding of environmental spectral information. We use data on oil droplets from different species [87–89] to implement a model of cone spectral responses that we incorporate in previously investigated zebrafish-like retinal networks [81] (see previous chapter). We investigate the combination of both strategies by quantifying the chromatic coding performance, and the corresponding gain in color space of these networks. To this end, we structure this chapter in three sections. In Section 1, we study the optical properties and spectral responses of droplets and cones from 18 different bird species, contrasting them with the few available data from aquatic species. In section 2, we introduce the Maxwell triangle for red, green and blue zebrafish cones to quantify changes in color space when oil droplets are included. Finally, in Section 3 we introduce the population model for the network with droplets, defining a cost function based on the hyperspectral images of the zebrafish environment. We model a variety of droplets with different optical properties to contrast color encoding performance in networks with and without such organelles.

3.1 Effects of oil droplets on cone sensitivity curves

Light transmission models [87, 88] have been used to estimate the responses of visual pigments with oil droplets, matching experimental observations. Here, we implement the transmittance model of Ref. [87] to estimate the sensitivity curves of zebrafish cones if oil droplets were present.

Oil droplets serve as an optical filter characterized by the transmittance function $T(\lambda)$. Depending on the droplet properties, some light wavelengths are either partially or completely blocked, leading to a narrower spectral response of cones. Oil droplet transmittance is fully characterized by two parameters: the cut wavelength λ_{cut} and the mid wavelength λ_{mid}, the latter being the wavelength at which the oil droplet has half of its maximum absorbance [87]. Such characterization leads to the function,

$$T(\lambda) = \exp\left[-2.93\, \exp\left[-2.89\, m(\lambda_{\text{mid}})\, (\lambda - \lambda_{\text{cut}})\right]\right], \tag{3.1}$$

with $m(\lambda_{\text{mid}}) = 0.5/(\lambda_{\text{mid}} - \lambda_{\text{cut}})$ [89]. Details on the derivation of Eq. (3.1) are provided in Appendix E.

We are interested in modeling colored droplets feasible for aquatic species, as is the case of zebrafish. Nevertheless, only a few of them develop these organelles, such as lungfish [89]. In contrast, oil droplets are common in birds, allowing a broad characterization of their optical properties. Here, we use bird data to model the wavelength parameter λ_{mid} as a function of λ_{cut}, and, as a proof of concept, we use the data available from lungfish to test whether the model is suitable to describe droplets from aquatic species. Fig. 12a

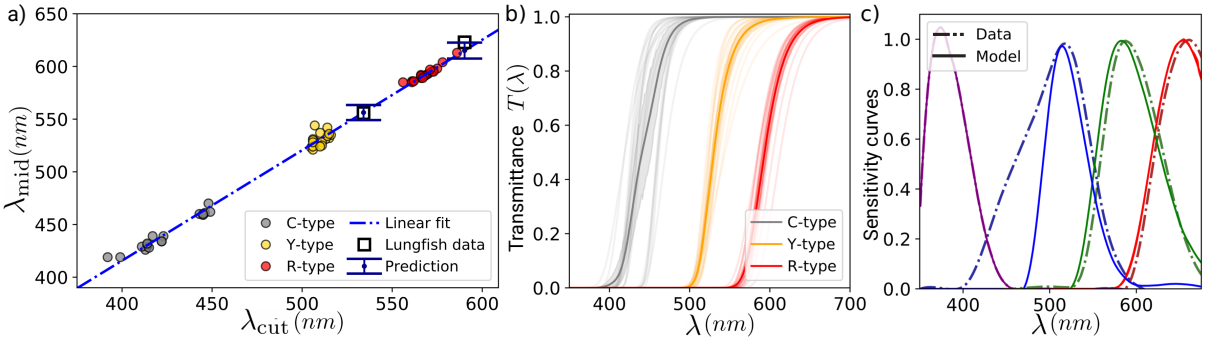

Figure 12 –

a) Relation between the parameters λ_{mid} and λ_{cut}. Circle markers correspond to the experimental data from 18 different bird species [89]. Each color corresponds to a different type of oil droplet; C-type (gray), Y-type (Yellow) and R-type (red). The blue line corresponds to the linear fit with line equation $\lambda_{\text{mid}} = 1.05 \pm 0.0097\,\lambda_{\text{cut}} - 3.57 \pm 4.92$. Square markers correspond to the experimental data of lungfish while the error bars correspond to predictions of the linear fit. b) Transmittance function $T(\lambda)$ as a function of the wavelength. Dark lines correspond to the mean transmittance curve for each type of oil droplet represented with light curves. c) Sensitivity curves with oil droplets. Dashed lines correspond to inferred curves from experiments with lungfish [89]. Solid lines correspond to the implementation of the transmittance model of Eq. (3.1).

shows experimental data on both λ_{cut} and λ_{mid} for different bird species [87], with gray, yellow and red circle markers representing C-, Y- and R-type droplets respectively. We fit a linear model to describe all bird droplets (dashed line), that we subsequently use to infer the wavelength parameters corresponding to C- and R-type droplets of lungfish [89]. More specifically, we use the experimental value of λ_{cut} of lungfish to estimate the relative error of the parameter λ_{mid}, corresponding to a linear model. As shown in Fig. 12a, experimental measurements of lungfish (square markers) are within the error bars of the model predictions, leading to relative errors of 0.1% and 1.26% for C and R droplets respectively. This result suggests that filter properties of droplet organelles might be unaffected by the specific environmental conditions of the species, in this case, aquatic or terrestrial habitats. Altogether, we expect the transmittance model of Eq. (3.1) and the linear relation between the parameters λ_{mid} and λ_{cut} to be suitable across species.

Figure 12b shows the mean transmittance curves of Eq. (3.1) corresponding to the C-, Y- and R-type oil droplets used in Fig. 12a. As expected, we observe that the larger the λ_{cut} the larger the zero transmittance region in the visible spectrum. To estimate the cone spectral function, we multiply the droplet transmittance function with the sensitivity function of the visual pigment. For instance, Fig. 12c shows the experimentally estimated sensitivity functions of lungfish cones (dashed lines) and the model predictions (solid lines). We observe that the model is in good agreement with both MWS and LWS sensitivity functions, but it fails at predicting the correct shape of the SWS cone in the

short-wavelength range –UV sensitivity remains unchanged due to their lack of colored droplets. Still, the model predicts correctly the wavelength peaks of maximum sensitivity as well as the overlap regions between the three red, green and blue cone responses.

So far, we have shown that by using the transmittance model of Eq. (3.1), we can estimate the sensitivity functions of different cone types with a variety of oil droplet combinations. For instance, Fig. 13a shows the zebrafish sensitivity curves without (above) and with (below) oil droplets, illustrating the narrowing and decrease of the response profile; we relate red, green and blue cones with R-, Y- and C-type droplets as found in birds. Considering that oil droplets cause both a red-shift and narrowing in the sensitivity functions, we use the wavelength of maximal sensitivity, $\lambda_{\max}^{\mathcal{O}_i}$, and the full width at half maximum (FWHM), $\sigma^{\mathcal{O}_i}$, to characterize the response changes as a function of the droplet cut-wavelength parameter. Fig. 13b shows the normalized red-shift effect for red, green and blue cones, which follows a linear relation with slope ≈ 1 regardless of the droplet type. By contrast, for the FWHM, we find that red and green cones suffer the strongest narrowing compared to the original sensitivity function, while blue cones only lose sensitivity in the ultraviolet spectral range. This means that the longer the peak-sensitivity wavelength, the more significant the droplet effect on the cone spectral response.

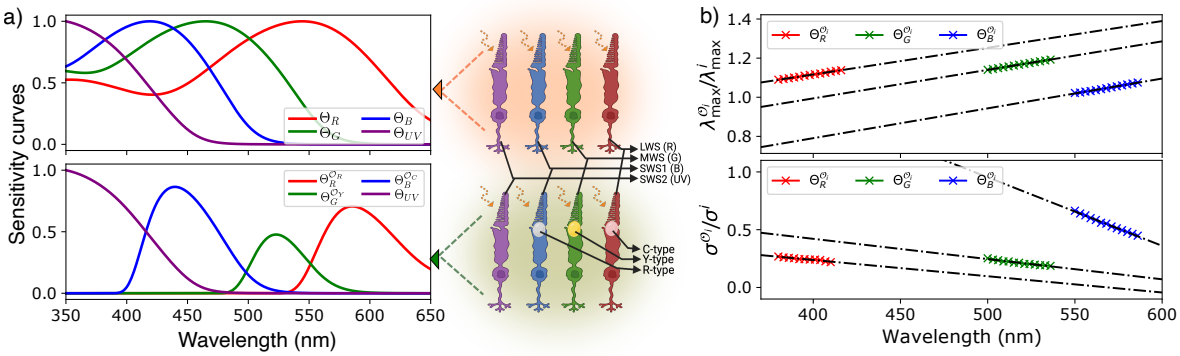

Figure 13 –

a) Predicted sensitivity curves in zebrafish a) without oil droplets. and b) with oil droplets. We used the cut wavelength parameters $\lambda_{\text{cut}} = \{400, 490, 540\}$ for the C-, Y- and R-type oil droplets respectively. b) Normalized maximal sensitivity wavelength, $\lambda_{\max}^{\mathcal{O}_i}$, and c) Normalized FWHM, $\sigma^{\mathcal{O}_i}$, as a function of the oil droplet parameter λ_{cut} for red, green and blue zebrafish cones.

3.2 Color space and color encoding

To discriminate colors, spectral features of visual stimuli must be distinguised from brightness. Experimental works on functional retinal circuits show that different species

implement color-opponent responses to achieve this task [67]. For instance, experiments in zebrafish retina [40] suggests that color circuits involving red, green and blue cones in the outermost layers yield functional chromatic channels that exhibit opponent responses. Furthermore, when compared to the statistics of hyperspectral images of zebrafish environment, these responses are found to be adapted to optimize color information encoding [81] (see previous chapter). Broadly, color-opponent channels combine the spectral-wide cone responses, preserving the relevant spectral information and decreasing response redundancies [37]. Interestingly, the overlap reduction between cone responses caused by droplets also leads to a direct redundancy decrease. To investigate the combination of droplets and adapted retinal circuits, we add the droplet model to the zebrafish outer retinal circuit, focusing in the color encoding performance of the network as well as in the gain in color-space.

For some species, such as birds, the expansion of color space allows the possible discrimination of more colors playing a key role in their ecological context [35, 90]. Nevertheless, for other species, such an expansion might not be critical given the environmental conditions. To examine the case of zebrafish, we use the Maxwell color triangle for red, green and blue cones. Fig. 14a shows the case of cones without droplets, in which the colored curve inside the triangle corresponds to the locus, limiting the zebrafish distinguisable color-space. Star markers indicate some typical colors extracted from the hyperspectral natural images in Ref [57]. We observe that most colors are distinguishable by zebrafish, but still, many are quite near to the locus boundary. As shown in Fig. 14b, the insertion of oil droplets leads to a larger locus with a less restrictive color-space; including the three oil droplets increases the zebrafish color-space volume approximately four times.

Figure 14c shows the color-space volume variation with the parameters λ^C_{cut} and λ^R_{cut}, corresponding to the C- and R-type droplets respectively, and the parameter λ^Y_{cut} fixed. We observe that once all three oil droplets are included in the system, the color space is only slightly affected by variations of the droplet parameter λ^i_{cut}. Variations including the parameter λ^Y_{cut} lead to the same results. Some retinas, tough, are found to have mixed combinations of cones with and without oil droplets, leading to the narrowing of only some sensitivity curves. We study these possible combinations for the zebrafish sensitivity curves. Fig. 14d shows the area estimation for different combinations represented by the colored code above. We observe that the red cone droplet (R-type) plays a key role in increasing the color space by decreasing considerably the overlap between cone responses. For instance, we find that the combination of only Y- and R-type droplets lead to an expansion similar to that generated by all three types of droplets. In contrast, combinations without R-type droplets lead to smaller area expansions and in some cases to an area decrease, as is the case of the combination 3 with only Y-type droplets. As shown in Fig. 14e, this combination with only Y droplets red-shifts the green-cone sensitivity, increasing

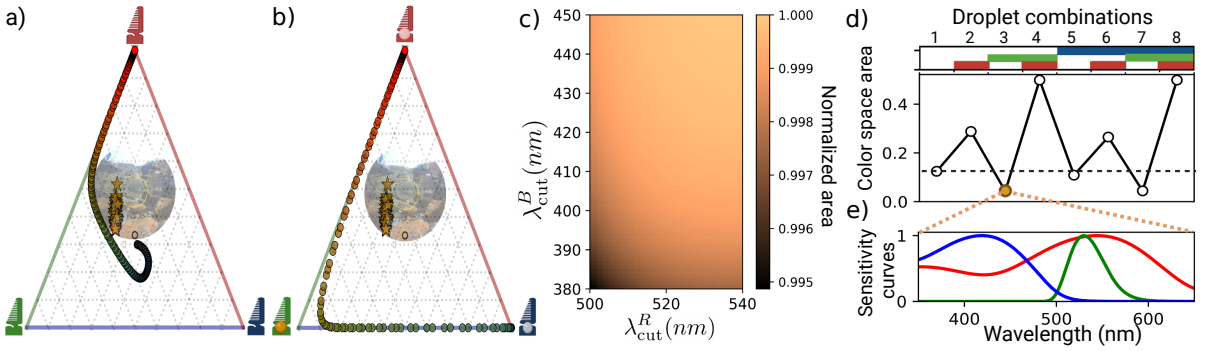

Figure 14 –

(a, b) Maxwell triangle of zebrafish trichromatic system a) wihout oil driplets, and b) with C-, Y- and R-type of oil droplets with the cut wavelengths $\lambda_{\text{cut}} = \{460, 510, 530\}$nm respectively. c) intensity plot of the color-space normalized area for the fixed value $\lambda_{\text{cut}G} = 530$nm. d) Color-space volume for different combinations of cones with and without droplets. The x-axis shows the combination of droplets in each case, such that colored (empty) spaces represent the presence (absence) of an oil droplet in the corresponding cone; blue, green and red cones are represented by the first, second and third row respectively. When present, the droplets have the same parameters used on the plot a). e) Sensitivity functions corresponding to the combination 3 (orange marker) in plot d.

the response redundancy between green and red cones, which decreases considerably the color-space volume.

3.3 Functional retinal circuits for color discrimination

Including oil droplets can also modify the outputs of the underlying chromatic channels in the visual system. In the previous chapter we discussed the network structures capable of efficient color encoding at early retinal stages (see Fig. 15a), which allows a systematic study of retinal networks with different cone types. We use the trichromatic model describing the dynamic responses of the network shown in Fig. 15a. For convenience we recall the corresponding equations:

$$\begin{aligned} \frac{\partial h_i}{\partial t} &= -h_i + I_i + \sum_{j \neq i} (w_{ij}\, F_E[h_j]) - w_{iH} \sum_j (w_{Hj}\, F_E[h_j])\,, \\ \frac{\partial h_B}{\partial t} &= -h_B + \sum_i w_{H_2 i} h_i \end{aligned} \tag{3.2}$$

with $F_E[h] = \tanh(h)+1$ the gain function and i the index representing the cone population; red (R), green (G) or blue (B'); the second equation corresponds to the output of the blue cone via the second inhibitory layer (see Fig. 15a and previous chapter). The parameters w_{ij} represent the coupling strength between populations, with $w_{ij} > 0$ ($w_{ij} < 0$) for

excitatory (inhibitory) pre-synaptic populations. The term I_i represents the independent response of cones to external visual stimuli. In our previous work, such a response is estimated with the cone sensitivity function, $\Theta_i(\lambda)$, and the stimulus spectral function, $S(\lambda)$. Here, we include the effect of oil droplets by adding the transmittance function of Eq. (3.1), that is,

$$I_i = \int \Theta_i(\lambda)\, T(\lambda) S(\lambda)\, d\lambda. \tag{3.3}$$

We quantify the encoding performance of networks with different droplet combinations by comparing the statinary responses in Eq. (3.2) with the principal components of the hyperspectral data of zebrafish environments (see dashed lines in Fig. 15b) [40, 81]. First, for a given droplet combination, we optimize the coupling parameters w_{ij} that minimize the cost function

$$C_o^2 = (h_R - PC_1)^2 + (h_G - PC_2)^2 + (h_B - PC_3)^2\,, \tag{3.4}$$

that is, the squared distance between the principal components and the network responses; we use the same same minimization methods as in the previous chapter. For instance, Fig. 15b shows the responses (solid lines) of a network with only Y-type droplets after optimizing the coupling parameters. To quantify the effects of oil droplets in chromatic encoding, we use the cost function of Eq. (3.4) as a performance measure. We use the linear extrapolation in Fig. 12 to systematically calculate the transmittance function of different droplet types and combinations. More specifically, we use the data in Fig. 12a to delimit the ranges $\lambda^R_{\text{cut}} = [550, 600]$nm, $\lambda^G_{\text{cut}} = [500, 550]$nm and $\lambda^B_{\text{cut}} = [400, 450]$nm for the droplet parameters corresponding to red, green and blue cones respectively.

We begin by studying combinations of droplets with the minimal value of the parameter λ_{cut}, that is $\lambda^R_{\text{cut}} = 550, \lambda^Y_{\text{cut}} = 500$ and $\lambda^C_{\text{cut}} = 400$. Fig. 15c shows the corresponding cost function for the different droplet combinations represented by the colored plot below. We observe that, on average, including oil droplets decreases the performance of the network to fit the expected principal components, as indicated by the growing linear trend (dashed blue line). In more detail, we find that the presence of R-type droplets contributes considerably to such a decrease, reducing the network efficiency to encode color information from natural images.

Finally, to investigate the effects of the droplet properties, we study the role of the parameter λ^i_{cut}. Given a droplet combination (from the possible 2^3), we vary systematically the cut-wavelength parameters in steps of $\Delta\lambda_{\text{cut}} = 5$nm, within the intervals previously defined for each droplet. As before, for each droplet parameter combination, we infer the network coupling parameters w_{ij} that minimize the cost function of Eq. (3.4). Here, we show the results corresponding to the combination of only R and Y oil droplets,

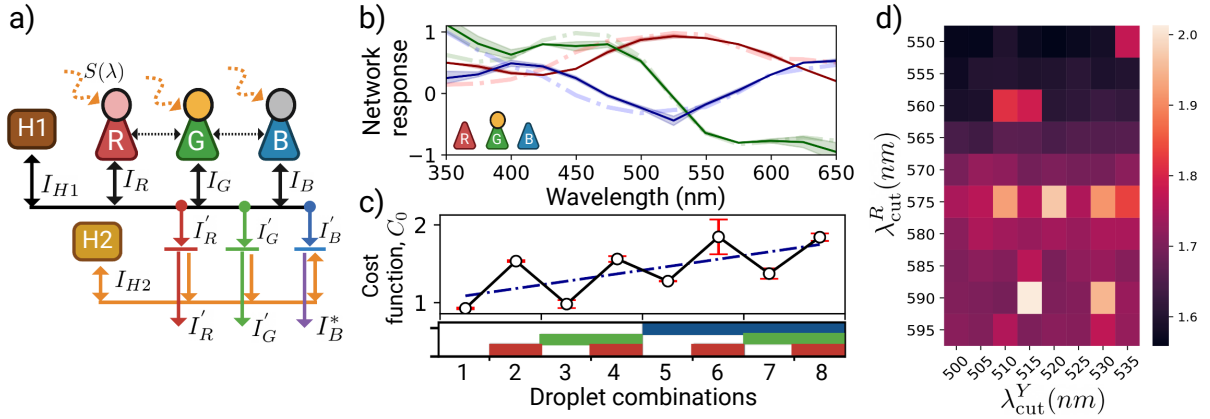

Figure 15 –

a) Trichromatic network with excitatory inter-cone couplings and inhibitory feedback from two layers of horizontal cells. Circles above cone symbols represent oil droplets. b) Stationary responses of a network with only Y-type droplets, as represented in the inset. Dashed lines correspond to the principal components from the hyperspectral images and solid lines correspond to the mean network responses for different coupling parameters minimizing the cost function in Eq. (3.4). c) Cost function of Eq.(3.4) for different combinations droplets represented by the colored plot below; colored (empty) spaces indicate the presence (absence) of droplets; red, green and blue colors represent R, Y and C droplets respectively. d) Intensity plot of the cost function for networks with R- and Y-type droplets. The cut-wavelength parameters of red (λ^R_{cut}) and green (λ^G_{cut}) cones were varied in steps of 5nm within the corresponding spectral intervals.

which as shown in the previous section, lead to the largest color-space expansion (see Fig. 14d). Nevertheless, we emphasize that, as expected from Fig. 15c, the results with other droplet combinations remain similar. Fig. 15d shows the intensity plot of the cost function, suggesting that the smaller the cut-wavelength the smaller the cost function and consequently, the better the chromatic encoding performance.

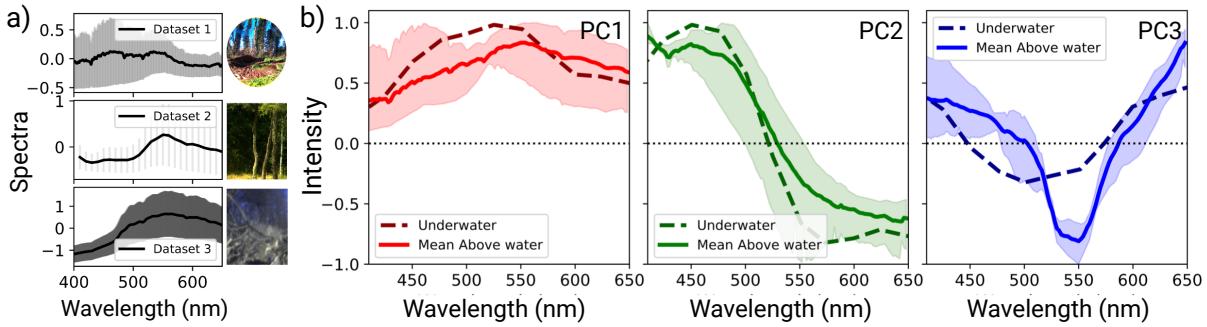

Figure 16 –

a) Spectra of three different datasets of images above water. Dataset1, Dataset2 and Dataset3 correspond to the hyperspectral images in Refs [40, 91, 92] respectively. b) Principal components of hyperspectral data. Solid lines correspond to the mean over all three datasets above water. Dashed lines correspond to hyperspectral data underwater.

Altogether, our results suggest that including oil droplets in a zebrafish-like network that exhibits efficient coding affects its performance, keeping it from matching the expected efficient responses. Depending on the specific droplet combination and their optical properties, the encoding performance differs. For instance, we found that including R-type droplets leads to a large decrease of encoding performance compared to other droplet types. One possible reason is that zebrafish-like retinal networks are adapted to use largely overlapping cone responses, facilitating the expected efficient responses. Since R-type droplets produce the largest effect, by blocking the sensitivity of a large spectral range in red cones, such network based in feedback mechanisms might exhibit encoding limitations. By contrast, regarding color space, R-type droplets lead to the largest expansion, increasing significantly the possible distinguishable colors of zebrafish. Similarly, Y-type droplets lead to a smaller but still significant effect compared to the R-type droplet. In fact, the combination of both R and Y droplets lead to the largest changes in both color space gain and coding performance decrease. This opposing effect of droplets in the zebrafish-like retinal network suggests the existence of a trade-off between color-space capacity and coding efficiency which is caused by the overlap decrease of cone responses in the spectral range.

Whether such a trade-off plays a role in animals with droplets, such as birds, depends on (i) the specific retinal circuits of color discrimination and (ii) the ecological conditions. Fig. 16a shows the spectra of three different data sets of natural images above water [40, 91, 92]. We perform a principal component analysis of these data to contrast the differences with the underwater scenario. Fig. 16b shows the first three principal components of aquatic environments (dashed lines) and the mean principal components over the three data-sets above water (solid lines). We observe differences in the wavelength crossing-points as well as in the intensity peaks. Nevertheless, the functional form of these curves remain quite similar, suggesting that zebrafish-like network architectures might be feasible to encode color information under the trade-off introduced by the presence of oil droplets. Although feasible, the extrapolation of zebrafish retinal circuitry to birds is not straightforward. In contrast to zebrafish, they also have double cones, which might provide broader spectral information decreasing the trade-off effects. Similarly, depending on the functional circuitry, UV cones could be involved at this early stage of color coding.

3.4 Discussion

As mentioned in the Introduction, several species rely in color vision to meet many of their ecological needs [35]. This has lead to the specialization and adaptation of retinas to optimize the encoding and processing of chromatic information at early visual stages. Among these possible specializations, we focused on two: the encoding of chromatic information via functionally efficient retinal circuits and the expansion of color

space via oil droplets. Experimental studies of functional retinal circuits are available only in few species, such as zebrafish, lacking oil droplets, while experiments in other species with oil droplets, such as birds, are still challenging. Nevertheless, over the last decades, experimental and theoretical works [77, 82, 85, 87, 88, 93–97] have elucidated some key optical and functional aspects of colored oil droplets in visual systems, allowing a computational study of the interplay between oil droplets and efficient retinal circuits. It is important to mention that color vision enhancement is one of the strongest advantages conferred by such organelles in photopic conditions, refining the spectral sensitivity of cones and consequently expanding the visual color space [75]; other optical properties regarding high-acuity vision have also been investigated and related mostly to colorless droplets [86].

We have shown that including droplets in zebrafish-like networks for chromatic coding poses a trade-off between efficient coding and color-space, being R and Y droplets the organelles inducing the largest effects. Our findings suggests that zebrafish retinal circuits for color efficient coding require largely overlapping cones responses. Consequently, their refining via oil droplets leads to a marked change of the incoming signal, keeping the system from exhibiting the expected responses. Regarding color space, we found that colors from the zebrafish natural environment are mostly embedded within the loci of cones without droplets, meaning that a potential expansion of color space might not lead to a direct benefit for zebrafish. Nevertheless, it is important to mention that the available hyperspectral images are limited and lack potentially relevant chromatic information. For instance, spectral information of zebrafish skin patterns is not included and might lead to color points outside of the dropless loci. Such a case might motivate the need for a color-space expansion as in the case of some bird species [90].

The reason why some species have oil droplets and others do not is not clear. A plausible explanation arises from phylogeny studies [75] suggesting that oil droplets are evolutionarily inherited. That is, taxa without colored oil droplets, such as *Actinopterygii*, are a consequence of a common ancestor lacking these organelles. As a consequence, some of such species, such as zebrafish, have adopted other biological strategies to improve color vision that seem to be obstructed by the presence of oil droplets. Our findings suggest that, If possible, the development of these organelles or other similar chromatic filters in zebrafish-like retinas might require circuit adaptations that optimize the trade-off between efficient color coding and color space capacity.

Part III

4 Compression as a path to simpler models of neuronal activity

As discussed in the Introduction, ganglion neurons perform the final step of retinal information processing by generating the spiking code which contains the extracted bits of information from the complex visual stimuli. Understanding this final stage can provide us with important insights into the general principles behind neuronal network encoding. In contrast to other neurons in the retina, ganglion cells (GCs) exhibit peaks of activity that are known as spikes. Fig. 17a shows a typical activity profile in which after certain threshold, θ, the neuron *fires* an action potential or spike that is transmitted to postsynaptic neurons. Each type of neuron has a characteristic threshold, θ_i, and spike duration, $\tau_s^{(i)}$. For instance, GCs in the salamander retina have a typical spike duration (including the refractory period) of $\tau_s \approx 20$ms [98].

Considering the characteristic spike duration, neuronal activity can be discretized into bins of size $\Delta t \approx \tau_s$, leading to two simplified states; $\sigma(t) = 1$ for spike and $\sigma(t) = 0$ otherwise. This simplification becomes relevant when a large number of neurons, N, are recorded simultaneously such that we can define the state of the population in a certain time bin j as $\sigma_{t_j} = \{\sigma_1(t_j), \sigma_2(t_j), \ldots, \sigma_N(t_j)\}$. Figures 17(b,c) show a raster plot representation in which the activity of N neurons is discretized into bins of $M = T/\tau_s$, with T the total time.

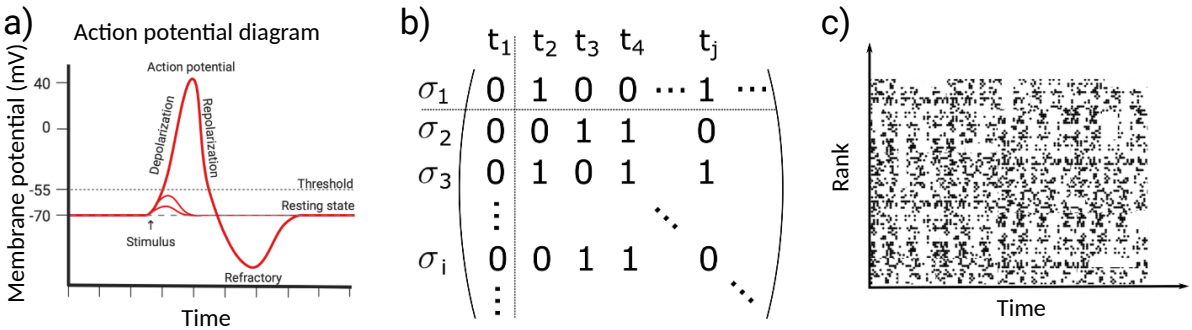

Figure 17 –

a) Action potential profile of a neuron with a threshold of $\theta = -55$mV. The action potential comprises three phases: depolarization, repolarization and refractory period. The typical temporal scale an action potential is $\approx$ 2ms. b) Neuronal activity discretization. The y-axis represents the neuron index and the x-axis the temporal samples; spikes (no spike) are represented by the state 1 (0). c) Raster plot representation of the discretized activity. Empty (filled) spaces represent absence (existence) of activity in some specific time window.

The resemblance of this two-state description to typical condensed matter systems has motivated the study of large neuronal populations from a theoretical perspective. For instance, spin models have been actively used to explain the global activity patterns and correlations observed in large neuronal population recordings [98–102]. Nevertheless, the number of parameters in such models grows with the number of neurons as 2^N, becoming impractical for real neuronal populations with at least thousands of neurons. Recent works have implemented inference methods, such as the maximum entropy, to model neuronal activity in the hippocampus with pairwise models [103]. However, these models fail at explaining the typical correlations observed experimentally in other neuronal populations, such as the ganglion cells of the salamander retina [98].

One possible path to tackle this problem is to develop more sophisticated inference methods that allow the implementation of models with higher order interactions that capture most of the correlations. Another path is to explore new models of neuronal activity that do not rely on assumptions of neuronal interactions. In this chapter, we focus in the latter by exploring ideas of compressibility of interactions [104] from an information theory perspective, as a path to build simpler models of neuronal activity.

5 Models of neuronal activity and compressibility of interactions in the retina: Salamander ganglion cells

We begin by introducing in more detail the ideas behind the compressibility of interactions which are inspired in spin systems and have been broadly discussed in Ref. [104]. Let us consider the N-spin lattice shown in Fig. 18a, in which each spin has two possible states s_i = up or down, leading to 2^N network states. One possible way to describe this system is via the joint probability $p(\{s_j\})$. By contrast, if we want to describe the spin interactions, we can focus in the conditional probability $p(s_i|\{s_j\})$, with s_i a single spin and $\{s_j\}$ the other $N-1$ spins.

In principle, we can solve this conditional problem by defining an effective field, $h_{\text{eff}}(\{s_j\})$, that describes the influence that each possible state in $\{s_j\}$ has on s_i. Nevertheless, following ideas of renormalization group (RG) [105], this solution with 2^{N-1} parameters might be inefficient if the system allows a scaling in which the number of degrees of freedom are reduced by some amount b while still preserving the physical properties of the system, such as the spin correlations. In our example, this would imply that the group $\{s_j\}$ can be scaled to a group $\{s_j^*\}$ with $(N-1)/b$ degrees of freedom. If possible, this scaling allows a vast simplification of the effective field describing the conditional probability $p(s_i|\{s_j\})$. In the context of neuronal networks, we ask whether we can extrapolate these ideas such that we can describe the state of a single neuron σ_0 conditional to the state of the rest of the network, $\boldsymbol{\sigma}$, with a significant reduction of the expected 2^{N-1} states (see Fig. 18b). That is, we want to find a compressed representation, $\tilde{\boldsymbol{\sigma}}$, of the variable $\boldsymbol{\sigma}$ for all possible choices of σ_0 such that we can determine the conditional distribution $p(\sigma_0|\tilde{\boldsymbol{\sigma}})$ that describes all neuronal interactions.

5.1 Compression model

In a nutshell, the RG formalism leads to a reduction in the number of degrees of freedom while preserving essential physical features, such as the correlations. In the context of neuronal populations with thousands of neurons, such a renormalization might lead to a vast simplification, making possible simpler models of neuronal activity. Nevertheless, in contrast to typical spin systems, neuronal networks do not have a well defined spatial interaction range that we can use to formalize our problem. As an alternative path, we propose a compression approach in which we can similarly decrease the total number of

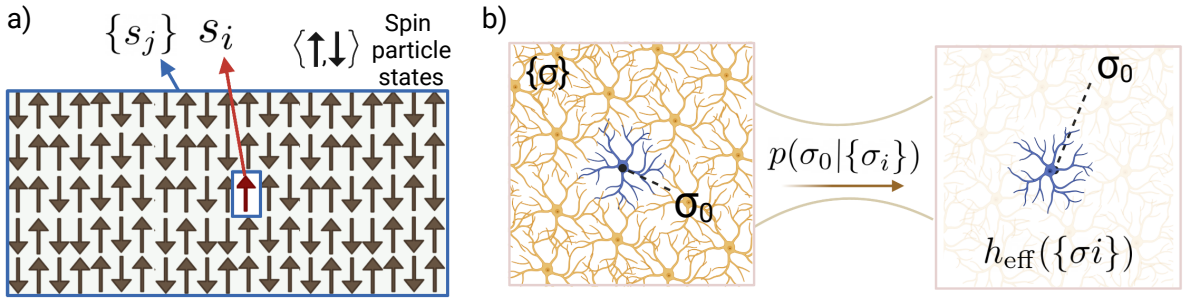

Figure 18 –

a) Sketch of a lattice of interacting spins. Arrow orientations represent spin states $s =$up or $s =$ down The highlighted area indicates the correlation length in the horizontal direction of the network. H_1 and H_2 represent the left and right halves of the lattice. b) Sketch of an N-neuron population. The variable σ_0 represents the state of a single neuron in the population and $\{\sigma\}$ represents the state of the complementary network, where the network states can be mathematically treated as an effective field with 2^{N-1} parameters.

2^{N-1} expected states while preserving all the neuron correlations.

To quantify such neuronal correlations from neuronal activity data, we use the mutual information that captures linear and non-linear correlations between neurons[1]. [44].For two non-overlapping groups of neurons, $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, the mutual information is defined as,

$$I(\boldsymbol{\sigma}_1;\boldsymbol{\sigma}_2) = \sum_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} p(\sigma_1,\sigma_2)\log_2\frac{p(\sigma_1,\sigma_2)}{p(\sigma_1)p(\sigma_2)} \tag{5.1}$$

where $p(\sigma_1,\sigma_2)$ is the joint probability distribution and $p(\sigma_1)$ and $p(\sigma_2)$ are the marginal probability distributions for σ_1 and σ_2.

Using this measure of neuronal correlations, we define our compression problem as follows: for all possible choices of σ_0 from a neuronal population with N neurons, we will determine the conditional probability $p(\sigma_0|\tilde{\boldsymbol{\sigma}})$, with $\tilde{\boldsymbol{\sigma}}$ a compressed representation of $\boldsymbol{\sigma}$ such that: I) the compressed variable has a significantly reduced number of states, that is $|\tilde{\boldsymbol{\sigma}}| \ll |\boldsymbol{\sigma}|$, and II. the mutual information or correlation between σ_0 and $\boldsymbol{\sigma}$ is preserved, that is $I(\sigma_0;\tilde{\boldsymbol{\sigma}})/I(\sigma_0;\boldsymbol{\sigma}) \approx 1$. To find these compressed representations, we will use the compression bottleneck method in the hard clustering limit.

5.1.1 Compression bottleneck method

As previously mentioned, the number of states in $\boldsymbol{\sigma}$ grows as 2^N, with N the number of neurons. For example, in a 1-neuron population, we have a total of 2^1 states, that is, $\sigma = \{1,0\}$. If we include another neuron, we get a 2-neuron population with 2^2

[1] Consider the following example: For the random variable $X[x]$ with $x = \{-1,0,1\}$ and probabilities $P[X = x] = 1/3$, we define the variable $Y = X^2$. If we use the covariance as a measure of correlation between X and Y we get $\text{Cov}(X,Y) = 0$ despite the fact that these variables are statistically dependent. In contrast, if we use the mutual information, $I(X;Y)$, we get a non-zero correlation as expected [44].

states, that is, $\boldsymbol{\sigma} = \{00, 01, 10, 11\}$. Regarding the conditional probability $p(\sigma_0|\boldsymbol{\sigma})$, we ask whether we can reduce this number of states while preserving an accurate description of σ_0.

Since we know that correlations can be generally quantified by the mutual information, $I(\sigma_0; \boldsymbol{\sigma})$, we can compress the states in $\boldsymbol{\sigma}$ by finding an optimal mapping of $\boldsymbol{\sigma} \to \tilde{\boldsymbol{\sigma}}$ that preserves the mutual information $I(\sigma_0; \boldsymbol{\sigma})$. This problem, known as the information bottleneck, is defined by the Markov chain $\sigma_0 \leftarrow \boldsymbol{\sigma} \leftarrow \tilde{\boldsymbol{\sigma}}$, such that the optimal mapping is determined though the conditional probability $p(\tilde{\sigma}|\sigma)$ [106]. That is, we attempt to minimize

$$\mathcal{L}[p(\tilde{\boldsymbol{\sigma}}|\boldsymbol{\sigma})] = I(\boldsymbol{\sigma}; \tilde{\boldsymbol{\sigma}}) - \beta I(\tilde{\boldsymbol{\sigma}}; \boldsymbol{\sigma}) \tag{5.2}$$

where $\tilde{\boldsymbol{\sigma}}$ is the compressed representation of $\boldsymbol{\sigma}$, and β is a Lagrange multiplier that determines the amount of mutual information preserved. The solution to this variational problem is,

$$\begin{cases} p(\tilde{\sigma}|\sigma) = \frac{p_t(\tilde{\sigma})}{Z_t(\beta,\sigma)} \exp(-\beta D_{KL}\left[p(\sigma_0|\sigma)||p(\sigma_0|\tilde{\sigma})\right]) \\ p(\sigma_0|\tilde{\sigma}) = \frac{1}{p(\tilde{\sigma})} \sum_{\sigma_0} p(\sigma_0, \sigma) p(\tilde{\sigma}|\sigma) \\ p(\tilde{\sigma}) = \sum_{\sigma} p(\sigma) p_t(\tilde{\sigma}|\sigma), \end{cases} , \tag{5.3}$$

where the D_{KL} is the Kullback–Leibler divergence [106]; the details on the derivation are in Appendix C. These solution depends on the specific value of β and the cardinality (number of states) of the mapping, $|\tilde{\boldsymbol{\sigma}}|$. The structure of these solutions is sketched in the information plane of Fig. 19a, where each curve corresponds to a fixed cardinality $|\tilde{\sigma}|$ and a continuous variation of the lagrange multiplier β. Note that at each cardinality, the solution saturates in the limit of $\beta \to \infty$, which is known as the hard clustering limit [107]. In this limit, each state in $\boldsymbol{\sigma}$ is assigned to a unique state in $\tilde{\boldsymbol{\sigma}}$, that is,

$$\begin{cases} p(\tilde{\sigma}|\sigma) = \begin{cases} 1 & \text{if } \sigma \in \tilde{\sigma} \\ 0 & \text{otherwise} \end{cases} & \forall \sigma \in \boldsymbol{\sigma} \\ p(\sigma_0|\tilde{\sigma}) = \frac{1}{p(\tilde{\sigma})} \sum_{i=1}^{|\tilde{\sigma}|} p(\sigma_i, \sigma_0) & \forall \sigma_0 \in \boldsymbol{\sigma_0} \\ p(\tilde{\sigma}) = \sum_{i}^{|\tilde{\sigma}|} p(\sigma_i). \end{cases} , \tag{5.4}$$

where we have taken the limit of $\beta \to \infty$ in Eq. 5.3 (see Appendix C). Each state in $\boldsymbol{\sigma}$ is assigned to one state in $\tilde{\boldsymbol{\sigma}}$, such that $|\tilde{\boldsymbol{\sigma}}| \ll |\boldsymbol{\sigma}|$. That is, the conditional probability $p(\tilde{\sigma}|\sigma)$ is nonzero only if the state σ is clustered in $\tilde{\sigma}$. Fig. 19b shows three different examples of typical solutions in this limit. Each curve shows the normalized information as a function of the cardinality of $\tilde{\sigma}$ – each point corresponds to the saturation value of each curve in the original problem sketched in Fig. 19a. The green squares show an example of sub-optimal

compression, where the information curve does not reach the plateau at small cardinalities, meaning that the compressed variable do not preserve all the information about σ_0. The blue curve shows an early saturation curve with an optimal preservation of information. The orange curve represents the intermediate case, in which compression is still significant.

In this work, we are interested in compressing σ while preserving most information about σ_0. Consequently, we will use the solution of the information bottleneck method in the hard clustering limit. To quantify the compression quality, we introduce the fractional information variable,

$$\mathrm{FI} = \frac{I(\sigma_0;\tilde{\boldsymbol{\sigma}})}{I(\sigma_0;\boldsymbol{\sigma})}, \tag{5.5}$$

with values ranging between 0 and 1, indicating total loss or preservation, respectively, of the original information of σ_0.

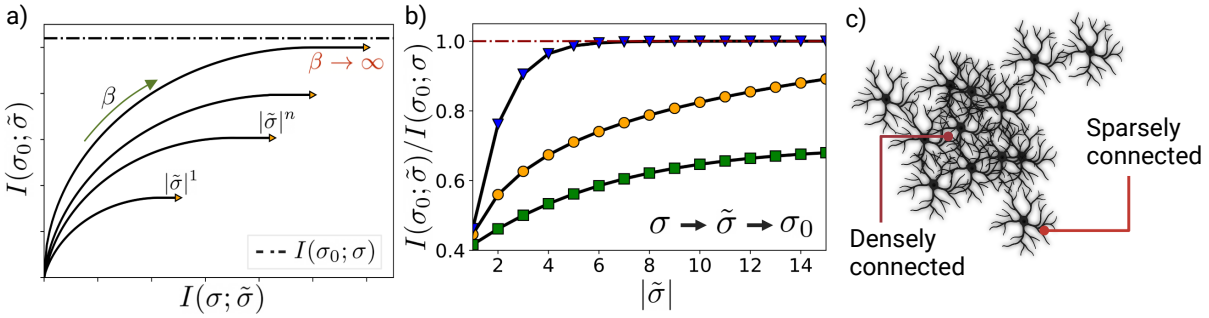

Figure 19 –

a) Structure of the IBM solutions in the information plane. Each curve corresponds to different values of β with a fixed cardinality $|\tilde{\sigma}|$. In the limit of $\beta \to \infty$ (hard clustering limit) the information curve saturates. b) Typical solutions of the hard clustering limit where the variable $\boldsymbol{\sigma}$ is compressed into the variable $\tilde{\boldsymbol{\sigma}}$ while maximizing the mutual information $I(\sigma_0;\boldsymbol{\sigma})$. The green curve corresponds to a compression with early saturation and poor preservation of the mutual information. The orange curve corresponds to a late saturation curve. The blue curve corresponds to an early and optimal preservation of the mutual information.

5.2 Neuronal interaction models

After compression, we reduce the number of states in $\boldsymbol{\sigma}$ and consequently, the number of parameters describing the effective field acting on σ_0, $h_{\mathrm{eff}}(\boldsymbol{\sigma})$ (see Fig. 18b). A fully detailed description of such an effective field is given by,

$$h_{\mathrm{eff}}(\boldsymbol{\sigma}) = \sum_{i|\sigma_i \in \boldsymbol{\sigma}} h_i \sigma_i. \tag{5.6}$$

Some inference methods, such as maximum entropy, are useful to infer these weight parameters, h_i. Nevertheless, using this detailed description has several inconveniences. First, the total number parameters grows quickly with the number of neurons making the inference problem computationally expensive. Second, the number of experimental samples needed to infer reliably these weights grows quickly with the population size. Third, there is a high risk of overfitting in the models.

Conventional models of neuronal activity use approximations of this effective field based on assumptions about neuronal interactions, that is,

$$h_{eff}\left(\{\sigma_{\mathrm{j}}\}\right) = h_0 + \sum_{\mathrm{j}} J^{(2)}_{0\mathrm{j}}\sigma_{\mathrm{j}} + \frac{1}{2}\sum_{\mathrm{j},k} J^{(3)}_{0\mathrm{j}k}\sigma_{\mathrm{j}}\sigma_{\mathrm{k}} + \cdots , \tag{5.7}$$

where each term determines a model with different order of interactions. For instance, the first term corresponds to the independent model, where there is a constant term that describes all neuronal interactions. A truncation up to the second term provides us with the pairwise model, where interactions between neuron pairs are considered. Higher order interactions are contained in the subsequent terms of the expansion.

Some studies of the mouse hippocampus [103] have shown that the pairwise model successfully describes and predicts the neuronal correlations observed experimentally in groups of up to hundreds of neurons. Nevertheless, similar approaches applied to populations of retinal ganglion cells [98] show that the pairwise model fails at predicting experimental observations, demanding higher-order interaction terms that are computationally expensive to deduce.

In addition to the computational and experimental limitations associated with this expansion, it is not clear that these models, inspired in condensed matter physics, provide an optimal fit for neuronal populations. The reason is that neuronal connectivity does not have a fixed spatial structure (as in the spin-lattice example), impeding the generalization of local interactions to the whole network. Figure 19c provides an example of a population with spatially disorganized, dense and sparse connections.

To test models of neuronal activity, we can use ideas of information theory. More specifically, each truncation of the series in Eq. (5.7) leads to a probabilistic model, $p(\sigma_0|h^{(i)}_{\mathrm{eff}})$, with some codification of the data that we can quantify using the mean code length definition,

$$L = \langle -\log_2 p(\sigma_0|h^{(i)}_{\mathrm{eff}})\rangle. \tag{5.8}$$

The optimal codification from the true model of the data, L_{min}, has a mean code length equal to the entropy. Other models lead to mean code length values, L_{approx}, larger

than the entropy. For instance, each truncation of order i in the series expansion of Eq. (5.7) provides an approximate model with the mean code length,

$$\begin{aligned} L^{(i)}_{\text{approx}} &= \left\langle -\log_2 \frac{\exp(-\sigma_0 h^{(i)}_{\text{heff}}(\sigma))}{\sum_{\langle\sigma,\sigma_0\rangle}\exp(-\sigma_0 h^{(i)}_{\text{heff}}(\sigma)} \right\rangle \\ &= \frac{-\langle\sigma_0 h_{\text{eff}}\rangle}{\log 2} + \langle\log_2(1+\exp h_{\text{eff}})\rangle\,. \end{aligned} \tag{5.9}$$

Several experimental studies [98, 100, 103] demonstrate that neurons from large populations are structurally correlated, meaning that their activity is far from being independent. Consequently, we expect that the truncation up to the first term in the expansion of Eq. (5.7) – the independent model – provides the poorest model of the data, with the mean code length,

$$L_{\text{approx}} = \frac{-h_{\text{eff}}\langle\sigma_0\rangle}{\log 2} + \log_2(1+\exp h_{\text{eff}})\,. \tag{5.10}$$

As previously mentioned, we know that $L^{(i)}_{\text{approx}} < L_{\text{min}}$. Similarly, $L_{\text{ind}} < L^{(i)}_{\text{approx}}$. Consequently, we can quantify the model quality by defining the coding cost variable,

$$C = \frac{L^{(i)}_{\text{approx}} - L_{min}}{L_{\text{ind}} - L_{min}}, \tag{5.11}$$

with values ranging from zero to unity.

Equations (5.5) and (5.11) are defined in terms of the same conditional probability $p(\sigma_0|h_{\text{eff}})$, that depends on the specific model of h_{eff}. This means that the compression approach provide us with a model that we can relate to the mean code definition as follows,

$$\begin{aligned} \text{FI} &= \frac{I(\sigma_0;\tilde{\sigma})}{I(\sigma_0;\sigma)} \\ &= \frac{\langle\log_2 p(\sigma_0|\tilde{\sigma})\rangle - \langle\log_2 p(\sigma_0)\rangle}{\langle\log_2 p(\sigma_0|\sigma)\rangle - \langle\log_2 p(\sigma_0)\rangle} \\ &= \frac{-L_{\text{comp}} + L_{\text{ind}}}{L_{\text{ind}} - L_{\text{min}}} \\ &= 1 - C + \frac{L_{\text{app}} - L_{\text{comp}}}{L_{\text{ind}} - L_{\text{min}}}. \end{aligned} \tag{5.12}$$

This means that in the limit of proper compression, $C = 1 - FI$, we expect a coding cost near zero, providing a means to quantitatively compare the performance between the series expansion and the compression approaches.

5.3 Experimental data: ganglion cells in the salamander retina

As previously discussed, we are interested in studying the neuronal activity of ganglion cell populations. In this section, we describe in greater detail the data we intend to analyze, which has been collected by our experimental collaborators in Berry's lab at Princeton University and published in Ref. [98].

The data correspond to the neuronal activity of GCs in the tiger salamander retina. More specifically, the population is composed of $N = 160$ neurons responding to naturalistic visual stimuli. The neuronal activity was recorded using a multi-electrode array which provides a continuous trace of the membrane potential as a function of time. As previously discussed, we can discretize these traces by considering the neuronal activity within specific time windows of size Δt. A statistical analysis of these data leads to the choice of $\Delta t \approx 20ms$; further details are provided in refs [98, 99, 108].

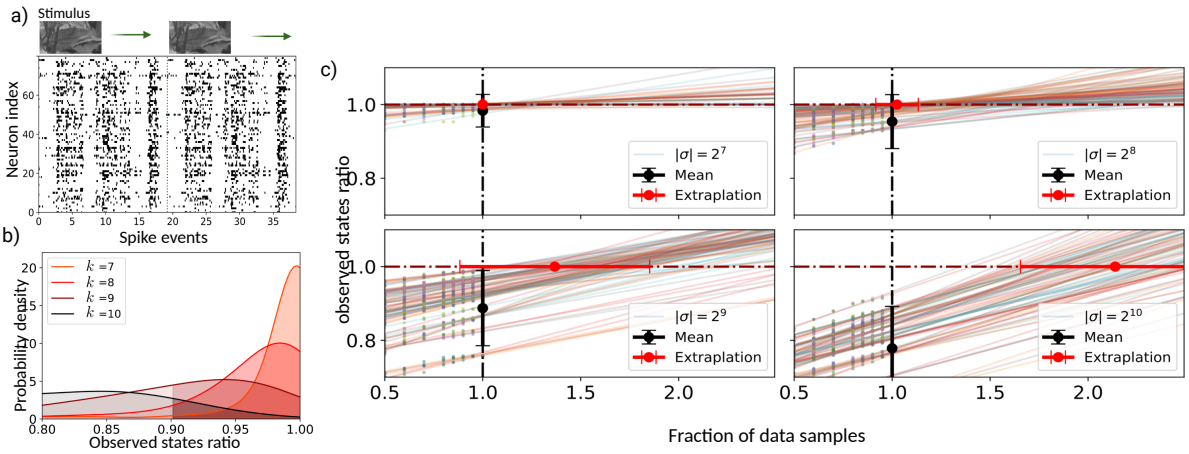

Figure 20 –

a) Raster plot of 80 neurons in the retinal population responding to repeated naturalistic movies of 19.2s duration each. The green dashed line marks the end of the first movie. b) Distribution of the ratio of observed states, $R_s = S_{\text{exp}}/S$, for different groups of neurons of size k. S_{exp} corresponds to the number of states observed in the data set and S to the number of expected states, 2^k. The distribution is calculated over random groups of k neurons from the retinal population. c) Finite size study of the states ratio as a function of the data sample fractions. The black marker corresponds to the mean value of the observed states over random groups of size k. The red marker corresponds to the linear extrapolation of the fraction of data samples needed to observe the expected number of states, 2^k.

The visual stimulus used in the experiment was a movie composed of grey-scale naturalistic images typical from the salamander habitat (see Fig. 20a). The movie was systematically repeated 297 times, leading to approximately two hours of neuronal activity recording time. After discretizing these neuronal activity traces, we got a total of $T_s = 283041$ experimental observation events per neuron.

The inference of the conditional probability, $p(\sigma_0|\boldsymbol{\sigma})$, requires information of the state probabilities $p(\sigma)$, which we can infer from the frequency of each state. If the number of states in $\boldsymbol{\sigma}$ is large, the number of experimental samples needed to get a reliable inference of these probabilities is also large. Consequently, having a limited experimental sample establishes a restriction in the network sizes that can be studied reliably. To establish the network sizes that we can investigate with our retinal data, we analyze different subnetworks of size k. First, we define the ratio of observed states as,

$$R_s = s_{\text{exp}}/s, \tag{5.13}$$

with s_{exp} the number of states observed in the data and s the number of expected states, 2^k. Fig. 20 shows the distribution of this ratio over 150 randomly chosen subnetworks of sizes $k = \{7, 8, 9, 10\}$. For networks of seven neurons, with an expected number of 128 states, we have good statistics. For networks of eight neurons, the statistics worsens, but still, for most subnetworks ($> 90\%$), we get samples of at least 90% of the expected states. For subnetworks of size $k > 8$ the statistics worsen quickly, such that the probability of observing more than 90% of the states decays to less than 0.5.

From these analysis, we conclude that subnetworks of size $k \leq 8$ have information of more than 90% of the expected states. These results are necessary but not sufficient to get a reliable statistics of the states in $\boldsymbol{\sigma}$. In addition, we need to ensure that the inferred state probabilities do not change drastically with the size of the experimental sample. In Fig. 20c, we show a finite size analysis of the ratio of observed states as a function of the fraction of data samples. First, we randomly choose a subnetwork of size k. For that group, we calculate the ratio R_s for different fractions of the data samples, $F = \{0.5, 0.6, \ldots, 0.9\}$. To get an unbiased result, we perform 300 random permutations of the samples, preserving the temporal data structure, per subnetwork. We run this analysis for the same 150 subnetworks previously studied. In the figures, solid lines correspond to the linear extrapolations of each independent case. The black marker corresponds to the mean value of the experimentally observed ratio for all the 150 subnetworks, with all the data. The red marker corresponds to the linear extrapolation of the number of samples needed to observe the total number of expected states, that is $R_s = 1$. We see that for subnetworks of size $k < 8$, the data provide reliable statistics on these states.

5.4 Compression implementation in the GC population

Considering the previous analysis, we reformulate our original problem such that we consider subnetworks instead of the full network of GCs. Since in our approach we are interested in compressing a large number of states, we will consider the largest group that we can reliably investigate, that is, $k = 8$; we denote these subnetworks as $\boldsymbol{\sigma}_i$, with

$i = \{1, \ldots, \frac{N-1}{k}\}$. If these variables $\boldsymbol{\sigma}_l$ are compressible, we can implement a coarse-graining procedure to increase the size of the subnetworks that we can reliably investigate. Fig. 21a shows a sketch of this implementation, where the first coarse-graining step is the compression of networks of size $k = 8$, the second step is the compression of the coarse-grained networks of size $k = 16$, and so on. The number of states of each coarse-grained variable depends on the number of states of the compressed variables in the previous step. This allows to get a reliable sampling of larger subnetworks with fewer expected states.

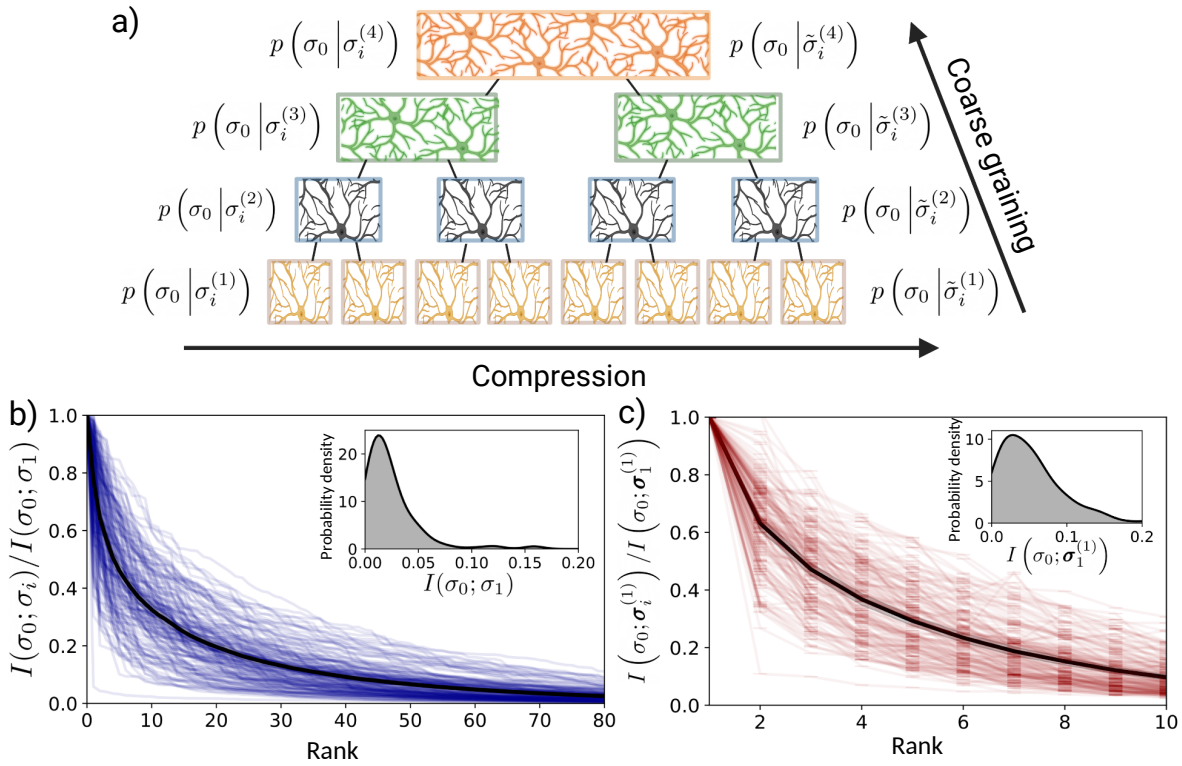

Figure 21 –

a) Coarse-graining and compression diagram. The base level (yellow) represents the groups of $k = 8$ neurons, $\sigma_i^{(1)}$, which are compressed into $\tilde{\sigma}_i^{(1)}$. Upper levels represent the coarse-grained variables, $\sigma_i^{(j)}$, which are compressed into $\tilde{\sigma}_i^{(j)}$. b) Normalized mutual information between pairs of neurons, $I(\sigma_0;\sigma_j)/I(\sigma_0;\sigma_1)$, for all choices of σ_0 in the population; blue lines correspond to the different choices of σ_0 and the black curve corresponds to the mean value. The inset corresponds to the distribution of the information between σ_0 and its most informative partner in the population, σ_1. c) Normalized mutual information between σ_0 and the first 10 groups of $k = 8$ neurons, $\boldsymbol{\sigma}_i$, sharing the most information with σ_0; red lines correspond to the different choices of σ_0 and the black curve corresponds to the mean value. The inset corresponds to the distribution of the information between σ_0 and its most informative partner in the population, $\boldsymbol{\sigma}_1$

As previously discussed, we are interested in compressing $\boldsymbol{\sigma}_l$ into $\tilde{\boldsymbol{\sigma}}_l$ while preserving the mutual information $I(\sigma_0;\boldsymbol{\sigma}_l)$. To form the subnetworks of $k = 8$ neurons, we first pick one neuron σ_0 from the population. Then, we calculate the mutual information

between pairs of neurons, that is, $I(\sigma_0 \mid \sigma_j)$, with $j \in \{1, 2, \ldots, N-1\}$. Fig. 21b shows the normalized mutual information for all choices of σ_0 and the 80 most informative neurons. The plot shows the probability density of the mutual information between each choice of σ_0 and its most informative partner in the population, σ_1.

To create the subnetworks, we rank neurons from the most informative to the least informative with respect to (w.r.t.) σ_0, and subsequently, we divide the sorted neurons into $(N-1)/k$ groups. For instance, the first subnetwork, $\boldsymbol{\sigma}_1$, contains the eight neurons that share the most information with σ_0, the subnetwork $\boldsymbol{\sigma}_2$ contains the next most informative eight neurons and so on. Similarly to the plot for the pairs of neurons, Fig. 21c shows the mutual information curve of each of these subnetworks. The mutual information between these subnetworks and the corresponding neuron σ_0 decreases quickly. This suggests that, on average, each neuron is correlated non-uniformly with the whole network.

5.4.1 Compression implementation

Based on the previous discussion, we implement the first coarse-graining step of compression by following the hard clustering limit of the information bottleneck solution in Eq. (5.4). We start by choosing one neuron σ_0 and ranking the $(N-1)/k$ groups according to the mutual information $I(\sigma_0; \boldsymbol{\sigma}_l)$. Each group has 2^k possible states that we denote as $\boldsymbol{\sigma}_l = \{\sigma_1, \sigma_2, \ldots, \sigma_{2^k}\}$ – note that the notation is the same used to refer the single neurons, but, we highlight that here we refer to the states contained in the groups $\boldsymbol{\sigma}_l$. For each of these groups, $\boldsymbol{\sigma}_l$, we start with some random assignment $\boldsymbol{\sigma}_l \to \tilde{\boldsymbol{\sigma}}_l$ and compute

$$P(\sigma_0; \tilde{\sigma}_i) = \sum_{\sigma_j \in \tilde{\sigma}_i} P(\sigma_0|\sigma_j) P(\sigma_j), \tag{5.14}$$

$$P(\tilde{\sigma}_i) = \sum_{\sigma_j \in \tilde{\sigma}_i} P(\sigma_j), \tag{5.15}$$

with $\sigma_j \in \boldsymbol{\sigma_l}$, $\tilde{\sigma}_i \in \tilde{\boldsymbol{\sigma}}_l$ and $P(\sigma_0|\tilde{\sigma}_i) = P(\sigma_0; \tilde{\sigma}_i)/P(\tilde{\sigma}_i)$ as usual. Then for each particular state we compute the Kullback–Leibller divergence

$$D_{KL}(\sigma_i; \tilde{\sigma}_i) = \sum_{\sigma_0} P(\sigma_0|\sigma_i) \ln \left[\frac{P(\sigma_0|\sigma_i)}{P(\sigma_0|\tilde{\sigma}_i)} \right], \tag{5.16}$$

and reassign

$$\sigma_i \to \arg\min_{\tilde{\boldsymbol{\sigma}}_i} D_{KL}(\sigma_i; \tilde{\sigma}_i). \tag{5.17}$$

Iterating, we arrive at a mapping $\boldsymbol{\sigma}_l \to \tilde{\boldsymbol{\sigma}}_l$ that maximizes $I(\sigma_0; \tilde{\boldsymbol{\sigma}}_l)$ for the corresponding cardinality $|\tilde{\boldsymbol{\sigma}}_l|$. We implement this solution for all cardinalities, $|\tilde{\boldsymbol{\sigma}}_i| \in \{2, 3, \ldots 2^k\}$, and define the optimal cardinality as the one satisfying $\mathrm{FI}(\tilde{\boldsymbol{\sigma}}_i) = 1 \pm \Delta_{\mathrm{FI}}$; with Δ_{FI} the experimentally determined error bars.

5.4.1.1 Two-neuron network

Before applying the compression algorithm to the retinal population, we analyze a simpler system of $k = 2$ neurons to develop a better intuition regarding the approach. This subnetwork $\boldsymbol{\sigma}$ has 2^2 states which we will compress, if possible, while preserving the information about the neuron σ_0. We study the two cases shown in Tables a and b in figure 22. In both cases, we set the state probabilities $p(\sigma_j) = 0.25$ and $p(\sigma_0) = 0.5$ with $\sigma_j \in \boldsymbol{\sigma}$ and $\sigma_0 \in \{0, 1\}$. The mutual information between the neuron σ_0 and the subnetwork $\boldsymbol{\sigma}$ is,

$$I(\sigma_0; \boldsymbol{\sigma} = \{\sigma_1, \sigma_2\}) = \sum_{\{\sigma\}} p(\sigma_0|\{\sigma_1, \sigma_2\}) \log_2 \frac{p(\sigma_0|\{\sigma_1, \sigma_2\})}{p(\sigma_0)p(\{\sigma_1, \sigma_2\})}, \tag{5.18}$$

that is 1 bit for case 1 (table a) and 0.57 bits for case 2 (table b). In case 1, we see that since the states $\sigma = \{(0, 0), (1, 0), (1, 1)\}$ are completely redundant about σ_0, we can compress them into a unique state preserving the single bit of information. Following equations (5.4), we find $|\tilde{\sigma}| = 2$, with

$$p(\tilde{\sigma} = 0) = p(\sigma = \{0, 0\}) + p(\sigma = \{1, 0\}) + p(\sigma = \{1, 1\}) = 0.75 \tag{5.19}$$

$$p(\tilde{\sigma} = 1) = p(\sigma = \{0, 1\}) = 0.25, \tag{5.20}$$

and the conditional probabilities shown in the table c in Fig. 22. The mutual information between σ_0 and the new compressed variable $\tilde{\sigma}$ remains equal, that is $I(\sigma_0; \tilde{\sigma}) = 1$ bit.

a)

$p(\sigma_0\|\sigma)$	$\sigma_0 = 0$	$\sigma_0 = 1$
$\sigma = \{0,0\}$	1	0
$\sigma = \{1,0\}$	1	0
$\sigma = \{0,1\}$	0	1
$\sigma = \{1,1\}$	1	0

b)

$p(\sigma_0\|\sigma)$	$\sigma_0 = 0$	$\sigma_0 = 1$
$\sigma = \{0,0\}$	1	0
$\sigma = \{1,0\}$	1	0
$\sigma = \{0,1\}$	0.2	0.8
$\sigma = \{1,1\}$	0.6	0.4

c)

$p(\sigma_0\|\tilde{\sigma})$	$\sigma_0 = 0$	$\sigma_0 = 1$
$\tilde{\sigma} = 0$	1	0
$\tilde{\sigma} = 1$	0	1

d)

$p(\sigma_0\|\tilde{\sigma})$	$\sigma_0 = 0$	$\sigma_0 = 1$
$\tilde{\sigma} = 0$	1	0
$\tilde{\sigma} = 1$	0.2	0.8
$\tilde{\sigma} = 2$	0.6	0.4

e)

$p(\sigma_0\|\tilde{\sigma})$	$\sigma_0 = 0$	$\sigma_0 = 1$
$\tilde{\sigma} = 0$	1	0
$\tilde{\sigma} = 1$	0.4	0.6

$p(\sigma_0\|\tilde{\sigma})$	$\sigma_0 = 0$	$\sigma_0 = 1$
$\tilde{\sigma} = 0$	0.8	0.2
$\tilde{\sigma} = 1$	0.2	0.8

Figure 22 –

(a, b) Table of conditional probabilities $p(\sigma_0|\sigma)$ for (a) Case 1 and (b) Case 2. (c, d, e) Table of conditional probabilities of the compressed states for (c) case 1 and (d, e) case 2.

Case 2 represents a non-trivial example with the conditional probabilities shown in Fig. 22b. As before, the states $\sigma = \{(0, 0), (1, 0)\}$ are redundant, and consequently, we can compress them into a unique state while preserving all the mutual information about

σ_0; the conditional probabilities for this compression step are shown in the table d in Fig. 22. Further compression of the remaining states is not trivial. The first table of Fig. 22e shows the conditional probabilities after compressing $\boldsymbol{\sigma}$ into $\tilde{\sigma}_0 = \{(0,0),(1,0)\}$ and $\tilde{\sigma}_1 = \{(0,1),(1,1)\}$. The second table of Fig. 22e shows the conditional probabilities after compressing $\boldsymbol{\sigma}$ into $\tilde{\sigma}_0 = \{(0,0),(1,0),(1,1)\}$ and $\tilde{\sigma}_1 = \{(0,1)\}$. The mutual information in both cases decreases to 0.51 bits and 0.46 bits respectively. This example shows that the outcome of the compression is strongly related to the network correlation structure.

5.5 Compression of GCs

The previous examples provide some intuition on how the compression algorithm works in neuronal networks with different correlation structures. Next, we apply our approach to the first coarse-graining step of the ganglion-cell population previously described. Specifically, we ask whether all the states in $\boldsymbol{\sigma}_i$ (see the raster plot of Fig. 23) are necessary to describe the conditional probability $p(\sigma_0|\boldsymbol{\sigma}_i)$. As previously shown, this compression is strongly related to the structure of interactions within the retinal population. We seek to compress the variable $\boldsymbol{\sigma}_i$ into the variable $\tilde{\sigma}$, while preserving all the relevant bits about σ_0.

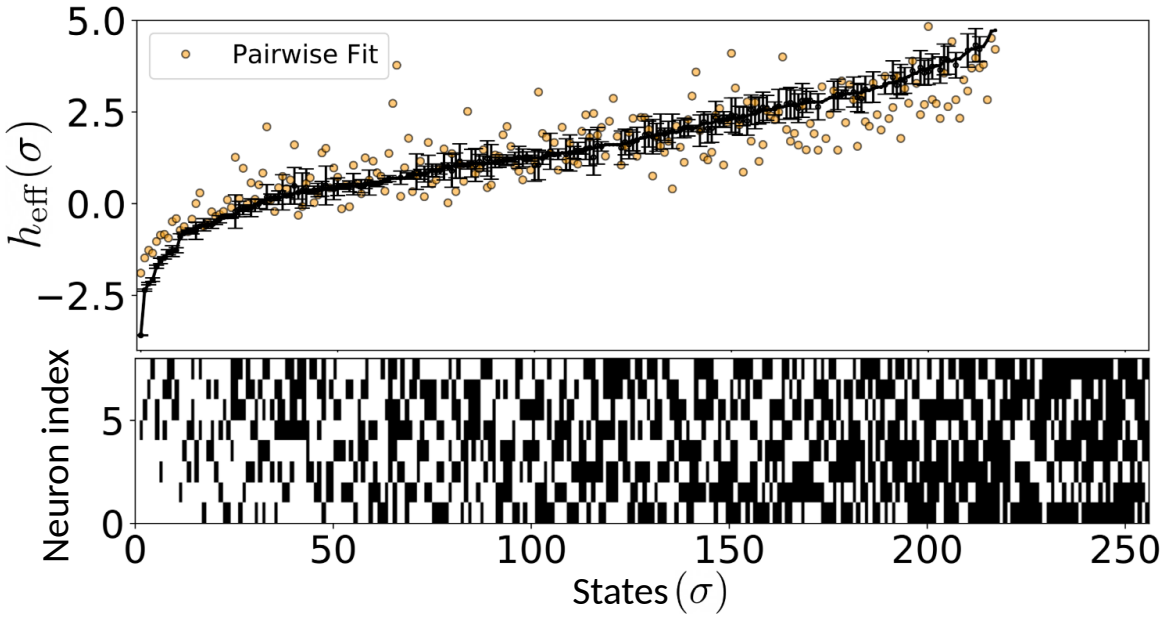

Figure 23 –

The effective field $h_{eff}(\{\sigma_\text{j}\})$ as a function of the states $\{\sigma_\text{j}\}$, in rank order. Mean, with error bars estimated from the standard deviation across random halves of data (black), and best least squares fit to Eq (5.7) truncated at third order (orange circles). Ranked states are represented in the figure below the trace, with black indicating that a neuron is "on," $\sigma_\text{j} = 1$. States at far right are not observed in data.

Fig. 24b shows the fractional information curve as a function of the number of states or cardinality, M, of the compressed variable $\tilde{\sigma_0}$. The error bars are calculated using a finite-size analysis of the data, as discussed in Appendix D. For this subnetwork $\boldsymbol{\sigma}_i$ and neuron σ_0, we find that the optimal number of states is $M = 8$, at which the fractional information reaches unity within experimentally intrinsic error bars. Now, we repeat this compression procedure for all possible choices of σ_0 and their corresponding first group of $k = 8$ most informative neurons. To show the overall result, we calculate the probability density of the fractional information using a weighted kernel density estimator (wKDE); details on this method are provided in Appendix A. As shown in Fig. 24c, the number of states leading to an optimal compression at the population level is $M \approx 10$.

Estimates of mutual information come with errors, and so statements about the number of states needed to capture a given fraction of the information also have uncertainty. For each choice of σ_0 and $\boldsymbol{\sigma}_l$, estimates of FI are accompanied by an error $\Delta_{FI}(\sigma_0, \boldsymbol{\sigma}_l)$, and as a global measure Δ_{FI} we take the median of these errors (see inset of Fig. 25b). If we choose a fixed number of states M for the compression, then across all choices of σ_0 and $\boldsymbol{\sigma}_l$ we will find a fraction D_{FI} for which the estimate of FI is larger than $1 - \Delta_{FI}$, i.e. the information captured is within errors of the information available. Fig. 25a highlights this interval $1 - \Delta_{FI}$ in the probability densities of the fractional information for different number of states M. Fig. 25b shows the plot of the fraction D_{FI} as a function of the number of states M. The colored markers correspond to the same probability densities shown in the plot of Fig. 25a.

To compare the performance of our compression method with the performance of the expansion defined in Eq.5.7, we now calculate the coding cost quantity of Eq. (5.11). For each subnetwork $\boldsymbol{\sigma}_l$ we can calculate the effective field as a function of the states following the equation,

$$h_{eff}\left(\boldsymbol{\sigma}_l\right) = \ln\left[\frac{P\left(\sigma_0 = 1|\boldsymbol{\sigma}_l\right)}{P\left(\sigma_0 = 0|\boldsymbol{\sigma}_l\right)}\right]. \tag{5.21}$$

The black markers in Fig. 23 show the effective field for a randomly chosen σ_0 and its group of $k = 8$ most informative neurons, $\boldsymbol{\sigma}_l$. The error bars are calculated using random halves of the data, as described in Appendix D. The raster plot below represents each state sorted by the corresponding effective field value above; the states at the far right are not observed in the data and consequently do not have any inferred effective field.

To calculate the coding cost, we infer the series parameters using 60% of the data and calculate the coding cost on the other 40%. We iterate 100 times to get the corresponding error bars. For instance, for the pairwise model in Eq. (5.7), the effective field is

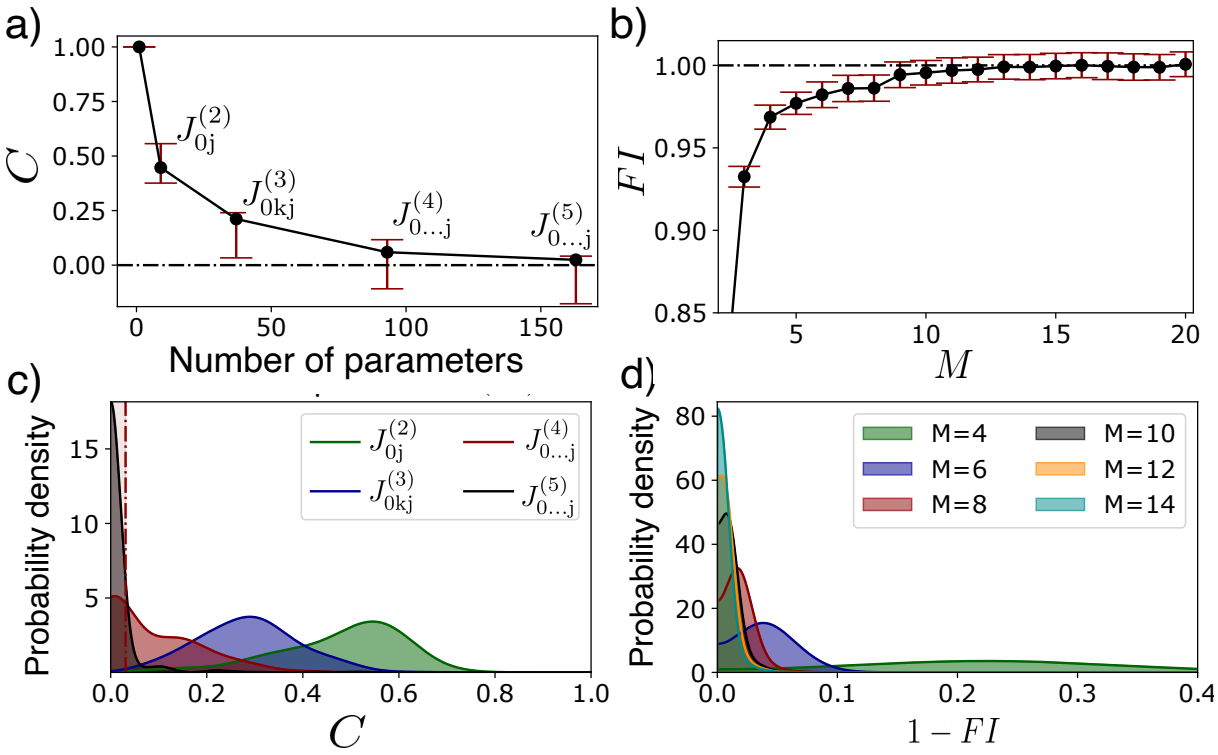

Figure 24 –

Series expansion [see Eq. 5.7] vs compression. (a) Coding cost [Eq (5.12)] as a function of the number of parameters for the series expansion in Eq (5.7). Black points from analysis with all data, error bars are standard deviation across random choices of learning from 60% of the data and testing on the remaining 40%. (b) Fraction of mutual information captured as a function of the number of states M in the compressed representation $\tilde{\sigma}_i$. Error bars from analyses of random subsets of the data. (c) Coding cost probability density over all possible choices of σ_0. Each curve corresponds to a different order of truncation, $J_{0j}^{(i)}$, of the expansion [Eq. (5.7)]. (d) Probability density of the fractional information over all possible choices of σ_0. Each curve correspond to a different value of M. We used a weighted-KDE method for the inference of the probability densities, considering the measured error bars of each choice of σ_0.

$$h_{eff}\left(\boldsymbol{\sigma}_i\right) = h_0 + \sum_{\mathrm{j}} J_{0\mathrm{j}}^{(2)} \sigma_{\mathrm{j}}. \tag{5.22}$$

We fit the $k+1$ parameters using a least square method. Orange markers in Fig. 23 show the model predictions. With these parameters, we calculate the coding cost on the other 40% of the data. We repeat this calculation for 100 random permutations of the data set to estimate uncertainties. Fig. 24a shows the coding cost curve for the same choice of σ_0 and $\boldsymbol{\sigma}_i$ used to calculate the compression of Fig. 24b. As previously discussed, at proper compression we expect the coding cost to be near zero, which in this case is accomplished with models with more than 150 parameters, representing high order interactions. To summarize, in this network we can describe the influence of $K = 8$ neurons on one neuron using just $M = 11 - 15$ parameters, but this most efficient description does not correspond

to a simple choice of pairwise or other low order interactions.

As shown in Fig. 24c, we calculated the probability density of the coding cost for all possible choices of σ_0 and their corresponding group of $k = 8$ most informative neurons. Similarly to the compression case, we find that our results for the single neuron also hold at the population level. On one hand, we find that only high order interaction models lead to a good encoding of the retinal activity data. On the other hand, with our compression approach we show that only a few degrees of freedom are necessary to describe this population.

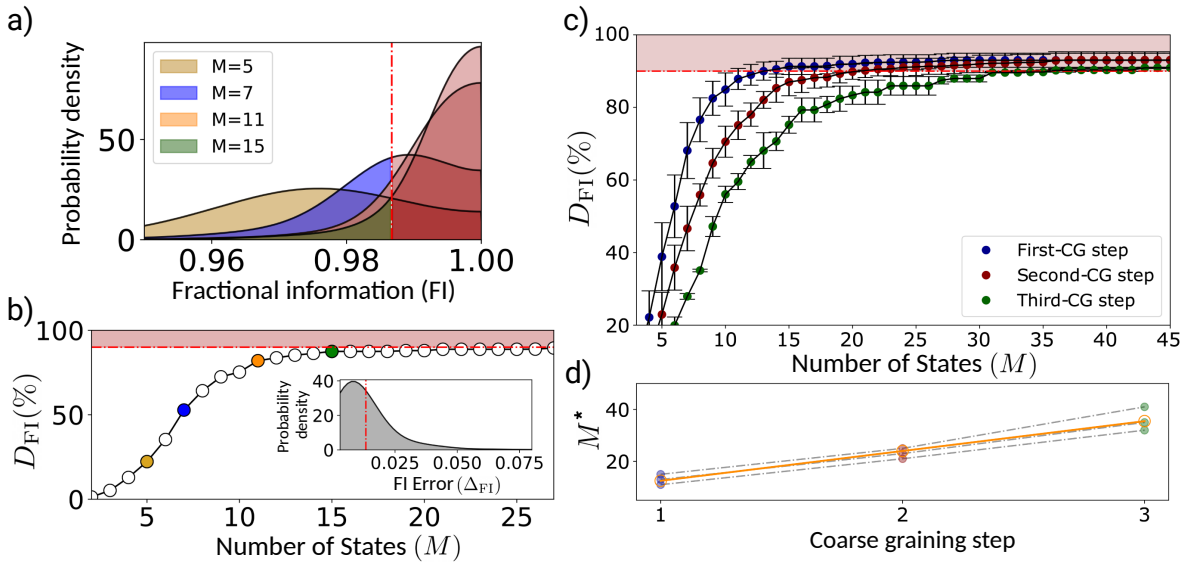

Figure 25 –

(a) Probability density of the fractional information of the first group, $\boldsymbol{\sigma}_1^{(1)}$, sharing the most information with σ_0, for all choices of σ_0. Each color represents a different cardinality of the compressed variable $\tilde{\boldsymbol{\sigma}}_i$. The red dashed line represents the threshold value $FI = 1 - \Delta_{\mathrm{FI}}$; the distribution red area represents the population percentage, D_{FI}, exhibiting a proper compression with the corresponding cardinality M. (b) Percentage of the population with proper compression, $D_{\mathrm{FI}}(\%)$, as a function of the number of states M. The inset shows the distribution of the different error bars, Δ_{FI}, over all choices of σ_0. (c) Fraction of cells σ_0 and groups $\boldsymbol{\sigma}_i$ such that compression into M states captures the available mutual information, within error bars. Successive coarse-graining steps as described in the text. We define M^* as the minimum number of states needed to achieve complete compression in 90% of the cases. (d) Minimum number of states M^* as a function of the coarse-graining step. Dashed curves (grey) correspond to different compression iterations, which vary because of noise in our estimates, together with linear fit (orange)

We can perform this analysis not only for interactions between a single cell σ_0 and its K most informative partners, $\boldsymbol{\sigma}_1 = \{\sigma_1, \ldots, \sigma_8\}$, but also for interactions with successively less informative groups $\boldsymbol{\sigma}_l = \{\sigma_{k(l-1)+1}, \ldots, \sigma_{k(l-1)+k}\}$, and the results are the same. The blue curve in the figure 25c shows the fraction of cells and groups for which

which we can capture the available information within errors with M states, and this fraction is computed across the first eight groups of eight neurons each, for all $N = 160$ choices of σ_0. Note that with $l = 8$ we are covering $0.4N$ of the cells in the entire population, and that $I(\sigma_0; \boldsymbol{\sigma}_l)$ is within error bars of zero for $l > 8$. We conclude that in 90% of all the relevant groups we achieve essentially perfect compression with $M^* \sim 11$ states. In a nutshell, we conclude that at the first coarse-graining step (see Fig. 21a), the data of neuronal activity in retinal populations is compressible, allowing a significant description simplification.

As previously mentioned and sketched in Fig. 21a, the compression $\boldsymbol{\sigma}_i \to \tilde{\boldsymbol{\sigma}}_i$ is reminiscent of the block spin construction in the renormalization group (RG) [109, 110]. We recall that block spins are coarse–grained variables that replace groups of spins. In the present context, it is important to remember that coarse–graining can be thought of as data compression, and vice versa. By analogy with RG, then, we would like to do iterative compression.

Concretely, we are focused on a variable σ_0 and have ordered the remaining variables $\sigma_{\rm j}$ by their mutual information with σ_0. Our first coarse-graining step has been to take these variables in groups of $K = 8$, and compress according to the solution of the optimization problem in Eq (5.4), which gives us

$$\begin{aligned} \{\sigma_1, \sigma_2, \cdots, \sigma_8\} &\to \tilde{\boldsymbol{\sigma}}_1^{(1)} \\ \{\sigma_9, \sigma_{10}, \cdots, \sigma_{16}\} &\to \tilde{\boldsymbol{\sigma}}_2^{(1)} \\ &\cdots \ , \end{aligned} \tag{5.23}$$

where each of the variables $\tilde{\boldsymbol{\sigma}}_n^{(1)}$ has M states. To iterate, we take pairs of these variables and compress again, e.g.

$$\left(\tilde{\boldsymbol{\sigma}}_1^{(1)}, \tilde{\boldsymbol{\sigma}}_2^{(1)}\right) \to \tilde{\boldsymbol{\sigma}}_1^{(2)}, \tag{5.24}$$

where again the mapping is chosen to maximize the mutual information $I(\sigma_0; \tilde{\sigma}_1^{(2)})$.

We can keep iterating,

$$\left(\tilde{\boldsymbol{\sigma}}_1^{(2)}, \tilde{\boldsymbol{\sigma}}_2^{(2)}\right) \to \tilde{\boldsymbol{\sigma}}_1^{(3)}, \tag{5.25}$$

always with the same principle of choosing the compression that maximizes the mutual information with σ_0.

It is not surprising that successive stages of compression or coarse-graining require more states to capture all the available mutual information, as shown in Fig 25(c,d). What

is surprising is that the minimal number of states M^* seems to grow linearly rather than exponentially as we proceed through multiple stages, as seen in Fig 25d. After three stages, we are describing the interactions of σ_0 with 32 other cells using only $M_3^* = 32$ states. The linear growth of M^* with the number of neurons is explicit evidence that we have tamed the combinatorial explosion, combining the compressibility of interactions with an RG–inspired iteration scheme.

The scaling of M^* is what we might expect in a model with pairwise interactions, or if single neurons coupled only to the total activity of other neurons, but neither of these simplifications is correct. Rather than digging into the details of the compressed states, which will be different for every group of cells, we can ask about the amount of information that we identify as shared between single neurons and their most informative partners, $I(\sigma_0; \tilde{\boldsymbol{\sigma}})$. Since the neurons we are analyzing are the output neurons of the retina, a natural comparison is between this information and the information that individual neurons carry about the visual stimulus, $I(\sigma_0; \mathbf{s})$.

As previously discussed, the experiments done to get the neuronal activity data from the salamander retial population include repeated presentations of the same naturalistic movie (see Fig. 20a). We expect these neuronal responses to contain relevant bits of information about this stimulus movie. Consequently, we can estimate $I(\sigma_0; \mathbf{s})$ directly without any assumptions about which features of the visual stimulus are being encoded, following [111].

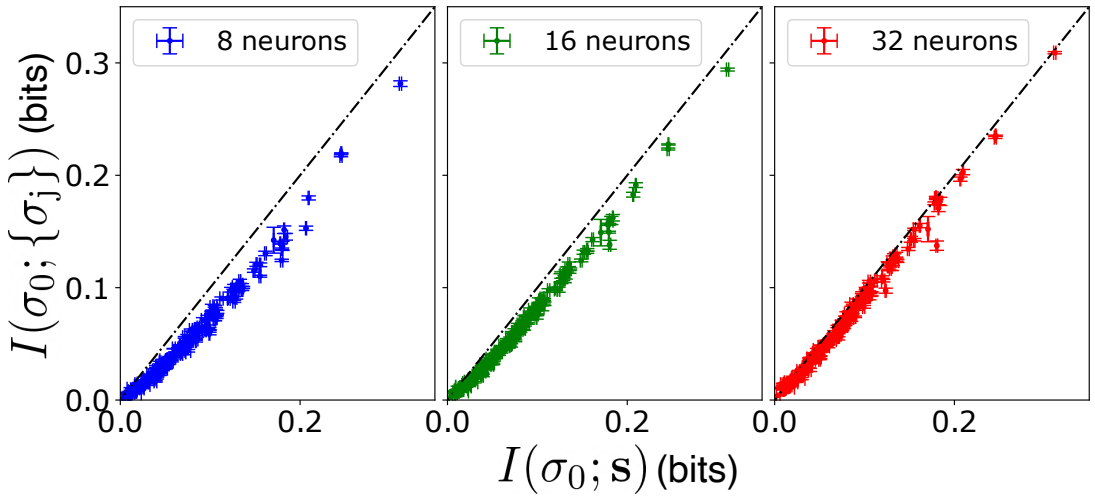

Figure 26 –

Information shared with the network vs information about the stimulus. Plot corresponds to the first (blue, 8 neurons), second (green, 16 neurons) and third (red, 32 neurons) coarse-graining steps. Estimates and errors as in Ref [111].

Figure 26 shows that as we consider larger groups of neurons, the information that single neurons share with the network, $I(\sigma_0; \tilde{\boldsymbol{\sigma}})$, approaches the information that these neurons carry about the visual stimulus, $I(\sigma_0; \mathbf{s})$, reaching equality within (small) error

bars. This has a surprising consequence for our understanding of neural coding. The k neurons that go into defining $\tilde{\boldsymbol{\sigma}}$ carry some information about the visual stimulus, $I(\boldsymbol{\sigma};\mathbf{s})$, and adding the extra neuron σ_0 provides additional information,

$$\begin{aligned}\Delta I(\sigma_0;\mathbf{s}) &= I(\boldsymbol{\sigma},\sigma_0;\mathbf{s}) - I(\boldsymbol{\sigma};\mathbf{s}) \\ &= I(\boldsymbol{\sigma},\sigma_0;\mathbf{s}) - I(\boldsymbol{\sigma};\mathbf{s}) - I(\sigma_0;\mathbf{s}) + I(\sigma_0;\mathbf{s}) \\ &= \sum p(\boldsymbol{\sigma},\sigma_0,s)\left[-\log_2\frac{p(\boldsymbol{\sigma},\sigma_0)}{p(\boldsymbol{\sigma})p(\sigma_0)} + \log_2\frac{p(\boldsymbol{\sigma},\sigma_0|\mathbf{s})}{p(\boldsymbol{\sigma}|\mathbf{s})p(\sigma_0|\mathbf{s})} + \log_2\frac{p(\sigma_0,\mathbf{s})}{p(\sigma_0)p(\mathbf{s})}\right] \\ &= I(\sigma_0;\mathbf{s}) - I(\boldsymbol{\sigma};\sigma_0) + I(\boldsymbol{\sigma};\sigma_0|\mathbf{s}), \end{aligned} \tag{5.26}$$

where $I(\sigma_0;\boldsymbol{\sigma}|\mathbf{s})$ is the (average) mutual information between σ_0 and the network given that visual stimulus is known. The fact that we achieve lossless compression means we can replace $I(\sigma_0;\boldsymbol{\sigma})$ by $I(\sigma_0;\tilde{\boldsymbol{\sigma}})$, and Fig 26 tells us that $I(\sigma_0;\tilde{\boldsymbol{\sigma}}) = I(\sigma_0;\mathbf{s})$, so that

$$\Delta I(\sigma_0;\mathbf{s}) = I(\sigma_0;\boldsymbol{\sigma}|\mathbf{s}). \tag{5.27}$$

If neurons respond independently to the visual inputs, so that all correlations are inherited from the stimulus, then this term would be zero and the neuron at the center of our analysis would be completely redundant with the other K neurons, $\Delta I(\sigma_0;\mathbf{s}) = 0$. Stated in a more positive way, the global correlation structure of the retinal population is such that the extra information carried by individual neurons depends entirely on their departure from conditional independence. Although there is ample experimental evidence for these correlations [112], conditional independence remains a widely held intuition; our analysis indicates that the retina is far from this regime.

As a further test of these ideas we have looked at experiments on a very different network of neurons, in the mouse hippocampus [103,113]. The results, to be described in the next chapter, are very much the same, but perhaps less surprising since maximum entropy models with only pairwise interactions already provide an excellent description of these data, matching the higher order correlations within experimental error [103]. In contrast, as emphasized in Ref [98] for the population of cells in the retina the pairwise models show small but significant deviations from the data, and this has led to the exploration of several alternatives [98, 114, 115]. The point of our discussion is not to identify the correct model, but to understand why *any* simple model can succeed. Previous work has focused on simple forms of the interactions, taking intuition both from physics and from neurobiology. The arguments given here suggest, strongly, that what is essential for a simplified description is the compressibility of interactions [104].

5.6 Discussion

As discussed in the Introduction, the study of the final stage of information processing in the retina might provide important insights into the principles of information

encoding by neuronal networks. The reason that the output signals of GCs are transmitted to downstream regions in the brain that might have similar decoding/encoding principles. Some works have suggested that information is encoded in the interspike time, while others consider that it is the spike which directly carries the encode information [116]. Here, we have focused our study using the latter approach, in which we have taken into account the spike count and the different patterns of activity in the whole retinal population.

Under this scenario, we discussed some conventional models that make assumptions on the neuronal interactions to predict the spike activity. More specifically, we discussed how depending on the specific network, these models require high order interactions to get an accurate prediction of the observed neuronal activity. Based on the limitations of these theoretical approaches, we motivated our work of compression of interactions as a path to build simpler models of neuronal activity. First, we described the population neuronal activity as a conditional problem for which we defined the compression approach allowing a direct comparison with the conventional interaction-based models. We demonstrated that the compression is feasible in GC-populations from the salamander retina, and, we show that the efficient description that we get from our approach does not correspond to a simple choice of pairwise or other low order interactions. As we illustrated in the two-neuron example, the outcome of compression strongly depends on the correlation structure among neurons. The fact that the investigated retinal subnetworks can be consistently compressed into a few states might indicate that there is an intrinsic structure of interactions and correlations that could be analogous to the spatial structure in the spin-lattice systems. Consequently, we conclude that simple models of neuronal activity are feasible but not based in assumptions on neuronal interaction motifs.

Finally, we introduced a coarse-graining procedure as a RG-analog. We showed that instead of an exponential growth in the number of parameters, typical from conventional models of neuronal activity, we find a linear growth. Such a simplification allowed us to investigate our conditional problem with larger subnetworks, providing important insights into the correlation structure of retinal GCs. More specifically, we concluded that the information carried by individual neurons is completely dependent on its departure from conditional independence. That is, conditionally independent neurons are expected to be completely redundant with the other neurons in the population.

To understand how general are our findings, we implement our approach in neuronal populations from the hippocampus, which are also characterized by their spiking activity. The results are described in the next chapter.

6 Compression in a population of hippocampus neurons

The hippocampus of vertebrates is located deep in the brain, such that neuronal activity is driven by inter-neuronal connections among different populations. In contrast with retinal populations, the relation between such neuronal activity and external stimuli is not quite evident. In mouse, nevertheless, experimental works have evidenced the key role of hippocampus in navigation-related tasks, bridging this gap between neuronal activity and external inputs. More specifically, these works have characterized different types of neurons, such as place cells, whose activity is largely correlated to specific physical locations of the moue environment [117].

The mouse hippocampus is composed of thousands of spiking neurons leading to the same theoretical and computational limitations discussed in the previous chapter. That is, network models based on neuronal interactions lead to a combinatorial growth of parameters that become impractical for real neuronal populations. In the previous chapter, we found that the number of states in retinal ganglion populations is significantly compressible, suggesting that we can avoid such combinatorial problem while still preserving a detailed description of neuronal correlations. In this chapter, we investigate whether this compressibility is characteristic of retinal populations or instead, it is a more general property that can be generalized to spiking neuronal populations.

The reader can find all the theoretical formalism of this work in the previous chapter, in which we described in detail the compression bottleneck method and its algorithm implementation in spiking neuronal networks. In this chapter, we first discuss the details of the hippocampus data-set as well as the relevant statistics to define our compression problem. Second, we show the results and discussion.

6.1 Experimental data: mouse hippocampus

In this section, we describe in detail the data that we intend to analyze, which has been collected by our experimental collaborators in Tank's lab at Princeton university and published in ref. [103].

The data correspond to the neuronal activity of $N = 1485$ hippocampus cells in the mouse brain. Neurons were recorded while the animal navigated in a virtual reality set-up, described in Ref. [103]. The mouse head is fixed, but the body is free to move on a rotating ball that is synchronized with a projected video of a maze scenario. This

experiment simulates real navigation and allows the recording of thousands of neurons via calcium imaging[1].

Similarly to the retinal population, the calcium activity was discretized into the two states $\sigma = \{1, 0\}$ for presence and absence of spike respectively [103]. The final number of experimental samples per neuron after discretization was $T_s = 70338$. As described in the previous chapter, our compression approach requires the inference of the state probabilities $p(\sigma)$, whose reliability depends on the data statistics. Consequently, to define our compression problem, first we must define the maximal number of neurons in the subnetworks that can be reliably described with these data. To do that, we calculate the ratio of observed states defined in Eq. (5.13). Fig. 27a shows the distribution of this ratio, over a thousand of randomly chosen subnetworks of sizes $k = \{6, 7, 8, 9\}$. We observe that the neuronal activity in the mouse hippocampus is sparse, keeping us from observing all the possible 2^k states. In Fig. 27b we show the probability of observing a ratio higher than 0.5 as a function of the group size. We see that this probability quickly decreases with the group size.

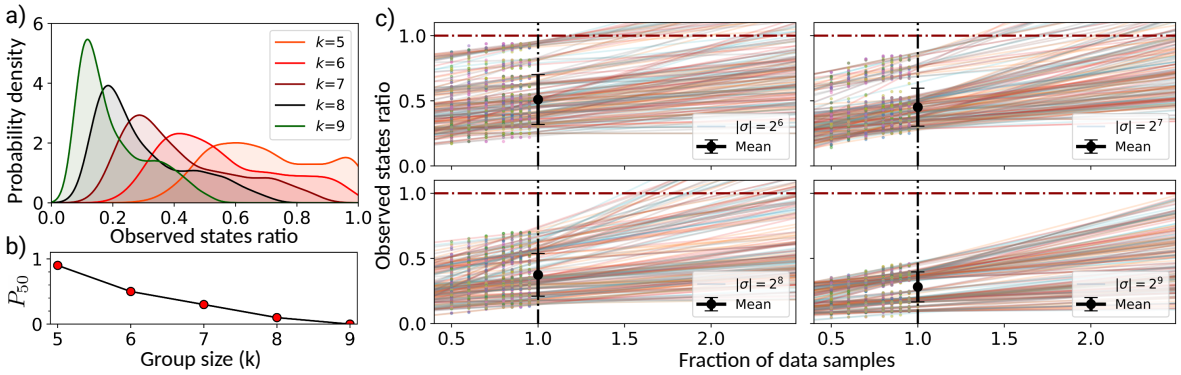

Figure 27 –

a) Distribution of the ratio of observed states over 1000 randomly chosen groups of size k. Different colors correspond to different group sizes. b) Probability, P_{50}, of observing more of the 50% of the states as a function of the group size, k. c) Finite size study of the states ratio as a function of the data sample fractions. The black marker corresponds to the mean value of the observed states over 1000 random groups of size k.

Compared to the retinal population, the number of observed states in the hippocampus is much smaller but still enough to admit a significant compression. To define the optimal group size, we study how sensitive is the ratio of observed states as a function of the number of experimental samples. In Fig. 27c, we show the finite size plots as a function of the fraction of data that we have. The black marker corresponds to the mean value, over all neurons, for the total amount of data that we have. Markers correspond

[1] This experimental technique allows the recording of calcium activity inside different regions of neurons, known as regions of interest (ROI). These recordings allow an indirect measure of the neuronal activity of thousands of ROI, simultaneously. Spikes are related to huge increments of calcium inside the ROI.

to the values calculated with fractions of the data and lines correspond to the linear extrapolation. We calculate the mean slope, a_k, of the normalized data for each group size to quantify the dependence of the states ratio on the number of samples. Similarly to the discussion in the previous chapter, we expect that groups with a strong dependence on the data size will have a poorer inference than groups with less dependence, and because larger groups have more states, we expect that the larger the group size, the larger the statistics variability for our data-set. Indeed, we find that the average slope of these linear extrapolations is $a_k < 0.5$ for $K \leq 8$, corresponding to the increase in variability with size. Larger values of K lead to larger variations and slope a_k. Based on these analysis, we conclude that the most reliable and large groups are $k = 7$ and $K = 8$, for which we observe a large number of states with a weak dependence on the data size. Consequently, we will investigate both cases, contrasting the compression differences when varying the number of neurons within the subnetworks $\boldsymbol{\sigma}$.

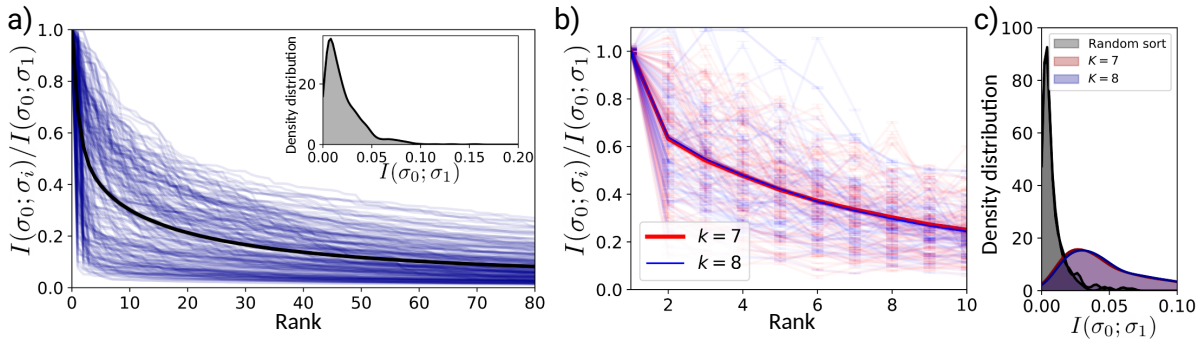

Figure 28 –

a) Normalized mutual information between pairs of neurons, $I(\sigma_0;\sigma_j)/I(\sigma_0;\sigma_1)$, for all choices of σ_0 in the population; blue lines correspond to 100 different choices of σ_0 and the black curve corresponds to the mean value over all 1485 neurons. The inset corresponds to the distribution of the information between σ_0 and its most informative partner in the population, σ_1. b) Normalized mutual information between σ_0 and the first 10 groups of $K = 8$ ($K = 7$) neurons in red (blue), $\boldsymbol{\sigma}_i$, sharing the most information with σ_0; marker lines correspond to 200 different choices of σ_0 and the solid curves in red and blue correspond to the mean value for $K = 7$ and $K = 8$ respectively. c) Distribution of the information between σ_0 and its most informative partner in the population, $\{\sigma\}_1^K$: red (blue) for $k = 7$ ($k = 8$). The distribution in black corresponds to the mutual information for groups with randomly chosen neurons.

6.2 Compression implementation in the hippocampus population

Considering the previous analysis, we will study subnetworks of sizes $k = 7$ and $k = 8$ instead of the full network of hippocampus neurons. Similarly to the case of ganglion neurons, we use the mutual information to define these subnetworks. That is, for a given neuron σ_0, we calculate its mutual information with all other neurons in the

whole population $I(\sigma_0;\sigma_j)$, with $j = \{1, 2, \ldots, N-1\}$. Then, we sort them such that we make disjoint groups starting with the most informative neurons about σ_0. Fig. 28a shows the normalized mutual information for all choices of σ_0 and the 80 most informative neurons. The inset plot shows the probability density of the mutual information between each choice of σ_0 and its most informative partner σ_1 in the population. Based on this mutual information values, we build the disjoint groups of k neurons. Fig. 28b shows the normalized mutual information, $I(\sigma_0;\boldsymbol{\sigma}_j)/I(\sigma_0;\boldsymbol{\sigma}_1)$, for the cases of $k = 7$ (in red) and $k = 8$ (in blue). Fig. 28c shows the distribution of the mutual information between σ_0 and its most informative group $\boldsymbol{\sigma}_1$, which overlaps for $k = 7$ (red) and $k = 8$ (blue). To compare how informative are these sub-networks, we compare them with sub-networks created at random as shown in Fig. 28c. We see that compared to random sub-networks (black), groups of ranked neurons share more information with σ_0.

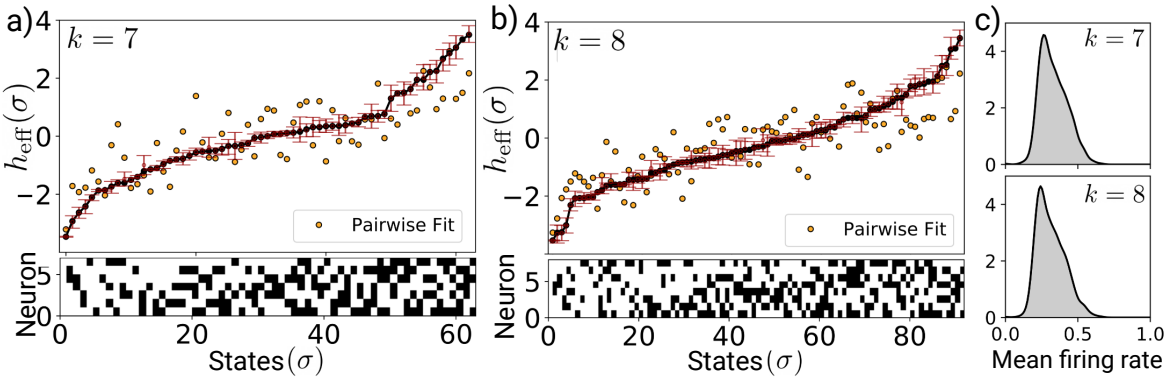

Figure 29 –

The effective field $h_{\text{eff}(\boldsymbol{\sigma})}$ as a function of the states in $\boldsymbol{\sigma}$, in rank order for a) for $k = 7$ and b) for $k = 8$. Mean, with error bars estimated from the standard deviation across random halves of the data (black) and best least squares fir to Eq (5.7) truncated at third order (orange). Ranked states are represented in the figure below the trace, with black indicating that a neuron is "on," $\sigma_{\text{j}} = 1$. Only experimentally observed states are shown. c) Mean firing rate for $k = 7$ (above) and $k = 8$ (below).

Compared to the retinal population (see Fig. 21b), in the hippocampus we observe a larger variability in the mutual information curves among different choices of σ_0. More specifically, we observe more neurons which are weakly correlated to most of the neurons in the population. Regarding the main curve of the groups of k neurons, we observe a slower decrease of the mutual information $I(\sigma_0;\boldsymbol{\sigma})$, meaning that there are more groups sharing significant information with σ_0 related to the most informative group $\boldsymbol{\sigma}_1$. By comparing the inset plots of figures 21c 28b, we observe that, on average, groups in the hippocampus share less bits of information with their corresponding σ_0 than retinal groups. Broadly, we have found that in terms of mutual information, both hippocampus and retinal neuron groups share similar properties. Nevertheless, the error bars, $\Delta_{FI}^{(\text{ret})}$ and $\Delta_{FI}^{(\text{hip})}$ of the retinal and hippocampus populations respectively, are significantly different, such that $\frac{\Delta_{FI}^{(\text{ret})}}{\Delta_{FI}^{(\text{hip})}} = 0.125$.

Because our compression approach is based on the mutual information values and their corresponding error bars, we expect to get a higher uncertainty in the number of optimal states for the hippocampus compared to the retinal population.

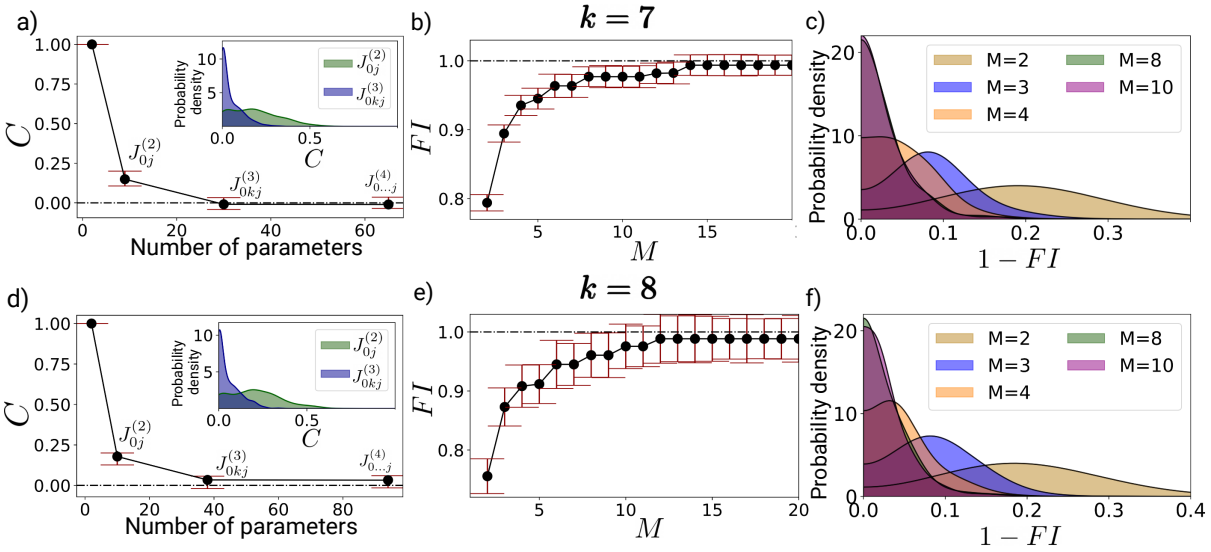

Figure 30 –

Series expansions vs compression for (a, b, c) $k = 7$ and (d, e, f) $k = 8$. (a,d) Coding cost [Eq (5.12)] as a function of the number of parameters for the series expansion in Eq (5.7). Black points from analysis with all data, error bars are standard deviation across random choices of learning from 60% of the data and testing on the remaining 40%. The inset shows the coding cost probability density over all possible choices of σ_0. Each curve correspond to a different order truncation, $J_{0j}^{(i)}$, of the series expansion [Eq. (5.7)]. (b, e) Fraction of mutual information captured as a function of the number of states M in the compressed representation $\tilde{\sigma}_i$. Error bars from analyses of random subsets of the data. (c, f) Probability density of the fractional information over all possible choices of σ_0. Each curve correspond to a different value of M. We used a weighted-KDE method for the inference of the probability densities, considering the measured error bars of each choice of σ_0.

Similarly to the previous chapter, we will compare our compression approach with the effective field expansion of Eq (5.7). The plots a and b in Fig. 29 show the effective field for a given choice of σ_0 with $k = 7$ and $k = 8$ respectively. For $k = 7$ we observe the 50% of the possible states, while for $k = 8$, we observe the 36%. Fig. 29c shows the mean firing rate distribution – the mean number of spikes per time window per neuron, evidencing that neurons in the hippocampus have a sparse neuronal activity with a mean firing rate of ≈ 0.3 spikes per neuron per time window.

Based on this effective field, we calculate the coding cost for different truncations of the series expansion of Eq. (5.7). Fig. 30a shows the result for a given choice of σ_0 and a group of $k = 7$ neurons. Similarly, Fig. 30d shows the coding cost for a group of $k = 8$ neurons. Compared to the retinal population (Fig. 24a,c), we observe that in

the hippocampus, the pairwise model has a better performance, yielding a coding that is nearest to zero. At the population level, we calculate the probability density over all possible choices of σ_0 as shown in the corresponding insets. This shows that, on average, the pairwise model achieves a better performance in describing the neuronal activity of the hippocampus when compared to the retinal population. Still, the coding cost is non zero, meaning that we can do better with other models.

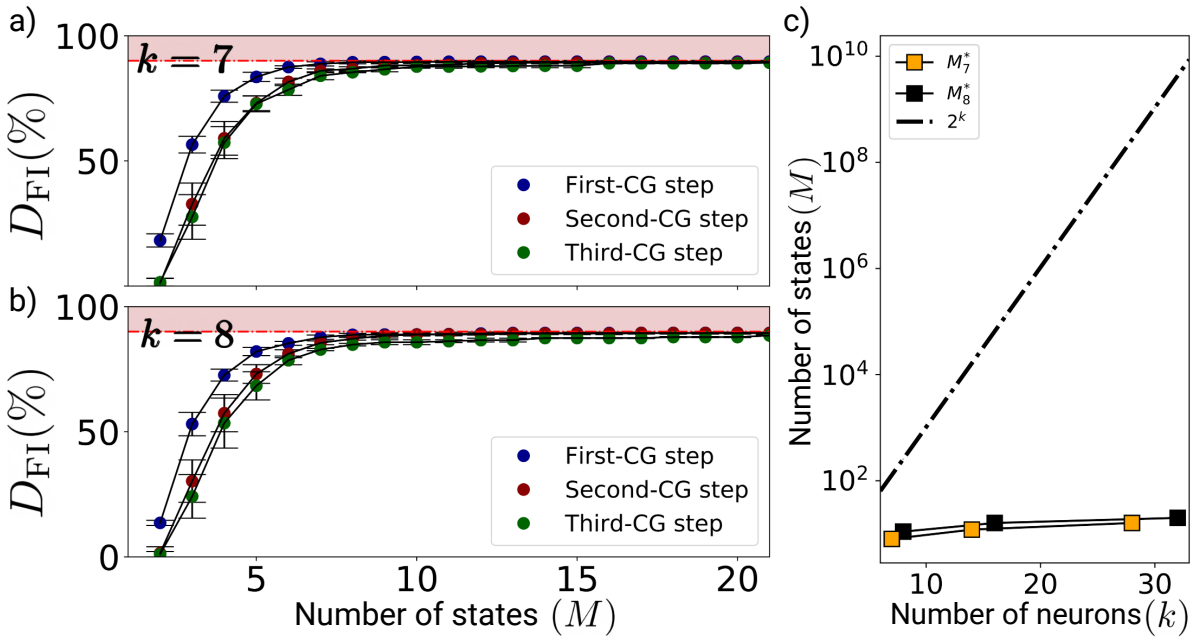

Figure 31 –

Fraction of cells σ_0 and groups $\boldsymbol{\sigma}_l^{(i)}$ such that compression into M states captures the available mutual information, within error bars for a) $k = 7$ and b) $k = 8$. Colors represent successive coarse-graining steps as described in the text. We define M^* as the minimum number of states needed to achieve complete compression in 90% of the cases. c) Number of states as a function of the number of neurons for $k = 7$ (orange markers) and $k = 8$ (black markers). The dashed line shows the possible number of states 2^k.

In the previous chapter, we shown that retinal populations are compressible, allowing us to estimate the minimal number of parameters that we need to efficiently describe the effective field $h_{\mathrm{eff}}(\boldsymbol{\sigma})$ while preserving all bits of information about a single neuron σ_0. Such a compressibility, though, depends on the network correlations and structure. Here, we test whether hippocampus populations are also compressible and whether the minimal number of degrees of freedom is similar to that one found in the retinal population. Similarly, we begin by choosing a neuron σ_0 and its group of K most informative neurons in the population. The plots b and e in Fig. 30 show the results for two randomly chosen σ_0 and the groups of $k = 7$ and $k = 8$ respectively. We observe that in both cases, compression with $FI \pm \Delta_{FI} \approx 1$ is possible with a small number of states. At the population level, we compress the groups of k neurons for all possible choices of σ_0, leading to the probability densities shown in Fig. 30c,f. We observe that with a number of $M \approx 8$ states we already

get a good compression. We highlight that the estimates of mutual information in the hippocampus have larger error bars, leading to larger values of Δ_{FI}. Subsequently, this leads to broad distributions of the fractional information.

These results suggest that compressibility is a more general property of neuronal populations, such that it does not depend on external conditions, e.g. stimuli, or neuronal features, such as brain regions. In addition, our results imply that our compression approach provides a description of the neuronal activity with an order of 10 parameters, which does not correspond to the conventional pairwise model.

As a last step, following equations (5.24)-(5.25), we implement the coarse graining compression, in which we can increase the number of neurons in $\boldsymbol{\sigma}$ while taming the combinatorial growth of parameters. The plots a and b in Fig. 31 show the result of the first three steps for $k = 7$ and $k = 8$ respectively. We find that the optimal number of states do not increase exponentially, rather, we get a saturating growth as shown in Fig. 31c. Compared to the compression results in the retina, we observe that states in the hippocampus population are similarly compressible while preserving most information about the chosen neuron σ_0.

6.2.1 Estimating errors in compression

If compression ideas work in the hippocampus population, several iterations of this compression should lead to a coarse-grained description of the original conditional problem with a much smaller number of parameters than the exponential 2^{N-1}. To get an intuition of our results, we discuss in greater detail the error estimation in compression as well as the expected error introduced by this coarse grained procedure.

Estimating the mutual information from data is done by using finite-size corrections while errors are estimated with variations among random halves of the data (see Appendix D for further details). As previously described, compression is quantified via the fractional information, FI, which is expected to be one when reaching perfect compression. Nevertheless, since estimates of mutual information come with errors, the fractional information also has an error Δ_{FI} propagated from the corresponding information ratio; plots a and b in Fig. **??** show the corresponding error distributions over different choices of σ_0. Consequently,a proper compression of $\{\sigma\}_i^K$ into $\tilde{\sigma}_i^K$ satisfies $FI = 1 \pm \Delta_{FI}$.

Following the ideas of coarse-graining, pairs of these compressed sub-networks $\tilde{\sigma}_i^{(K)}$ can be treated as new sub-networks of size $2K$ with a much fewer number of states than 2^{2K} that can be further compressed. Successive iterations of this coarse-grained compression lead to a large sub-network preserving most bits of information with σ_0. This construction works only if compressed parts of the network when merged preserve most of the information of the corresponding σ_0. More specifically if,

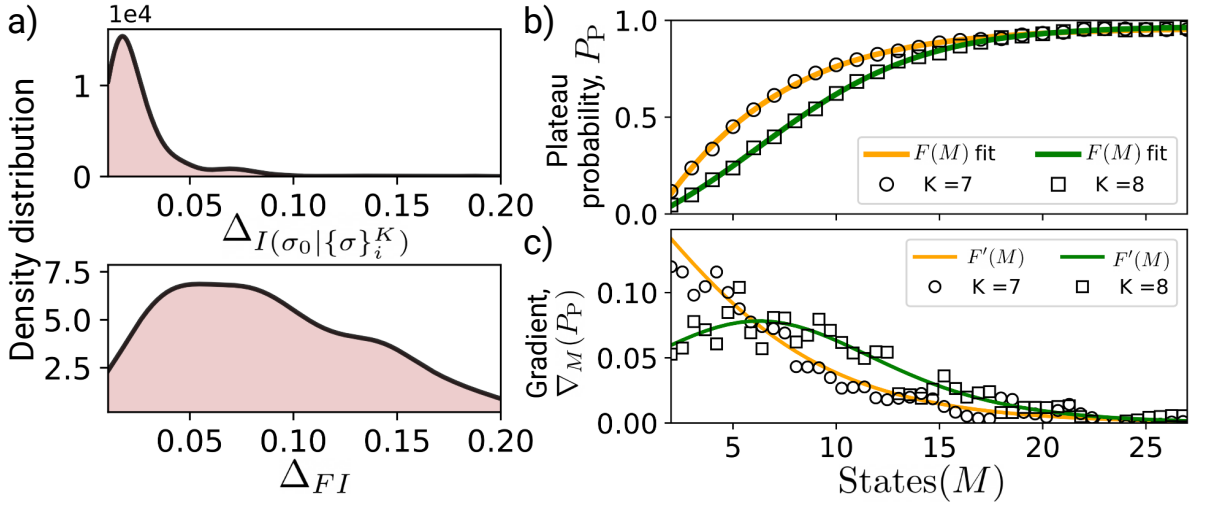

Figure 32 –

a) Error density distribution of (above) the mutual information $I(\sigma_0; \{\sigma\}_i^K)$ and (below) the fractional information; over thousands of groups of size $K = 8$. b) Plateau probability in Eq. (6.2) as a function of the number of states M for $K = 7$ and $K = 8$; markers correspond to data and solid lines correspond to the fit to a hyperbolic tangential function $F(M)$. c) Gradient $\nabla_M(P_P)$ as a function of M; markers correspond to numeric estimation from data and solid lines correspond to the derivative of the fitted function. $F'(M)$.

$$MFI = \frac{I(\sigma_0; \{\tilde{\sigma}_i^K, \tilde{\sigma}_{i+1}^K\})}{I(\sigma_0; \{\sigma\}_i^{2K})} \sim 1 \tag{6.1}$$

Estimating reliably the mutual information $I(\sigma_0; \{\sigma\}_i^{2K})$ is not possible for large K. Nevertheless, we expect MFI to be a saturating function of the number of states M, meaning that we can use the gradient of MFI to estimate if with a small number of states M the system reaches a plateau. As we did with the fractional information, we can estimate the probability density $\rho\,(MFI)$, over all possible choices of σ_0, with a plateau interval $MFI \sim 1 \pm \Delta_{MFI}$. Figure **??**b shows the plateau probability,

$$P_P = \int_{1-\Delta_{MFI}}^{1+\Delta_{MFI}} \rho\,(MFI)\,, \tag{6.2}$$

as a function of the number of states M for $K = 7$ and $K = 8$ respectively. Markers correspond to the data while solid lines correspond to the fit to an hyperbolic tangential function. Figure **??**c shows the gradient $\nabla_M(P_P)$ which quickly decays for $M > 15$. Such decay as a function of M means that few bits of information are gained when including more states in the compressed sub-networks. Consequently, we expect the coarse-grained compression to work in the hippocampus, preserving most relevant bits of σ_0.

Part IV

7 Discussion and perspectives

The rapid evolution of experimental techniques has allowed the observation and better understanding of the brain in a few model organisms, such as zebrafish or mice [118]. Nevertheless, the extrapolation of these techniques to other non-model organisms remains challenging, leading to a narrow range of observations across species. Such a limitation might lead to a potentially biased understanding of the principles underlying neuronal function [118]. To overcome such a limitation without fully relying in experiments, we can develop theoretical frameworks valid across species. One such example is the efficient coding hypothesis which proposes that systems adapt to their environmental conditions, optimizing information processing at a minimal energetic cost. Such a theory is species-independent, providing a powerful framework to investigate functional and anatomical features in sensory systems.

In this direction, we proposed a model of neuronal population dynamics of outer retinal circuits, using zebrafish experimental observations as a baseline. This data-driven approach provided a deepest understanding of the role of inhibitory and excitatory feedback mechanisms in circuits with N number of cones and one type of horizontal-cell. For instance, we shown that the emergence of bistability in networks with strong excitatory feedback implies a disadvantage for chromatic coding that networks with dominant inhibitory feedback do not exhibit, providing support to the predominant inhibitory couplings observed in zebrafish. In addition, such a model allows a straightforward extrapolation to species with similar functional circuits. For instance, other vertebrate species with interneuronal circuit motifs might exhibit a similar bistability, predicting weak or negligible inter-cone couplings for color coding. Similarly, other circuit motifs with different number of interneurons and time scales are feasible to investigate with this dynamical formalism, providing a broad data-driven approach to study outer retinal networks of color coding across species.

As a generalization, we used this neuronal population model to investigate the combination of two independent strategies for color coding identified in different clades. In more detail, we contrasted the coding efficiency of circuit motifs with and without oil-droplets, identifying a trade-off between coding performance and color-space gamut. From these studies, we identified the importance of independent overlapping cone responses in circuits exhibiting efficient encoding. Such overlap is considerably reduced when including droplet organelles, decreasing the encoding performance while increasing the total color-space gamut. Unfortunately, data on the functional responses of outer retinal circuits in birds or other species with droplets are yet to be collected. Nevertheless, based in our results, we might expect to observe a circuit motif optimizing such a trade-off, yielding

both an efficient processing of environmental chromatic information and a large number of distinguishable colors. In a nutshell, these data-driven studies on the outer retinal layer illustrate how theoretical approaches serve as a generalization of our understanding of sensory systems and neuronal circuits from a few species.

Along the retinal pathway, neuronal responses go from being smooth to exhibit spikes, allowing a signal diascretization that simplifies considerably theoretical approaches at a large scale. In this work, we proposed a new compression approach to describe neuronal activity without imposing constraints on the neuronal interactions as is done in conventional models inspired in condensed matter systems, where neurons are treated as interacting binary particles. This approach allows the accurate description of experimental observations with a number of parameters that grows much slower than the maximum 2^N, leading to more feasible descriptions of neuronal activity for large scale systems with typically thousands of neurons. For instance, using this compression coarse-grained approach, we were able to study large groups of neurons from the retina while preserving an accurate description of the experimental data, outperforming pairwise models with the same number of parameters. As a generalization, we tested our approach in the mouse hippocampus, suggesting that such compressibility is a more general property of neuronal populations, allowing the investigation of different brain areas across species. Beyond neuronal networks, our approach also shows promising for systems where the mutual information is subextensive, which seems to be a feature from natural systems [104].

Bibliography

[1] KUFFLER, S. W.: *Discharge patterns and functional organization of mammalian retina.* Journal of neurophysiology, 16(1):37–68, 1953, ISSN 00223077. Citado na página 15.

[2] Derrington, A. M. and P. Lennie: *Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque.* The Journal of Physiology, 357(1):219–240, 1984, ISSN 14697793. Citado na página 15.

[3] Valois, Russell L D E and Herman C Morgan: *Psychophysical studies of monkey vision–ii. squirrel m o n k e y wavelength a n d.* Vision Research, 14:69–73, 1974. Citado na página 15.

[4] Valois, Russell L D E and Herman Morgan: *SPATIAL LUMINANCE CONTRAST SENSITIVITY i.* Vision research, 14:75–81, 1974. Citado na página 15.

[5] Valos, Russell L D E, Herman C Morgan, Martha C Polson, William R Mead, and Elaine M Hull: *Psychophysical Studies of M O N K E Y Vision–I .* Russell The Journal Of The Bertrand Russell Archives, 14(1966), 1974. Citado na página 15.

[6] Wald, George and Paul K Brown: *Rhodopsin.* 127. Citado na página 15.

[7] Masland, Richard H.: *The Neuronal Organization of the Retina.* Neuron, 76(2):266–280, 2012, ISSN 08966273. `http://dx.doi.org/10.1016/j.neuron.2012.10.002`. Citado na página 15.

[8] Seung, H. Sebastian and Uygar Sümbül: *Neuronal cell types and connectivity: Lessons from the retina.* Neuron, 83(6):1262–1272, 2014, ISSN 10974199. Citado na página 15.

[9] Schiller, P. H. and J. G. Malpeli: *Properties and tectal projections of monkey retinal ganglion cells.* Journal of Neurophysiology, 40(2):428–445, 1977, ISSN 00223077. Citado na página 15.

[10] Lee, Barry B., Joel Pokorny, Paul R. Martin, Arne Valbergt, and Vivianne C. Smith: *Luminance and chromatic modulation sensitivity of macaque ganglion cells and human observers.* Journal of the Optical Society of America A, 7(12):2223, 1990, ISSN 1084-7529. Citado na página 15.

[11] Nelson, Ralph and Helga Kolb: *ON and OFF pathways in the vertebrate retina and visual system.* The visual neurosciences, (September):260–278, 2003. Citado na página 15.

[12] Hartline, H.K.: *the Response of Single Optic Nerve Fibers*. American Journal of Physiology, 121:400–415, 1937. Citado na página 15.

[13] Krieger, Brenna, Mu Qiao, David L. Rousso, Joshua R. Sanes, and Markus Meister: *Four alpha ganglion cell types in mouse retina: Function, structure, and molecular signatures*. PLoS ONE, 12(7):1–21, 2017, ISSN 19326203. Citado na página 15.

[14] Parthasarathy, Nikhil, Eleanor Batty, William Falcon, Thomas Rutten, Mohit Rajpal, E. J. Chichilnisky, and Liam Paninski: *Neural networks for efficient Bayesian decoding of natural images from retinal neurons*. Advances in Neural Information Processing Systems, 2017-Decem(Nips):6435–6446, 2017, ISSN 10495258. Citado 2 vezes nas páginas 15 and 20.

[15] Zhou, Mingyi, John Bear, Paul A. Roberts, Filip K. Janiak, Julie Semmelhack, Takeshi Yoshimatsu, and Tom Baden: *Zebrafish Retinal Ganglion Cells Asymmetrically Encode Spectral and Temporal Information across Visual Space*. Current Biology, 30(15):2927–2942.e7, 2020, ISSN 18790445. `https://doi.org/10.1016/j.cub.2020.05.055`. Citado na página 15.

[16] Baden, Tom, Timm Schubert, Le Chang, Tao Wei, Mariana Zaichuk, Bernd Wissinger, and Thomas Euler: *A tale of two retinal domains: Near-Optimal sampling of achromatic contrasts in natural scenes through asymmetric photoreceptor distribution*. Neuron, 80(5):1206–1217, 2013, ISSN 08966273. `http://dx.doi.org/10.1016/j.neuron.2013.09.030`. Citado 3 vezes nas páginas 15, 20, and 21.

[17] Balasubramanian, Vijay and Michael J. Berry: *A test of metabolically efficient coding in the retina*. Network: Computation in Neural Systems, 13(4):531–552, 2002, ISSN 0954898X. Citado na página 15.

[18] Gjorgjieva, Julijana, Haim Sompolinsky, and Markus Meister: *Benefits of pathway splitting in sensory coding*. Journal of Neuroscience, 34(36):12127–12144, 2014, ISSN 15292401. Citado 2 vezes nas páginas 15 and 20.

[19] Ratliff, Charles P., Bart G. Borghuis, Yen Hong Kao, Peter Sterling, and Vijay Balasubramanian: *Retina is structured to process an excess of darkness in natural scenes*. Proceedings of the National Academy of Sciences of the United States of America, 107(40):17368–17373, 2010, ISSN 00278424. Citado 2 vezes nas páginas 15 and 20.

[20] Atick, Joseph J. and A. Norman Redlich: *Towards a Theory of Early Visual Processing*. Neural Computation, 2(3):308–320, 1990, ISSN 0899-7667. Citado 2 vezes nas páginas 15 and 20.

[21] Baden, Tom, Timm Schubert, Philipp Berens, and Thomas Euler: *The Functional Organization of Vertebrate Retinal Circuits for Vision*. 2018, ISBN 9780190264086. Nenhuma citação no texto.

[22] Szatko, Klaudia P, Maria M Korympidou, Yanli Ran, Philipp Berens, Deniz Dalkara, Timm Schubert, Thomas Euler, and Katrin Franke: *Neural circuits in the mouse retina support color vision in the upper visual field*. Nature communications, 11(1):1–14, 2020. Citado 3 vezes nas páginas 15, 20, and 21.

[23] Segev, Ronen, Joe Goodhouse, Jason Puchalla, and Michael J Berry: *Recording spikes from a large fraction of the ganglion cells in a retinal patch*. Nature Neuroscience, 7(10):1154–1161, 2004, ISSN 10976256. Citado na página 15.

[24] Jun, James J, Nicholas A Steinmetz, Joshua H Siegle, Daniel J Denman, Marius Bauza, Brian Barbarits, Albert K Lee, Costas A Anastassiou, Alexandru Andrei, Çağatay Aydin, Mladen Barbic, Timothy J Blanche, Vincent Bonin, João Couto, Barundeb Dutta, Sergey L Gratiy, Diego A Gutnisky, Michael Häusser, Bill Karsh, Peter Ledochowitsch, Carolina Mora Lopez, Catalin Mitelut, Silke Musa, Michael Okun, Marius Pachitariu, Jan Putzeys, P Dylan Rich, Cyrille Rossant, Wei Lung Sun, Karel Svoboda, Matteo Carandini, Kenneth D Harris, Christof Koch, John O'Keefe, and Timothy D Harris: *Fully integrated silicon probes for high-density recording of neural activity*. Nature, 551(7679):232–236, 2017, ISSN 14764687. `http://dx.doi.org/10.1038/nature24636`. Citado na página 15.

[25] Ahrens, Misha B., Michael B. Orger, Drew N. Robson, Jennifer M. Li, and Philipp J. Keller: *Whole-brain functional imaging at cellular resolution using light-sheet microscopy*. Nature Methods, 10(5):413–420, 2013, ISSN 15487091. Citado na página 15.

[26] Holy, T. E., M. J. Berry, F. Soo, K. Sadeghi, D. Amodei, N. Deshmukh, and O. Marre: *Mapping a Complete Neural Population in the Retina*. Journal of Neuroscience, 2012, ISSN 0270-6474. Citado na página 15.

[27] Ramirez, Luisa and Ronald Dickman: *Data-driven models of optimal chromatic coding in the outer retina*. bioRxiv, page 2022.02.07.479405, 2022. `http://biorxiv.org/content/early/2022/02/09/2022.02.07.479405.abstract`. Citado na página 18.

[28] Field, G D and E J Chichilnisky: *Information Processing in the Primate Retina : Circuitry and Coding*. 2007. Citado 2 vezes nas páginas 18 and 20.

[29] Demb, Jonathan B and Joshua H Singer: *Functional Circuitry of the Retina*. pages 263–291. Citado na página 18.

[30] Thoreson, Wallace B and Dennis M Dacey: *DIVERSE CELL TYPES , CIRCUITS , AND MECHANISMS FOR COLOR VISION IN THE VERTEBRATE RETINA ORIGINS OF SPECTRAL SENSITIVITY.* (498):1527–1573, 2022. Citado na página 18.

[31] Jiggins, Chris D: *Ecological Speciation in Mimetic Butterflies.* 58(6), 2008. Citado na página 18.

[32] Melin, A D, C Hiramatsu, N A Parr, Y Matsushita, S Kawamura, and L M Fedigan: *The Behavioral Ecology of Color Vision : Considering Fruit Conspicuity , Detection Distance and Dietary Importance.* pages 258–287, 2014. Citado na página 18.

[33] Kooi, Casper J Van Der, Doekele G Stavenga, Kentaro Arikawa, Gregor Belušič, and Almut Kelber: *Evolution of Insect Color Vision : From Spectral Sensitivity to Visual Ecology.* pages 435–463. Citado na página 18.

[34] Regan, B. C., C. Julliot, B. Simmen, F. Viénot, P. Charles-Dominique, and J. D. Mollon: *Fruits, foliage and the evolution of primate colour vision.* Philosophical Transactions of the Royal Society B: Biological Sciences, 356(1407):229–283, 2001, ISSN 09628436. Citado na página 18.

[35] Cronin, Thomas W, Sönke Johnsen, N Justin Marshall, and Eric J Warrant: *Visual Ecology.* Princeton University Press, 2014, ISBN 9781400853021. `https://doi.org/10.1515/9781400853021`. Citado 6 vezes nas páginas 18, 22, 30, 40, 44, and 48.

[36] Shevell, Steven K and Paul R Martin: *Color opponency: tutorial.* JOSA A, 34(7):1099–1108, 2017. Citado na página 18.

[37] Buchsbaum, Gershon and Allan Gottschalk: *Trichromacy, opponent colours coding and optimum colour information transmission in the retina.* Proceedings of the Royal society of London. Series B. Biological sciences, 220(1218):89–113, 1983. Citado 4 vezes nas páginas 18, 31, 32, and 44.

[38] De Valois, Russell L, Israel Abramov, and Gerald H Jacobs: *Analysis of response patterns of LGN cells.* JOSA, 56(7):966–977, 1966. Citado na página 18.

[39] Patterson, Sara S, Maureen Neitz, and Jay Neitz: *Reconciling color vision models with midget ganglion cell receptive fields.* Frontiers in neuroscience, 13:865, 2019. Citado na página 18.

[40] Yoshimatsu, Takeshi, Philipp Bartel, Cornelius Schröder, Filip K. Janiak, François St-Pierre, Philipp Berens, and Tom Baden: *Ancestral circuits for vertebrate color vision emerge at the first retinal synapse.* Science Advances, 7(42):6815–6828, oct 2021. Citado 10 vezes nas páginas 18, 21, 31, 35, 38, 40, 44, 46, 47, and 48.

[41] Packer, Orin S, Jan Verweij, Peter H Li, Julie L Schnapf, and Dennis M Dacey: *Blue-yellow opponency in primate S cone photoreceptors.* Journal of Neuroscience, 30(2):568–572, 2010. Citado na página 18.

[42] Barlow, Horace B and Others: *Possible principles underlying the transformation of sensory messages.* Sensory communication, 1(01), 1961. Citado na página 20.

[43] Shannon, Claude Elwood: *The mathematical theory of communication, by CE Shannon (and recent contributions to the mathematical theory of communication), W. Weaver.* University of illinois Press Champaign, IL, USA, 1949. Citado na página 20.

[44] Cover, Thomas M: *Elements of information theory.* John Wiley & Sons, 1999. Citado 2 vezes nas páginas 20 and 54.

[45] Atick, Joseph J and A Norman Redlich: *Predicting ganglion and simple cell receptive field organizations.* International Journal of Neural Systems, 1(04):305–315, 1991, ISSN 0129-0657. Citado na página 20.

[46] Shepard, Roger N.: *Perceptual-cognitive universals as reflections of the world.* Behavioral and Brain Sciences, 24(4):581–601, 2001, ISSN 0140525X. Citado na página 20.

[47] Shepard, Roger N: *Psychological Review.* Mind, XXV(3):415–b–416, 1916, ISSN 0026-4423. Citado na página 20.

[48] Simoncelli, Eero P and Bruno A Olshausen: *Natural Image Statistics And Neural Representation.* 2001. Citado na página 20.

[49] Kersten, Daniel: *Predictability and redundancy of natural images.* Journal of the Optical Society of America A, 4(12):2395, 1987, ISSN 1084-7529. Citado na página 20.

[50] Laughlin, Simon: *A simple coding procedure enhances a neuron's information capacity.* Zeitschrift fur Naturforschung - Section C Journal of Biosciences, 36(9-10):910–912, 1981, ISSN 18657125. Citado na página 20.

[51] Field, David J and Since Hubel: *Relations between the statistics of natural images and the.* America, 4(12):2379–2394, 1987. Citado na página 20.

[52] Olshausen, B A and D J Field: *Network : Computation in Neural Systems Natural image statistics and efficient coding.* Network: Computation in Neural Systems, 7(October):333–339, 1996. Citado na página 20.

[53] Bialek, William, Fred Rieke, Rob R. De Ruyter Van Steveninck, and David Warland: *Reading a neural code.* Science, 252(5014):1854–1857, 1991, ISSN 00368075. Citado na página 20.

[54] Barlow, Horace: *Redundancy reduction revisited.* Network: Computation in Neural Systems, 12(3):241–253, 2001, ISSN 0954898X. Citado na página 20.

[55] Gollisch, Tim and Markus Meister: *Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina.* Neuron, 65(2):150–164, 2010, ISSN 08966273. `http://dx.doi.org/10.1016/j.neuron.2009.12.009`. Citado na página 20.

[56] *Yoshimatsu, Takeshi et al. (2021), Two-photon calcium recordings of cones, Dryad, Dataset, https:\/\/doi.org\/10.5061\/dryad.pzgmsbcmk.* Citado na página 22.

[57] *Cornelius. (2021). berenslab/cone_colour_tuning: v0.2(v0.2). Zenodo. https:\/\/doi.org\/10.5281\/zenodo.5502804.* Citado 2 vezes nas páginas 22 and 44.

[58] *Data on opsin and in-vivo photoreceptor spectral responses were from \cite{Cornelius2021}, hyperspectral data of naturalistic aquatic images from \cite{YoshimatsuData}.* Citado 2 vezes nas páginas 22 and 32.

[59] Kelber, Almut and Daniel Osorio: *From spectral information to animal colour vision: experiments and concepts.* Proceedings of the Royal Society B: Biological Sciences, 277(1688):1617–1625, 2010. Citado na página 22.

[60] Burkhardt, Dwight A.: *Light adaptation and photopigment bleaching in cone photoreceptors in situ in the retina of the turtle.* Journal of Neuroscience, 14(3 I):1091–1105, 1994, ISSN 02706474. Citado na página 23.

[61] Gerstner, Wulfram, Werner M Kistler, Richard Naud, and Liam Paninski: *Neuronal dynamics: From single neurons to networks and models of cognition.* Cambridge University Press, 2014. Citado na página 25.

[62] Chapot, Camille A, Thomas Euler, and Timm Schubert: *How do horizontal cells 'talk'to cone photoreceptors? Different levels of complexity at the cone–horizontal cell synapse.* The Journal of physiology, 595(16):5495–5506, 2017. Citado na página 25.

[63] Zimmermann, Maxime J Y, Noora E Nevala, Takeshi Yoshimatsu, Daniel Osorio, Dan Eric Nilsson, Philipp Berens, and Tom Baden: *Zebrafish differentially process color across visual space to match natural scenes.* Current Biology, 28(13), 2018. Citado na página 31.

[64] Hurvich, Leo M and Dorothea Jameson: *Some quantitative aspects of an opponent-colors theory. II. Brightness, saturation, and hue in normal and dichromatic vision.* JOSA, 45(8):602–616, 1955. Citado na página 32.

[65] Chittka, Lars and Randolf Menzel: *The evolutionary adaptation of flower colours and the insect pollinators' colour vision.* Journal of Comparative Physiology A, 171(2):171–181, 1992. Citado 2 vezes nas páginas 36 and 39.

[66] Fasick, JEFFRY I and PHYLLIS R Robinson: *Spectral-tuning mechanisms of marine mammal rhodopsins and correlations with foraging depth.* Visual neuroscience, 17(5):781–788, 2000. Citado 2 vezes nas páginas 36 and 39.

[67] Baden, T. and D. Osorio: *The Retinal Basis of Vertebrate Color Vision.* Annural Reviews in Vision Science, 5:177–200, 2019, ISSN 23744650. Citado 3 vezes nas páginas 37, 40, and 44.

[68] Huang, Yanping and Rajesh P N Rao: *Predictive coding.* Wiley Interdisciplinary Reviews: Cognitive Science, 2(5):580–593, 2011. Citado na página 37.

[69] Belušič, Gregor, Marko Ilić, Andrej Meglič, and Primož Pirih: *Red-green opponency in the long visual fibre photoreceptors of brushfoot butterflies (Nymphalidae).* Proceedings of the Royal Society B, 288(1961):20211560, 2021. Citado na página 37.

[70] Raviola, Elio and Norton B Gilula: *Gap junctions between photoreceptor cells in the vertebrate retina.* Proceedings of the National Academy of Sciences, 70(6):1677–1681, 1973. Citado na página 38.

[71] DeVries, Steven H, Xiaofeng Qi, Robert Smith, Walter Makous, and Peter Sterling: *Electrical coupling between mammalian cones.* Current Biology, 12(22):1900–1907, 2002. Citado na página 38.

[72] Laughlin, Simon B: *Retinal function: coupling cones clarifies vision.* Current biology, 12(24):R833—-R834, 2002. Citado na página 38.

[73] Tsukamoto, Yoshihiko, Patricia Masarachia, Stanley J Schein, and Peter Sterling: *Gap junctions between the pedicles of macaque foveal cones.* Vision research, 32(10):1809–1815, 1992. Citado na página 38.

[74] Jackman, Skyler L, Norbert Babai, James J Chambers, Wallace B Thoreson, and Richard H Kramer: *A positive feedback synapse from retinal horizontal cells to cone photoreceptors.* PLoS biology, 9(5):e1001057, 2011. Citado na página 38.

[75] Toomey, Matthew B. and Joseph C. Corbo: *Evolution, development and function of vertebrate cone oil droplets.* Frontiers in Neural Circuits, 11(December):1–17, 2017, ISSN 16625110. Citado 2 vezes nas páginas 40 and 49.

[76] Vorobyev, Misha: *Coloured oil droplets enhance colour discrimination*. Proceedings of the Royal Society B: Biological Sciences, 270(1521):1255–1261, 2003, ISSN 14712970. Citado na página 40.

[77] Goldsmith, Timothy H., James S. Collins, and Sherry Licht: *The cone oil droplets of avian retinas*. Vision Research, 24(11):1661–1671, 1984, ISSN 00426989. Citado 2 vezes nas páginas 40 and 49.

[78] Seifert, M., T. Baden, and D. Osorio: *The retinal basis of vision in chicken*. Seminars in Cell and Developmental Biology, 106(March):106–115, 2020, ISSN 10963634. `https://doi.org/10.1016/j.semcdb.2020.03.011`. Citado na página 40.

[79] Vorobyev, M., D. Osorio, A. T.D. Bennett, N. J. Marshall, and I. C. Cuthill: *Tetrachromacy, oil droplets and bird plumage colours*. Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology, 183(5):621–633, 1998, ISSN 03407594. Citado na página 40.

[80] Baden, Tom: *Circuit mechanisms for colour vision in zebrafish*. Current Biology, 31(12):R807–R820, 2021, ISSN 18790445. `https://doi.org/10.1016/j.cub.2021.04.053`. Citado na página 40.

[81] Ramirez, Luisa and Ronald Dickman: *Data-driven models of optimal chromatic coding in the outer retina*. bioRxiv, page 2022.02.07.479405, jan 2022. `http://biorxiv.org/content/early/2022/02/09/2022.02.07.479405.abstract`. Citado 4 vezes nas páginas 40, 41, 44, and 46.

[82] Wilby, David and Nicholas W. Roberts: *Optical influence of oil droplets on cone photoreceptor sensitivity*. Journal of Experimental Biology, 220(11):1997–2004, 2017, ISSN 00220949. Citado 2 vezes nas páginas 40 and 49.

[83] Wilby, David, Matthew B. Toomey, Peter Olsson, Rikard Frederiksen, M. Carter Cornwall, Ruth Oulton, Almut Kelber, Joseph C. Corbo, and Nicholas W. Roberts: *Optics of cone photoreceptors in the chicken (Gallus gallus domesticus)*. Journal of the Royal Society Interface, 12(111), 2015, ISSN 17425662. Citado na página 40.

[84] Goldsmith, T. H. and B. K. Butler: *The roles of receptor noise and cone oil droplets in the photopic spectral sensitivity of the budgerigar, Melopsittacus undulatus*. Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 189(2):135–142, 2003, ISSN 03407594. Citado na página 40.

[85] Bowmaker, J. K. and G. R. Martin: *Visual pigments and colour vision in a nocturnal bird, Strix aluco (tawny owl)*. Vision Research, 18(9):1125–1130, 1978, ISSN 00426989. Citado 2 vezes nas páginas 40 and 49.

[86] Hart, Nathan S.: *Variations in cone photoreceptor abundance and the visual ecology of birds.* Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology, 187(9):685–697, 2001, ISSN 03407594. Citado 2 vezes nas páginas 40 and 49.

[87] Hart, Nathan S. and Misha Vorobyev: *Modelling oil droplet absorption spectra and spectral sensitivities of bird cone photoreceptors.* Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 191(4):381–392, 2005, ISSN 03407594. Citado 4 vezes nas páginas 40, 41, 42, and 49.

[88] Ives, Jeffrey T., Richard A. Normann, and Peter W. Barber: *Light intensification by cone oil droplets: electromagnetic considerations.* Journal of the Optical Society of America, 73(12):1725, 1983, ISSN 0030-3941. Citado 3 vezes nas páginas 40, 41, and 49.

[89] Hart, Nathan S., Helena J. Bailes, Misha Vorobyev, N. Justin Marshall, and Shaun P. Collin: *Visual ecology of the Australian lungfish (Neoceratodus forsteri).* BMC Ecology, 8:1–14, 2008, ISSN 14726785. Citado 2 vezes nas páginas 41 and 42.

[90] Nordén, Klara Katarina, Chad M. Eliason, and Mary Caswell Stoddard: *Evolution of brilliant iridescent feather nanostructures.* eLife, 10:1–26, 2021, ISSN 2050084X. Citado 2 vezes nas páginas 44 and 49.

[91] Foster, David H., Kinjiro Amano, Sérgio M.C. Nascimento, and Michael J. Foster: *Incidence of metamerism in natural scenes.* Final Program and Proceedings - IS and T/SID Color Imaging Conference, pages 36–39, 2005, ISSN 10831304. Citado 2 vezes nas páginas 47 and 48.

[92] Baden, Tom, Timm Schubert, Le Chang, Tao Wei, Mariana Zaichuk, Bernd Wissinger, and Thomas Euler: *A tale of two retinal domains: Near-Optimal sampling of achromatic contrasts in natural scenes through asymmetric photoreceptor distribution.* Neuron, 80(5):1206–1217, 2013, ISSN 08966273. `http://dx.doi.org/10.1016/j.neuron.2013.09.030`. Citado 2 vezes nas páginas 47 and 48.

[93] Arnold, Karin and Christa Neumeyer: *Wavelength discrimination in the turtle Pseudemys scripta elegans.* Vision Research, 27(9):1501–1511, 1987, ISSN 00426989. Citado na página 49.

[94] Bowmaker, J. K., L. A. Heath, S. E. Wilkie, and D. M. Hunt: *Visual pigments and oil droplets from six classes of photoreceptor in the retinas of birds.* Vision Research, 37(16):2183–2194, 1997, ISSN 00426989. Citado na página 49.

[95] Muntz, W R A: *Inert absorbing and reflecting pigments.* In *Photochemistry of vision*, pages 529–565. Springer, 1972. Citado na página 49.

[96] Olsson, Peter, David Wilby, and Almut Kelber: *Quantitative studies of animal colour constancy: Using the chicken as model*. Proceedings of the Royal Society B: Biological Sciences, 283(1830):1–8, 2016, ISSN 14712954. Citado na página 49.

[97] Stavenga, Doekele G. and Bodo D. Wilts: *Oil droplets of bird eyes: Microlenses acting as spectral filters*. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1636), 2014, ISSN 14712970. Citado na página 49.

[98] Tkačik, Gašper, Olivier Marre, Dario Amodei, Elad Schneidman, William Bialek, and Michael J. Berry: *Searching for Collective Behavior in a Large Network of Sensory Neurons*. PLoS Computational Biology, 2014, ISSN 1553734X. Citado 6 vezes nas páginas 51, 52, 57, 58, 59, and 70.

[99] Tkacik, Gasper, Elad Schneidman, Michael J Berry, and William Bialek: *Ising models for networks of real neurons*. (3):3–6, 2006. `http://arxiv.org/abs/q-bio/0611072`. Citado 2 vezes nas páginas 52 and 59.

[100] Schneidman, Elad, Michael J Berry II, Ronen Segev, and William Bialek: *Weak pairwise correlations imply strongly correlated network states in a neural population*. Nature, 440(7087):1007, 2006. Citado 2 vezes nas páginas 52 and 58.

[101] Bialek, William and Rama Ranganathan: *Rediscovering the power of pairwise interactions*. 2007. `http://arxiv.org/abs/0712.4397`. Citado na página 52.

[102] Weigt, Martin, Robert A. White, Hendrik Szurmant, James A. Hoch, and Terence Hwa: *Identification of direct residue contacts in protein-protein interaction by message passing*. Proceedings of the National Academy of Sciences of the United States of America, 106(1):67–72, 2009, ISSN 00278424. Citado na página 52.

[103] Meshulam, Leenoy, Jeffrey L. Gauthier, Carlos D. Brody, David W. Tank, and William Bialek: *Collective Behavior of Place and Non-place Neurons in the Hippocampal Network*. Neuron, 96(5):1178–1191.e4, 2017, ISSN 10974199. `https://doi.org/10.1016/j.neuron.2017.10.027`. Citado 6 vezes nas páginas 52, 57, 58, 70, 72, and 73.

[104] Bialek, William, Stephanie E. Palmer, and David J. Schwab: *What makes it possible to learn probability distributions in the natural world?* (4):1–5, 2020. `http://arxiv.org/abs/2008.12279`. Citado 4 vezes nas páginas 52, 53, 70, and 82.

[105] Goldenfeld, Nigel: *Lectures on phase transitions and the renormalization group*. CRC Press, ISBN 0-429-49349-5. Citado na página 53.

[106] Tishby, Naftali, Fernando C Pereira, and William Bialek: *The information bottleneck method*. pages 1–16, 2000. `http://arxiv.org/abs/physics/0004057`. Citado 2 vezes nas páginas 55 and 101.

[107] Slonim, Noam and Naftali Tishby: *Agglomerative information bottleneck*. In *Advances in Neural Information Processing Systems*, 2000, ISBN 0262194503. Citado na página 55.

[108] Tkacik, Gasper: *Information flow in biological networks*. Princeton University, 2007, ISBN 0549136525. Citado na página 59.

[109] Cardy, John: *Scaling and renormalization in statistical physics*, volume 5. Cambridge university press, 1996, ISBN 0521499593. Citado na página 68.

[110] Kadanoff, Leo P: *Scaling laws for Ising models near T c*. Physics Physique Fizika, 2(6):263, 1966. Citado na página 68.

[111] Strong, S. P., Roland Koberle, Rob R. De Ruyter Van Steveninck, and William Bialek: *Entropy and information in neural spike trains*. Physical Review Letters, 80(1):197–200, 1998, ISSN 10797114. Citado na página 69.

[112] Tkačik, Gašper, Thierry Mora, Olivier Marre, Dario Amodei, Stephanie E Palmer, Michael J Berry, and William Bialek: *Thermodynamics and signatures of criticality in a network of neurons*. Proceedings of the National Academy of Sciences, 112(37):11508–11513, 2015. Citado na página 70.

[113] Meshulam, Leenoy, Jeffrey L. Gauthier, Carlos D. Brody, David W. Tank, and William Bialek: *Coarse graining, fixed points, and scaling in a large population of neurons*. Physical Review Letters, 2019, ISSN 10797114. Citado na página 70.

[114] Humplik, Jan and Gašper Tkačik: *Probabilistic models for neural populations that naturally capture global coupling and criticality*. PLoS Computational Biology, 13(9):1–26, 2017, ISSN 15537358. Citado na página 70.

[115] Ganmor, Elad, Ronen Segev, and Elad Schneidman: *Sparse low-order interaction network underlies a highly correlated and learnable neural population code*. Proceedings of the National Academy of Sciences of the United States of America, 108(23):9679–9684, 2011, ISSN 00278424. Citado na página 70.

[116] Eggermont, Jos J: *Is there a neural code?* 22(2):355–370, ISSN 0149-7634. Publisher: Elsevier. Citado na página 71.

[117] Moser, Edvard I, Emilio Kropff, and May Britt Moser: *Place cells, grid cells, and the brain's spatial representation system*. 31(1):69–89. Citado na página 72.

[118] Baden, Tom: *Vertebrate vision: lessons from non-model species*. In *Seminars in cell & developmental biology*, volume 106, pages 1–4, 2020, ISBN 1096-3634. Citado na página 81.

[119] Chen, Xiaohong, Victor Chernozhukov, Sokbae Lee, and Whitney K Newey: *Local identification of nonparametric and semiparametric models*. Econometrica, 82(2):785–809, 2014, ISSN 0012-9682. Citado na página 96.

[120] Gramacki, Artur: *Nonparametric kernel density estimation and its computational aspects*, volume 37. Springer, 2018. Citado na página 96.

[121] Purves, Dale, Roberto Cabeza, Scott A Huettel, Kevin S LaBar, Michael L Platt, Marty G Woldorff, and Elizabeth M Brannon: *Cognitive neuroscience*, volume 6. Sunderland: Sinauer Associates, Inc, 2008. Citado na página 100.

Appendix

APPENDIX A – Kernel density estimation (KDE)

The kernel density estimation is a smooth method to estimate the probability distribution of a given data-set X; we use them as a smooth version of the histogram. Instead of producing a discrete distribution, the KDE provides a continuous estimation of the data distribution; for our purposes, we restrict our discussion to Gaussian KDE. The general idea consist in assigning a kernel distribution, $K(x) = (x - X_i)$, to each sample X_i from the data-set. For a Gaussian Kernel that is,

$$K(x) \equiv K_h(x - X_i) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{(x - X_i)^2}{2h^2}\right), \tag{A.1}$$

where h is the standard deviation of the Gaussian kernel, also known as the bandwidth parameter of the KDE; this is a free parameter whose value is optimized for each data-set (see below). The estimated distribution is calculated as,

$$\begin{aligned} f(x,h) &= \frac{1}{n}\sum_{i}^{n} K_h(x - X_i) \\ &= \frac{1}{n}\sum_{i}^{n} \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{(x - X_i)^2}{2h^2}\right) \end{aligned} \tag{A.2}$$

with n the total number of samples. As mentioned above, the bandwidth is a free parameter adapted to the specific data-set. The larger the parameter the naiver the final distribution and the smaller the parameter the more complex the distribution; both extremes produce bad estimations. There are different methods to calculate an optimal bandwidth based on the specific data-set; here, we restrict our discussion to the rule-of-thumb method, which is the one we use to develop our work. The main assumption of this method is to consider that the unknown probability distribution belongs to the family of Gaussian distributions (see Ref. [119]). Under this assumption, the *optimal* bandwidth value is,

$$h = \frac{1.06\tilde{\sigma}}{\sqrt{n}} \tag{A.3}$$

where $\tilde{\sigma}$ is the standard deviation of the data. This method is the more straightforward and computationally cheap. Nevertheless, depending on the data-set, this method could yields a poor bandwidth estimation. More details are provided in Ref. [120].

When data is accompanied of error bars, we cannot consider all data equal anymore. Data with small error bars should get a strongest contribution to the KDE than data with

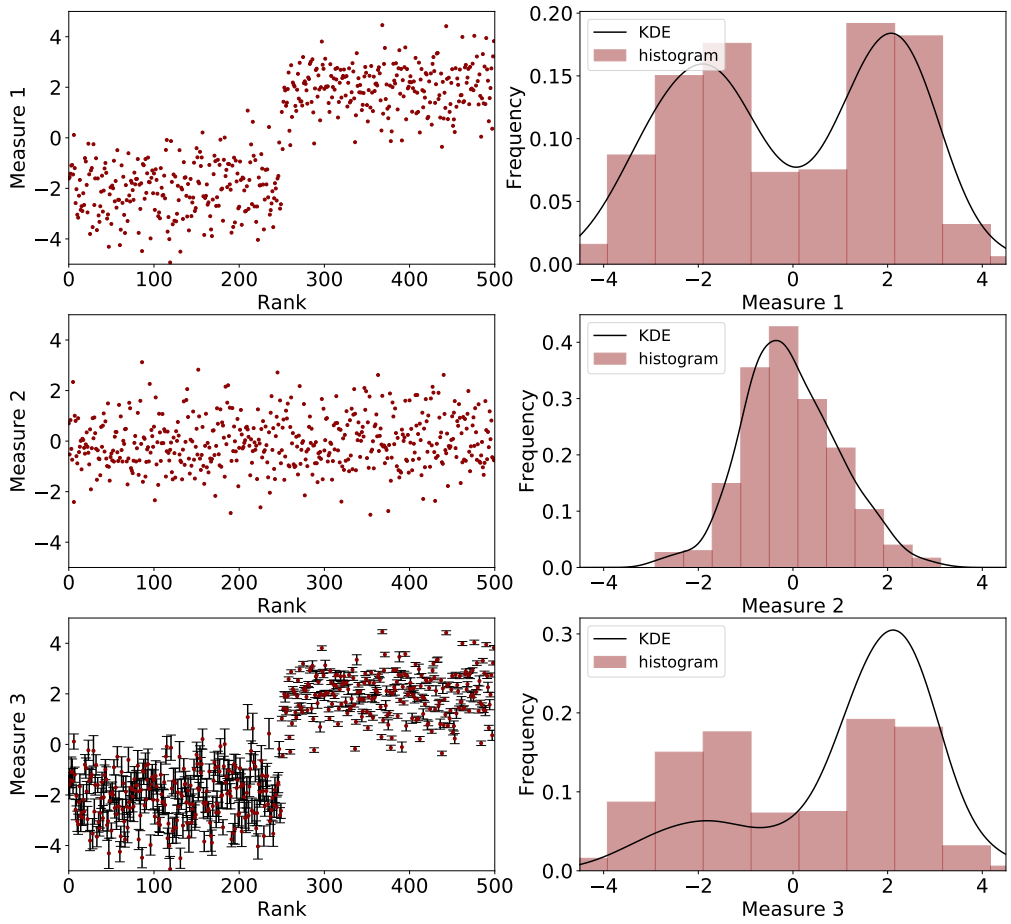

Figure 33 – Kernel distribution estimation (KDE) of C and FI for the whole population of neurons in the hippocampus data-set. The intensity plots show the KDE per M in the case of FI and for each number of parameters in the case of C. Both figures correspond to the case of the first 8 neurons sharing the most information with each independent σ_0.

large error bars (see the example of Fig. 33 c). The weighted-KDE considers a different optimal bandwidth for each datum depending on its error bar. A full implementation is available in python, by scipy.

On the left panel of Fig. 33, we show three examples of data-sets, generated from normal distributions. The first data-set (measure 1) is composed of two data clusters drawn from the distributions $\mathcal{N}(-\in, \infty)$ and $\mathcal{N}(\in, \infty)$. The second data-set (measure 2) is drawn from the distribution $\mathcal{N}(\prime, \infty)$. The third data-set (measure 3) is equal to the first data-set, but with error bars included – data drawn from the distribution $\mathcal{N}(-\in, \infty)$

have larger error bars than those ones of the data drawn from $\mathcal{N}(\in, \infty)$. On the right side of each data plot, we observe the superposition of both the histogram and KDE associated to the data-set. Note that in the first two cases, the KDE is a smooth approximation of the histogram. In the last case, we used a weighted-KDE to get a better estimation based on the uncertainties of data. Note that the larger the error-bar of a datum, the smaller its contribution.

The KDEs used to develop our work were compared with the corresponding histograms to avoid large statistical errors. We used the function *stats.gaussian_ KDE* from the scipy library in python. This library uses the kernel of Eq. A.1 and calculated the optimal bandwidth with the rule-of-thumb.

APPENDIX B – Neurons

In this appendix, we discuss the basic biology of neurons. In the first section, we describe general anatomical and functional aspects. Following this, we describe the interaction mechanisms among neurons, known as synapses.

B.1 Basic concepts

Neurons are the unit cells of the nervous system, ramifying all over the animal's body. The anatomy of these cells vary across regions, but most of them have three principal parts: soma, axon and dendrites (see Fig. 34a). Typically, the dendrites feed the neuron with external stimulation from other neurons. The soma contains the nuclei and other organelles of the neuron, and it is in charge of integrating the inputs from the dendrites to generate a unique response. Finally, the axon send the responses to other neurons via synapses.

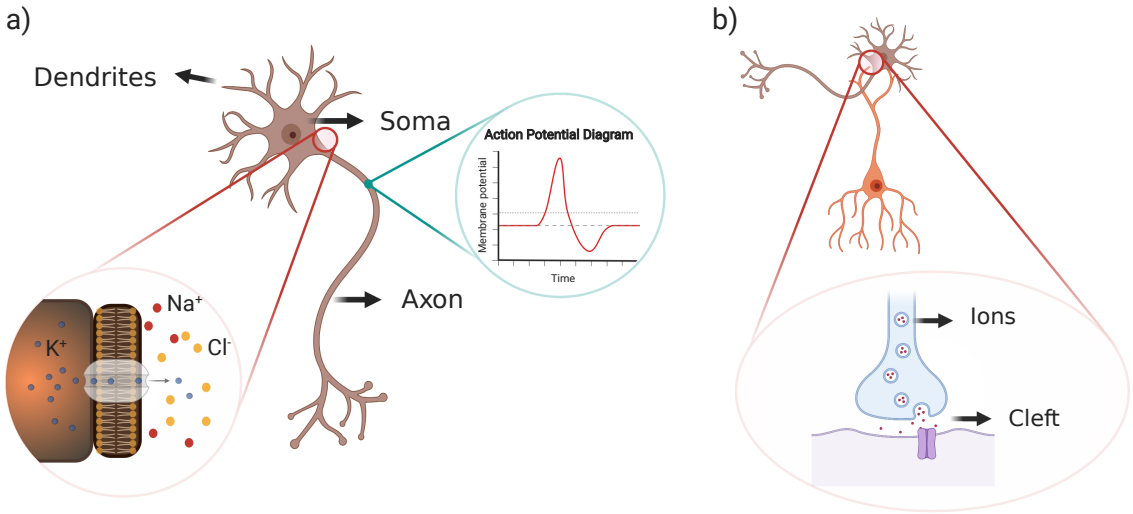

Figure 34 –

Typical neurons and synapses. a) Sketch of a typical neuron. The inset, in red, shows both the intracellular and extracellular ionic concentrations. Ions move across the membrane using specific ion-channels. The inset, in blue, shows a typical action potential of a spike neuron. b) Sketch of a dendrite-soma synapse. The activation of the pre-synaptic neuron generates an ionic flux to be released into the cleft. The post-synaptic neuron has ion-specific channels opening when the ionic concentration at the cleft increases. Created with BioRender

All neurons have a celular membrane allowing the flux of ions in specific sites called ion-channels. In general, the ions with the larger concentration around the membrane are

Potassium K^+, Calcium Ca^- and Sodium Na^- (see Fig. 34a). The charge difference on both the inside and outside of the membrane defines the membrane potential which at rest -when there is no external stimulation- equals the Nernst potential [121].

The external stimulation of a neuron produces an ionic flux causing a depolarization of membrane potential. If the depolarization is continuous in time, the neurons are non-spiking. On the other hand, when the depolarization suddenly increases after a threshold[1], we define the neurons as spiking. The inset in figure 34a shows the typical behavior of the membrane potential of spiking neurons. Notice that after reaching the threshold (solid grey line), the membrane potential increases up to a maximum and starts decreasing. This phenomenon, known as action potential or spike, is triggered by an external stimulation, but once the threshold is reached, the behavior becomes independent of external stimuli. After reaching the maximum, the neurons initiates a process of hyperpolarization known as refractory period; the typical duration of this period is approximately $2ms$.

B.1.1 Synapses

So far, we described the general anatomy and dynamics of neurons. Now, we will describe how these neurons communicate among them[2]. As previously discussed, a neuron (post-synaptic) activates when receiving stimuli from other neurons (pre-synaptic) via ionic fluxes. The place at which this stimulation happens is known as synapse. There are diverse types of synapses in the nervous systems. For practical purposes, we will discuss only chemical synapses, where neurons do not interact physically. Figure 34b shows a sketch of a typical synapse between a neuron's dendrite and a neuron's soma[3].

When a pre-synaptic neuron spikes, the ionic flux produced flows throughout the axon. The synapses are in charge of releasing this ionic concentration into specific places outside the neuron called clefts. When the ionic concentration at the cleft changes, the ion channels of the post-synaptic cell open allowing the entrance of the ionic flux. Synapses are plastic, meaning that neurons can create and delete synapses changing the neuronal connectivity; it is the basis of learning.

[1] The membrane potential threshold varies across neurons, but it is fixed for each neuron.
[2] For practical purposes, we restrict our discussion only to spiking neurons.
[3] In general, synapses can be created in any place over the neuron. The ionic flux always goes from the pre-synaptic neuron to the post-synaptic neuron.

APPENDIX C – Solution IBM

As discussed in the main text, the goal of the information bottleneck method is to find an optimal mapping of $\sigma \to \tilde{\sigma}$ with the constraint on the mutual information $I(\sigma_0; \tilde{\sigma})$. This problem is defined by the Markov chain $\sigma_0 \leftarrow \sigma \leftarrow \tilde{\sigma}$, such that the optimal mapping is determined though the conditional probability $p(\tilde{\sigma}|\sigma)$ [106]. That is,

$$\mathcal{L}[p(\tilde{\sigma}|\sigma)] = I(\tilde{\sigma};\sigma) - \beta I(\tilde{\sigma};\sigma_{\mathbf{0}}) - \sum_{\sigma,\tilde{\sigma}} \lambda(\sigma) p(\tilde{\sigma}|\sigma), \tag{C.1}$$

where we have included the normalization of the conditional probability $p(\tilde{\sigma}|\sigma)$. Note that this solution depends on the specific mapping $\sigma \to \tilde{\sigma}$. We minimize Eq. (C.1) such that we get

$$0 = \frac{\partial}{\partial p(\tilde{\sigma}|\sigma)} \left[\sum_{\sigma,\tilde{\sigma}} p(\sigma,\tilde{\sigma}) \log_2 \frac{p(\sigma,\tilde{\sigma})}{p(\sigma)p(\tilde{\sigma})} - \beta \sum_{\tilde{\sigma},\sigma_0} p(\sigma_0,\tilde{\sigma}) \log_2 \frac{p(\sigma_0,\tilde{\sigma})}{p(\sigma_0)p(\tilde{\sigma})} - \sum_{\sigma,\tilde{\sigma}} \lambda(\sigma) p(\tilde{\sigma}|\sigma) \right]. \tag{C.2}$$

For the first term on the right hand side (r.h.s.), we get

$$\begin{aligned} \frac{\partial}{\partial p(\tilde{\sigma}|\sigma)} \left[\sum_{\sigma,\tilde{\sigma}} p(\sigma,\tilde{\sigma}) \log_2 \frac{p(\sigma,\tilde{\sigma})}{p(\sigma)p(\tilde{\sigma})} \right] &= p(x) \log_2 \left(\frac{p(\tilde{\sigma}|\sigma)}{p(\tilde{\sigma})} \right) \\ &\quad + p(\sigma)p(\tilde{\sigma}) \left(\frac{1}{p(\tilde{\sigma})} - \frac{p(\tilde{\sigma}|\sigma)}{p(\tilde{\sigma})^2} \frac{\partial p(\tilde{\sigma})}{\partial p(\tilde{\sigma}|\sigma)} \right) \\ &= p(x) \log_2 \left(\frac{p(\tilde{\sigma}|\sigma)}{p(\tilde{\sigma})} \right). \end{aligned} \tag{C.3}$$

For the second term on the r.h.s. we get,

$$\begin{aligned} \frac{\partial}{\partial p(\tilde{\sigma}|\sigma)} \left[\beta \sum_{\tilde{\sigma},\sigma_0} p(\sigma_0,\tilde{\sigma}) \log_2 \frac{p(\sigma_0,\tilde{\sigma})}{p(\sigma_0)p(\tilde{\sigma})} \right] &= \beta \left[\sum_{\sigma_0} \log_2 \frac{p(\tilde{\sigma}|\sigma_0)}{p(\tilde{\sigma})} \frac{\partial p(\tilde{\sigma}|\sigma_0)}{\partial p(\tilde{\sigma}|\sigma)} \right. \\ &\quad \left. \sum_{\sigma_0} p(\sigma_0)p(\tilde{\sigma}) \frac{\partial}{\partial p(\tilde{\sigma}|\sigma)} \left(\frac{p(\tilde{\sigma}|\sigma_0)}{p(\tilde{\sigma})} \right) \right] \\ &= \beta \left[\sum_{\sigma_0} p(\sigma_0|\sigma)p(\sigma) \log_2 \frac{p(\sigma_0|\tilde{\sigma})}{p(\sigma_0)} \right]. \end{aligned} \tag{C.4}$$

For the third term on the r.h.s. we get,

$$\frac{\partial}{\partial p(\tilde{\sigma}|\sigma)} \sum_{\sigma,\tilde{\sigma}} \lambda(\sigma) p(\tilde{\sigma}|\sigma) = \lambda(\sigma) \tag{C.5}$$

Replacing Eqs. (C.3)-(C.5) into Eq. (C.2) we get,

$$0 = p(\sigma) \left[\log_2 \frac{p(\tilde{\sigma}|\sigma)}{p(\tilde{\sigma})} - \beta \sum_{\sigma_0} p(\sigma_0|\sigma) \log_2 \frac{p(\sigma_0|\tilde{\sigma})}{p(\sigma_0)} - \frac{\lambda(\sigma)}{p(\sigma)} \right]. \tag{C.6}$$

that leads to the solution

$$p(\tilde{\sigma}|\sigma) = \frac{p(\tilde{\sigma})}{Z(\sigma,\beta)} \exp\left(-\beta D_{KL}\left[p(\sigma_0|\sigma)||p(\sigma_0|\tilde{\sigma})\right]\right), \tag{C.7}$$

with

$$Z(\sigma,\beta) = \exp[\beta\tilde{\lambda}(x)] = \sum_{\tilde{\sigma}} p(\tilde{\sigma}) \exp\left(-\beta D_{KL}\left[p(\sigma_0|\sigma)||p(\sigma_0|\tilde{\sigma})\right]\right) \tag{C.8}$$

the partition function.

This problem follows the Markov chain $\sigma_0 \leftarrow \sigma \leftarrow \tilde{\sigma}$, consequently we have,

$$p(\sigma_0|\tilde{\sigma}) = \sum_{\sigma} p(\sigma_0|\sigma) p(\sigma|\tilde{\sigma}). \tag{C.9}$$

Similarly, using baye's rule,

$$p(\tilde{\sigma}) = \sum_{x} p(\tilde{\sigma}|\sigma) p(\sigma). \tag{C.10}$$

This solution leads to an iterative solution in which the probabilities $p(\tilde{\sigma}|\sigma)$, $p(\sigma_0|\tilde{\sigma})$ and $p(\tilde{\sigma})$ are determined such that they simultaneously minimize the free energy functional, leading to the iterative solution,

$$\begin{cases} p_t(\tilde{\sigma}|\sigma) = \frac{p_t(\tilde{\sigma})}{Z_t(\beta,\sigma)} \exp(-\beta D_{KL}\left[p(\sigma_0|\sigma)||p(\sigma_0|\tilde{\sigma})\right]) \\ p_{t+1}(\sigma_0|\tilde{\sigma}) = \frac{1}{p(\tilde{\sigma})} \sum_{\sigma_0} p(\sigma_0,\sigma) p(\tilde{\sigma}|\sigma) \\ p_{t+1}(\tilde{\sigma}) = \sum_{\sigma} p(\sigma) p_t(\tilde{\sigma}|\sigma). \end{cases},$$

In the limit of $\beta \to \infty$, the Kullback–Leibler divergence $D_{KL}\left[p(\sigma_0|\sigma)||p(\sigma_0|\tilde{\sigma})\right]$ is minimized, and the solution reduces to

$$\begin{cases} p(\tilde{\sigma}|\sigma) = \begin{cases} 1 & \text{if } \sigma \in \tilde{\sigma} \\ 0 & \text{otherwise} \end{cases} & \forall \sigma \in \boldsymbol{\sigma} \\ p(\sigma_0|\tilde{\sigma}) = \frac{1}{p(\tilde{\sigma})} \sum_{i=1}^{|\tilde{\sigma}|} p(\sigma_i,\sigma_0) & \forall \sigma_0 \in \boldsymbol{\sigma_0} \\ p(\tilde{\sigma}) = \sum_{i}^{|\tilde{\sigma}|} p(\sigma_i). \end{cases}. \tag{C.11}$$

APPENDIX D – Error estimation

In this appendix we describe the finite-size approach that we used to determine the error bars of the second part of this thesis.

D.1 Fractional information

Estimating the mutual information between a single neuron, σ_0, and a group of K neurons, $\boldsymbol{\sigma}_i$, means inferring first all the 2^{K+1} state probabilities. As previously discussed, choosing $K = 8$ provide us with a good statistics to reliably infer such probabilities. Figure 20 shows the probability density of the ratio of observed states in data for all choices of σ_0, indicating that in the retinal population, we have statistics of more than 90% of the possible states.

Estimating the mutual information directly from data leads to finite-size sampling errors that must be corrected. To do so, we implement an extrapolation method to estimate the truth value. We calculate the mutual information as a function of the inverse of the sample size, by using fractions of the data, $f = (50, 60, 70, 80, 90)\%$. We use a linear fit to estimate the intercept value which is equivalent to the extrapolated mutual information. We repeat this procedure a hundred times from random permutations of the data and take the mean value as the truth mutual information. To calculate the error of the extrapolated mutual information, we calculate the mutual information for random halves of the data and calculate the standard deviation.

To calculate the fractional mutual information, FI, we propagate the error of $I_{\max}(M)$ divided into $I(\sigma_0, \boldsymbol{\sigma})$. Doing a proper compression means reaching the plateau $FI = 1$ within error bars Δ_{FI}, coming from finite-size corrections. Note that each choice of σ_0 leads to a different value of mutual information, and consequently, to a different error estimator – figure 25b shows the Δ_{FI} distribution over all choices of σ_0.

D.2 Coding cost

Calculating the coding cost as a function of the number of parameters requires performing several least-squares fits of the effective field for different truncations of Eq. 5.7. To calculate the error bars, we perform this fit using 40% of the data, and with the inferred parameters, we estimate C in the other 60% of the data. We repeat this procedure a hundred times to get an estimation of the standard deviation that we use as an error estimator. We use all the data to estimate the mean value of the coding cost, ensuring it

is always within the estimated error bars – figure 24a shows the error bars in red and the main values in black.

D.3 Error bar estimation of the extrapolated mutual information for a system following Zipf law

To get a better understanding about the correct way to estimate the error bars in our data, we study a simple case in which we can estimate analytically the mutual information. Lets assume the same problem of having one neuron σ_0 and a group of K neurons, $\boldsymbol{\sigma}_i$, with the mutual information $I(\sigma_0, \boldsymbol{\sigma}_i)$. Assuming that the probability mass function of the 2^K states in $\boldsymbol{\sigma}_i$ follows a Zipf law, we get,

$$p(\boldsymbol{\sigma}_i) = \frac{1/k}{\sum_k 1/k}, \tag{D.1}$$

where k represent the rank of the state. In addition, assigning to each of these states k an effective field h_k acting on the single neuron σ_0, we get,

$$p(\sigma_0|\boldsymbol{\sigma}_i) = \frac{\exp(\sigma_0 h_k)}{1+\exp(h_k)}. \tag{D.2}$$

We draw the effective fields, h_k, from a normal distribution and we use Bayes theorem to estimate the mutual information, $I(\sigma_0;\boldsymbol{\sigma}_i) = 0.1559$ bits[1]. Assuming now that we do not have access to the probability distribution, but instead to a finite number of samples, N, we expect to get an estimation error proportional to $1/\sqrt{N}$. Considering $N = 10000$, we expect an error of ≈ 0.01 bits.

Applying the extrapolation method discussed in the previous subsections, we can estimate the mutual information in the limit $N \to \infty$ by iterating fractions of the data (see Fig. 35). Using a linear fit to calculate the intercept values, we estimate the mutual information to be ≈ 0.1546 bits – we also used a quadratic fit without significantly changes. Comparing the effective-size estimation with the truth mutual information give us a relative error of $\approx 1\%$, which agrees with the expected error.

So far, we can conclude that the linear extrapolation method provides a good estimation of the mutual information, within the theoretical expected error. Estimating the error magnitude directly from data can be achieved by calculating several times the mutual information on random halves of data and taking the standard deviation. Figure 35 shows that the magnitude of the expected error bar (red line) is the same as the magnitude of the estimated error bar (green line).

[1] Calculations were made in python using the numpy random generator with the initialized seed zero

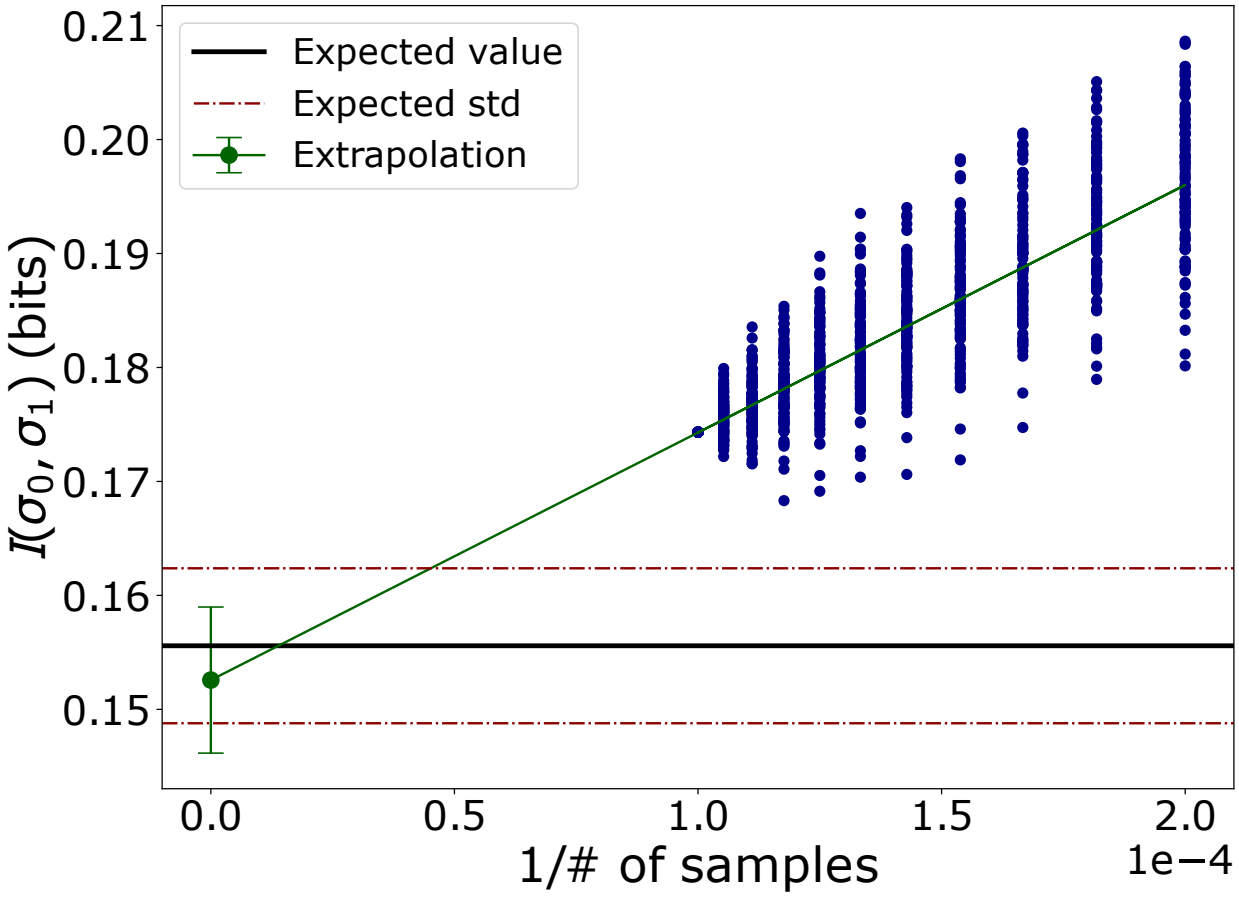

Figure 35 –

Estimation of the mutual information $I(\sigma_0, \{\sigma_j\})$ for the Zipf system via linear extrapolation. (black) Expected mutual information (red) Expected error from finite-size sampling. (blue) Mutual information estimation of one hundred iterations from fractions of the sample as a function of the fraction size. (green) Linear extrapolation fit of all the iterations; the intercept value correspond to the extrapolated mutual information and, the error bar corresponds to the standard deviation from random halves of the data.

Despite we know that neuronal populations do not follow an exact Zipf law, we know that there is a strong dependence between both the state frequency and the rank. Also, when observing the extrapolation curves of real data, we observe a similar linear trend as the one observed in the Zipf example. Consequently, we apply this linear extrapolation method to estimate the mutual information between groups of neurons, and we estimate the magnitude of the error by calculating several times the mutual information from random halves of the data.

APPENDIX E – Transmittance model

As shown in Fig. 36a, carotenoides inside an oil droplet absorb a portion of the incident light, I_0, transmitting I_1. Such a process is described by the transmittance defined as,

$$T = \frac{I_1}{I_0}, \tag{E.1}$$

that is related to the material absorbance as,

$$A = -\log_{10} T. \tag{E.2}$$

Since carotenoids are the molecules responsible for light absorption in the droplet, the transmittance depends on the molecular density inside the droplet. Assuming that carotenoids are homogeneously distributed in the droplet, we can use the Beer-Lambert law to describe light absorbance, that is,

$$A = \alpha(\lambda)\, l\, c, \tag{E.3}$$

where l is the mean light path, c is the carotenoid concentration and $\alpha(\lambda)$ is the absorption coefficient. Following, the logarithmic absorbance has an exponential decay for wavelengths far from the absorption peak, that is,

$$\log A = -a\lambda + b. \tag{E.4}$$

With equations (E.2), (E.3) and (E.4) we get,

$$T = \exp(-lc\beta \exp(-a(\lambda - \lambda_s))) \tag{E.5}$$

with $\alpha(\lambda) = \beta \exp(-a(\lambda - \lambda_s)$ and λ_s the minimum wavelength at which the linear relation of Eq. (E.4) holds. That is, $\alpha(\lambda_s)$ is the maximum droplet absorption. For wavelengths far from the absorption peak, the transmittance must be large, meaning that $lc\alpha(\lambda_s) >> 1$. The case of dense carotenoids is the one satisfying $c >> 1/(l\alpha(\lambda_s))$.

Experiments on droplets provide information of the wavelength parameters λ_{cut} and λ_{mid}, which are related to the transmittance function as shown in Fig. 36B. The parameter λ_{mid} corresponds to the wavelength at which the transmittance has half of its maximum value, while λ_{cut} is related to the wavelength at which the linear extrapolation from the mid point is zero. We rewrite Eq. (E.5) as,

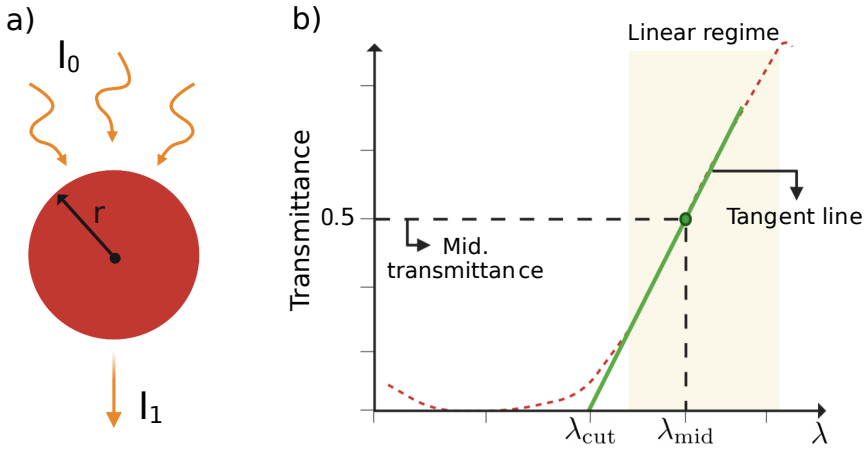

Figure 36 –

a) Sketch of a droplet of radius r with incident and transmitted light, I_0 and I_1 respectively. b) Sketch of a typical transmittance function (red dashed line). The green line shows the tangent line corresponding to the point (λ_{mid}), with λ_{mid} the wavelength at which the transmittance has half of its maximum value (black dashed line). The parameter λ_{cut} is defined as the wavelength at which the tangent line cuts the λ-axis.

$$T = \exp(-\exp(-a(\lambda - \lambda_0))) \tag{E.6}$$

with $\lambda_0 = \lambda_s + 1/a \log(lc\beta)$. From Eq. (E.6), we calculate the mid wavelength parameter to be,

$$\lambda_{\text{mid}} = -\frac{\log(\log(2))}{a} + \lambda_0 \tag{E.7}$$

$$\approx \frac{0.37}{a} + \lambda_0 \tag{E.8}$$

We calculate the parameter λ_{cut} using the tangent line at $T(\lambda_{\text{mid}})$. We calculate the derivative of the function in Eq. (E.6) to estimate the slope of the tangent line, that is,

$$m = T'(\lambda_{\text{mid}}) = \frac{\log 2}{2} a \tag{E.9}$$

$$\approx 0.35a. \tag{E.10}$$

Using this slope m and the line point $(\lambda_{\text{mid}}, 1/2)$, we get the line equation,

$$T = \frac{a \log 2}{2}\lambda - \frac{a \log 2}{2}\lambda_0 + \frac{\log 2}{2}\log(\log 2) + \frac{1}{2} \tag{E.11}$$

$$\approx 0.35\, a\, \lambda - 0.35\, a\, \lambda_0 + 0.37. \tag{E.12}$$

From this equation, we calculate the cut wavelength,

$$\lambda_{\text{cut}} \approx \lambda_0 - \frac{1.06}{a}. \tag{E.13}$$

With Equations (E.8) and (E.13) we write the parameters in Eq. E.6 as

$$\begin{aligned} a &\approx \frac{1.43}{\lambda_{\text{mid}} - \lambda_{\text{cut}}} && \text{(E.14)} \\ \lambda_0 &\approx 0.74\lambda_{\text{mid}} + 0.26\lambda_{\text{cut}} && \text{(E.15)} \\ m &= \frac{0.5}{\lambda_{\text{mid}} - \lambda_{\text{cut}}}, && \text{(E.16)} \end{aligned}$$

such that we can rewrite Eq. (E.13) as,

$$\begin{aligned} T(\lambda) &= \exp\left[-\exp\left(-\frac{1.43}{\lambda_{\text{mid}} - \lambda_{\text{cut}}}(\lambda - 0.74\lambda_{\text{mid}} + 0.26\lambda_{\text{cut}})\right)\right] && \text{(E.17)} \\ &= \exp\left[-\exp\left(-\frac{1.43}{\lambda_{\text{mid}} - \lambda_{\text{cut}}}(\lambda - \lambda_{\text{cut}})\right)\exp(1.06)\right] && \text{(E.18)} \\ &= \exp\left[-1.82\exp(-2.85\, m\, (\lambda - \lambda_{\text{cut}}))\right] && \text{(E.19)} \end{aligned}$$