FUNCTIONAL DECOMPOSITION OF RETINAL GANGLION CELL RECEPTIVE FIELDS

Dissertation

for the award of the degree
"Doctor rerum naturalium"
of the Georg-August-Universität Göttingen

within the doctoral program
Sensory and Motor Neuroscience
of the Georg-August University School of Science (GAUSS)

submitted by

SÖREN JOHANNES ZAPP

from Darmstadt, Germany

Göttingen, 2023

THESIS COMMITTEE

FIRST REFEREE AND SUPERVISOR
Prof. Dr. Tim Gollisch
Department of Ophthalmology, University Medical Center Göttingen

SECOND REFEREE
Dr. Jan Clemens
European Neuroscience Institute, Göttingen

Prof. Dr. Hansjörg Scherberger
Neurobiology Laboratory, German Primate Center, Göttingen

FURTHER MEMBERS OF THE EXAMINATION BOARD

Prof. Dr. Alexander Ecker
Institute of Computer Science, University of Göttingen

Dr. Viola Priesemann
Max Planck Institute for Dynamics and Self-Organization, Göttingen

Dr. Caspar M. Schwiedrzik
European Neuroscience Institute, Göttingen

Date of oral examination: January 25, 2023

doi:10.53846/goediss-9740

ABSTRACT

The retina has the fascinating ability to extract various visual features from our surroundings. This feature detection is enabled by nonlinear operations in the retinal circuitry. The nonlinear processing is in part attributed to retinal ganglion cells and to their signal integration of the presynaptic circuitry. Of particular interest are nonlinear subunits in the receptive field of ganglion cells that are thought to correspond to the excitatory inputs of presynaptic bipolar cells. As the access to bipolar cells is experimentally limited, subunits are typically inferred computationally from ganglion cell responses. However, current computational methods are demanding in both time and computational hardware. This prevents them from being applied easily in routine analyses.

Based on an extensive review of the current landscape of subunit analysis and its challenges, I developed computational methods that recover subunits without high demands for time or computational power. With the development of tools for hyperparameter selection, the methods identify subunits in different ganglion cell types from various species under variable experimental conditions. Furthermore, I conducted multielectrode-array recordings in the isolated marmoset retina to investigate the spatial arrangement of subunits using the developed methods. The population analyses of parasol and midget ganglion cells reveal that their subunits are arranged in distinct mosaics that tile the visual space. The subunit mosaics of ON and OFF midget cells align spatially, as do subunit mosaics of ON and OFF parasol cells. Facilitated by the methods, these findings suggest that the systematic spatial coordination of ON and OFF channels can be traced back computationally to the level of bipolar cells. Together, this thesis provides computational methods of subunit inference that overcome prevalent challenges and demonstrate new insights into the nonlinear signal integration in the primate retina.

ZUSAMMENFASSUNG

Die Netzhaut besitzt die faszinierende Fähigkeit, eine Vielzahl an visuellen Eigenschaften aus unserer Umgebung zu extrahieren. Dies wird durch nichtlineare Operationen im retinalen Netzwerk gewährleistet. Die nichtlineare Verarbeitung wird zum Teil retinalen Ganglienzellen und ihrer Verrechnung von präsynaptischen Signalen zugeschrieben. Von besonderer Bedeutung sind nichtlineare Einheiten, sogenannte Subunits, im rezeptiven Feld von Ganglienzellen, welche den erregenden Signalen von präsynaptischen Bipolarzellen entsprechen sollen. Da Bipolarzellen experimentell schwer zugänglich sind, werden die Subunits üblicherweise rechnerisch von neuronalen Antworten von Ganglienzellen abgeleitet. Gegenwärtige rechnerische Methoden haben jedoch hohe Ansprüche an Zeit und rechnerischer Leistung. Das erschwert ihre Verwendung innerhalb von Routineanalysen.

Ausgehend von einer umfangreichen Aufarbeitung der Herausforderungen von gegenwärtigen Subunitanalysen, habe ich rechnerische Verfahren entwickelt, die Subunits ermitteln, ohne große Anforderungen an Zeit oder Computerausstattung. Mit der Erstellung von Werkzeugen zur Wahl von Hyperparametern können die Methoden Subunits in verschiedenen Ganglienzelltypen, von verschieden Spezies und unter variierenden experimentellen Konditionen identifizieren. Des Weiteren habe ich Ableitungen mit Mikroelektrodenarrays in der isolierten Marmosetnetzhaut durchgeführt, um die räumliche Anordnung von Subunits mit den entwickelten Methoden zu untersuchen. Die Populationsanalysen von Parasol- und Midget-Ganglienzellen zeigen, dass deren Subunits in separaten Mosaiken strukturiert sind, die das visuelle Feld gleichmäßig bedecken. Die Subunitmosaike von ON- und OFF-Midget-Ganglienzellen und die der ON- und OFF-Parasol-Ganglienzellen sind jeweils räumlich zueinander ausgerichtet. Diese Befunde weise auf eine systematische räumliche Koordination von ON- und OFF-Informationskanälen hin, die sich rechnerisch bis zu Bipolarzellen zurückführen lässt. Zusammenfassend bietet diese Dissertation analytische Methoden zur Subunitableitung, welche gegenwärtige Herausforderungen überwindet und neue Einblicke in die nichtlineare Signalverarbeitung der Primatennetzhaut gewährt.

CONTENTS

1 GENERAL INTRODUCTION

In a society designed around the ability to see, vision is an essential aspect of every part of life. It is the most demanded of the five senses and represents an integral factor of social interaction, information exchange, and spatial navigation in a fast paced environment (Desrosiers et al. 2009; Heine and Browning 2002). With this in mind, vision impairment has severe consequences for individuals, their careers, and relationships. This is of particular significance from a medical perspective. The most common eye-related inherited disease is retinitis pigmentosa, a degenerative disease leading to photoreceptor loss that is considered untreatable (Hartong et al. 2006).

To provide effective vision restoration, it is crucial to understand the implications of the defect, the workings of a healthy eye, and in particular, how the eye processes visual information. When light enters the eye, it is projected onto the retina, a thin sheet of neural tissue at the back of the eyeball. In the healthy eye, the light is detected by photoreceptors, and the visual information is processed by the neuronal circuitry of the retina before it is sent to downstream areas of the brain in form of electrical signals, so-called action potentials or spikes. The sequences of spikes are generated by ganglion cells, representing the output of the retina, encoding all visual information that gives rise to the sense of vision. Without photoreceptors, the interface with the outside world is missing and no meaningful visual information can be encoded by the retina.

A promising treatment is optogenetic therapy that inserts light-sensitive proteins in targeted cells, effectively turning them into substitute light detectors (Deisseroth et al. 2006). With the aim to restore visual capabilities, often the targeting of retinal ganglion cells (RGCs) is studied (see Lindner et al. 2022). However, the primate retina is rich in diversity of ganglion cell types (Dacey 2004; Masri et al. 2019a; Y.-R. Peng et al. 2019; Yan et al. 2020b) that form parallel channels that encode distinct features like luminance, local contrast, and color (Grünert and Martin 2020). Their extraction involves various nonlinear computations (Gollisch and Meister 2010; G. W. Schwartz and Rieke 2011; Silveira and Roska 2011). Consequently, with a non-selective optogenetic targeting of RGCs, multiple types are affected. Their parallel and nonlinear processing of visual information in their presynaptic circuitry is by-passed and lost.

As alternative candidates, bipolar cells, the excitatory interneurons with converging inputs to RGCs, have been investigated recently (Cehajic-Kapetanovic et al. 2015; Kralik et al. 2022; Lagali et al. 2008). They remain largely unaffected by retinitis pigmentosa (Santos et al. 1997; J. L. Stone et al. 1992) and functionally intact (M. H. Berry et al. 2019; Cehajic-Kapetanovic et al. 2015; Lagali et al. 2008; Wyk et al. 2015). Consequently, directly imparting light sensitivity to bipolar cells instead, preserves the inner retinal computations. Nevertheless, to produce retinal spiking patterns that are coherent to the downstream visual system, mindful optogenetic stimulation is essential. A thorough understanding of the connections between bipolar cells and RGCs is necessary. While morphological studies have provided generous insights into the convergence between

the two (Dacey 1993; Kolb et al. 1992), physiological investigation is necessary to quantify their functional connectivity. It is imperative to understand how bipolar cell signals are integrated by RGCs with respect to nonlinear operations and different cell types. Because bipolar cell activity is difficult to access experimentally, computational approaches have moved into focus. With statistics and model-based techniques, the activation of RGCs is decomposed into its presynaptic inputs. By modelling the interplay between the identified components, the methods aim to decipher how the retina encodes visual input. These approaches have proven successful in recovering bipolar cell contributions for various species (J. K. Liu et al. 2017; Maheswaranathan et al. 2018; Shah et al. 2020). Nevertheless, to aid medical development, they lack flexibility in application to large cell populations and have yet to provide actionable insight into the cell-type specific wiring in the retina. To investigate, a first step is a closer look at the circuitry from which the nonlinear operations emerge.

1.1 RETINAL CIRCUITRY AND ENCODING

The retina (Figure 1.1) is segregated into alternating layers of cell bodies and synaptic connections (Masland 2012; Wässle 2004). The photoreceptors are situated in the outer retina, furthest away from the center of the eye. They consist of rods and cones, primarily associated with scotopic vision in low-light conditions and with color vision, respectively. Photoreceptors convert light signals into electrical signals which are passed on to bipolar cells that reside in the inner nuclear layer. Generally, several of these excitatory interneurons synapse onto one RGC in the ganglion cell layer at the front-facing side of the retina. The RGCs generate patterns of spikes sent along the optic nerve to the lateral geniculate nucleus (LGN) in the thalamus, the next stage in the sensory visual pathway. The excitatory feed-forward network of bipolar cells and RGCs is complemented by an inhibitory network of horizontal cells and amacrine cells, located in the inner nuclear layer. While horizontal cells receive input from photoreceptors and primarily feed back onto them, amacrine cells mediate signals on the level of bipolar cells and RGCs.

This seemingly complete rule set of connectivity is riddled with exceptions that complicate this simple model. Examples are gap junctions between photoreceptor cells (Raviola and Gilula 1973), cone-specific bipolar cells being mediated indirectly by rod inputs (Nelson 1977), and intrinsically photosensitive RGCs (Foster et al. 1991). Nevertheless, nonlinear processing arises already in the simple feed-forward signal transduction described here (Demb et al. 2001a). It is therefore a valid starting point to cautiously ignore the heterologous coupling and to rather gradually extend an intepretable model of visual encoding.

Patterns of light that enter the eye are projected onto the retina like on a canvas, such that regions in the visual field are mapped to distinct locations on the retina (Drasdo and Fowler 1974). By their arrangement on the retina, cells encode constrained areas in visual space: their receptive fields. Originally conceived by Sherrington (1906) in a different sensory modality, Hartline (1938) experimentally characterized the receptive

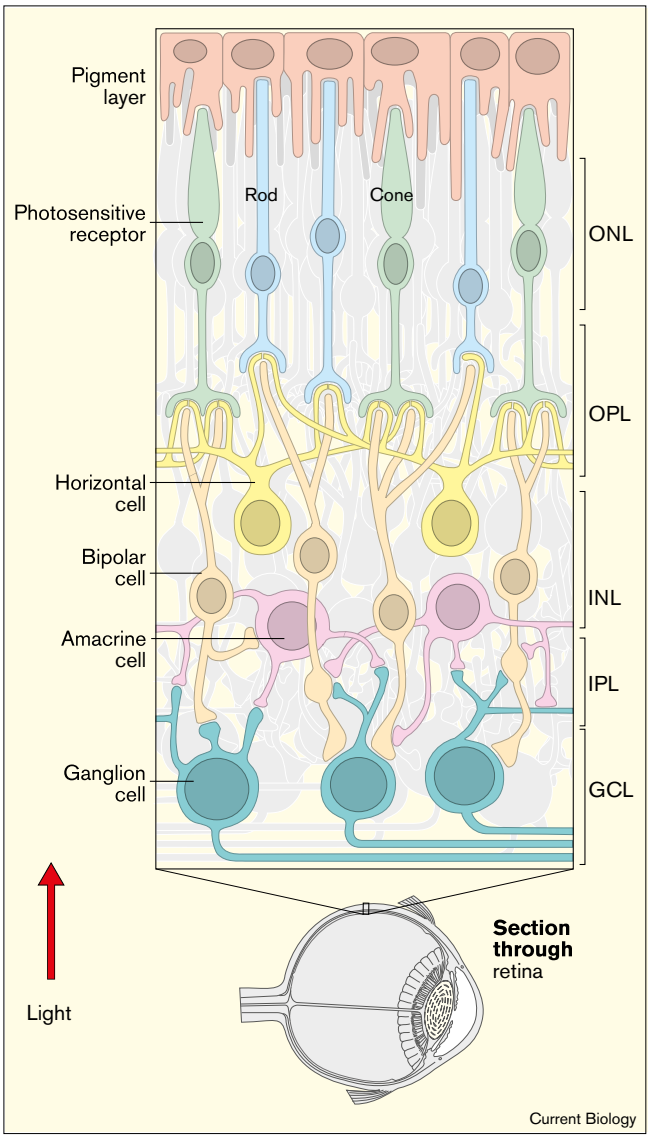


Figure 1.1: **The layers and cells of the retina.** Photoreceptors reside in the outer nuclear layer (ONL) synapsing onto bipolar cells and horizontal cells in the outer plexiform layer (OPL). The cell bodies of horizontal cells, bipolar cells, and amacrine cells are situated in the inner nuclear layer (INL). Connections between bipolar cells, amacrine cells and ganglion cells are formed in the inner plexiform layer (IPL). Ganglion cells are located in the ganglion cell layer (GCL). Light travels in the reverse direction (bottom to top).

field for ganglion cells in the retina. Later, the concept of the receptive field was adopted beyond the retina in other brain areas of visual processing, most notably, with the targeted stimulation of cells in the cat primary visual cortex (Hubel and Wiesel 1962). The RGC receptive field constitutes a circle surrounded by an antagonistic concentric ring, arranging in a center-surround structure (Barlow 1953; Kuffler 1953). Center and surround are tuned oppositely to light, resulting in ON-center and OFF-center receptive fields. ON-center cells respond to light increments in the center and decrements in the surround, while OFF-center cells exhibit the opposite behavior. To model a RGC receptive field adequately, the difference of two Gaussian functions has proven fruitful (Rodieck 1965). This resonates with efficient coding theory of the visual system (Atick and Redlich 1992; Barlow 1961), as encoding algorithms trained on natural images converge to a collection of difference-of-Gaussians models (Roy et al. 2021). Often complemented with a temporal description of impulse response, the spatiotemporal receptive field (Stevens and Gerstein 1976) has become an integral part in visual neuroscience (DeAngelis et al. 1995) and beyond (Aertsen and Johannesma 1981; DiCarlo and Johnson 2000).

A general view is that neurons in the visual system encode a visual scene by comparing how similar it is to a preferred pattern, defined by their receptive field. This enables characterizing a cell's receptive field as a stimulus filter (Enroth-Cugell and Pinto 1970; Movshon et al. 1978). For the retina to produce a response, an incoming stimulus is filtered by the RGCs' receptive fields and is subsequently subjected to nonlinear spike generation (Kistler et al. 1997), such as non-negativity, thresholding, and saturation in spike frequency (Chichilnisky 2001; Simoncelli et al. 2004), also called firing rate. This chain of a linear and a nonlinear stage has allowed predicting firing rate responses of RGCs successfully across species (Chichilnisky 2001; Keat et al. 2001) and has been adequately termed the linear-nonlinear (LN) model (De Boer and Kuyper 1968; Meister and M. J. Berry 1999; Victor and Shapley 1980).

1.2 NONLINEAR SPATIAL INTEGRATION

The LN model fails to be universally complete on two accounts. On the one hand, the simple description of linear stimulus filtering does not match the reported complexity of retinal morphology (Cajal 1893; see Rodieck 1973; Shkolnik-Yarros 1971) and the vast abundance of morphologically identified RGC types in primates (Dacey 2004; Masri et al. 2019a; Y.-R. Peng et al. 2019; Yan et al. 2020b) and across other mammals as reported in mouse (Bae et al. 2018; Sanes and Masland 2015; Yan et al. 2020a), cat (Kolb et al. 1981), rabbit (Rockhill et al. 2002), and salamander (Toris et al. 1995; J. Wang et al. 2016). On the other hand, it is insufficient to explain all aspects of retinal encoding.

For instance, in an investigation of the effects of contrast variations across the receptive field of RGCs in the cat retina, Enroth-Cugell and Robson (1966) identified characteristics diverging from the classical linear description (Kuffler 1953; Rodieck and J. Stone 1965). They discovered differences in the response to an alternating grating of dark and bright presented at different spatial phases (Figure 1.2A). For one group, the X cells, the response ceased for certain stimulus positions, as is to be expected if they summed their

receptive fields linearly over space, such that bright and dark portions negate each other (Figure 1.2B). The other group of cells persisted to respond even in the absence of a change in net luminance across their receptive fields (Figure 1.2C). These were termed Y cells identifiable by their nonlinear spatial summation (Hochstein and Shapley 1976a). The nonlinear integration has been attributed to distinct substructures acting as nonlinear subunits within the Y-cell receptive field (Derrington et al. 1979; Hochstein and Shapley 1976b). These localized subunits operate as individual spatial filters smaller than the receptive field center and their signals are nonlinearly transformed before the summation in the receptive field (Figure 1.2E). The idea about a linear-nonlinear-linear-nonlinear (LNLN) model emerged, a cascade of LN models, with the first stage representing the nonlinear subunits (Victor and Shapley 1979b). The LNLN model has since become a popular concept of nonlinear signal integration and has enjoyed success in demonstrating the nonlinear nature of receptive fields in the retina of various species with appropriate model variations and extensions (J. Freeman et al. 2015; Maheswaranathan et al. 2018; McFarland et al. 2013; Real et al. 2017; Vintch et al. 2012).

The subunits had been hypothesized to originate from activation of nonlinear amacrine cells (Derrington et al. 1979; Fischer et al. 1975; Hochstein and Shapley 1976b; Victor et al. 1977; Victor and Shapley 1979b). But selectively blocking bipolar cell and amacrine cell input revealed that excitatory input of bipolar cells suffices to construct a nonlinear receptive field if the signal is rectified (Demb et al. 1999, 2001a). This nonlinear signal transmission was confirmed in bipolar cells (Borghuis et al. 2013; G. W. Schwartz et al. 2012). Although an interplay of amacrine cells and other bipolar cells (Kuo et al. 2016) and nonlinear computation in bipolar cells (Schreyer and Gollisch 2021) cannot be excluded, the focus has shifted to 'effective' bipolar cell input, the combined ensemble of bipolar cell input modulated by inhibitory interactions (Vaney et al. 2001). This may provide a good description of a subunit, as inhibitory effects shape the response of bipolar cells largely before integration at the ganglion cell (Asari and Meister 2012; Zaghloul et al. 2007).

Together, these findings allow establishing an intuitive understanding of the functional connectivity of X cells and Y cells. Much like RGCs, the response of a bipolar cells can be characterized by their receptive fields (Werblin and Dowling 1969). As a ganglion cell pools the inputs from several bipolar cells, its receptive field integrates the narrower receptive fields of the presynaptic neurons and consequently spans a larger spatial area on the visual field. To facilitate the emergence of X cells and Y cells, the bipolar cell receptive fields are integrated in either linear or nonlinear fashion. This resonates with the proposal of Hochstein and Shapley (1976b) of linear and rectified (nonlinear) subunits.

1.3 SUBUNIT FUNCTION

Many phenomena of RGC computation originate from subunits within the receptive field (Gollisch and Meister 2010). Various functional characteristics of subunits across cell types and species have been explained (Bölinger and Gollisch 2012; Enroth-Cugell

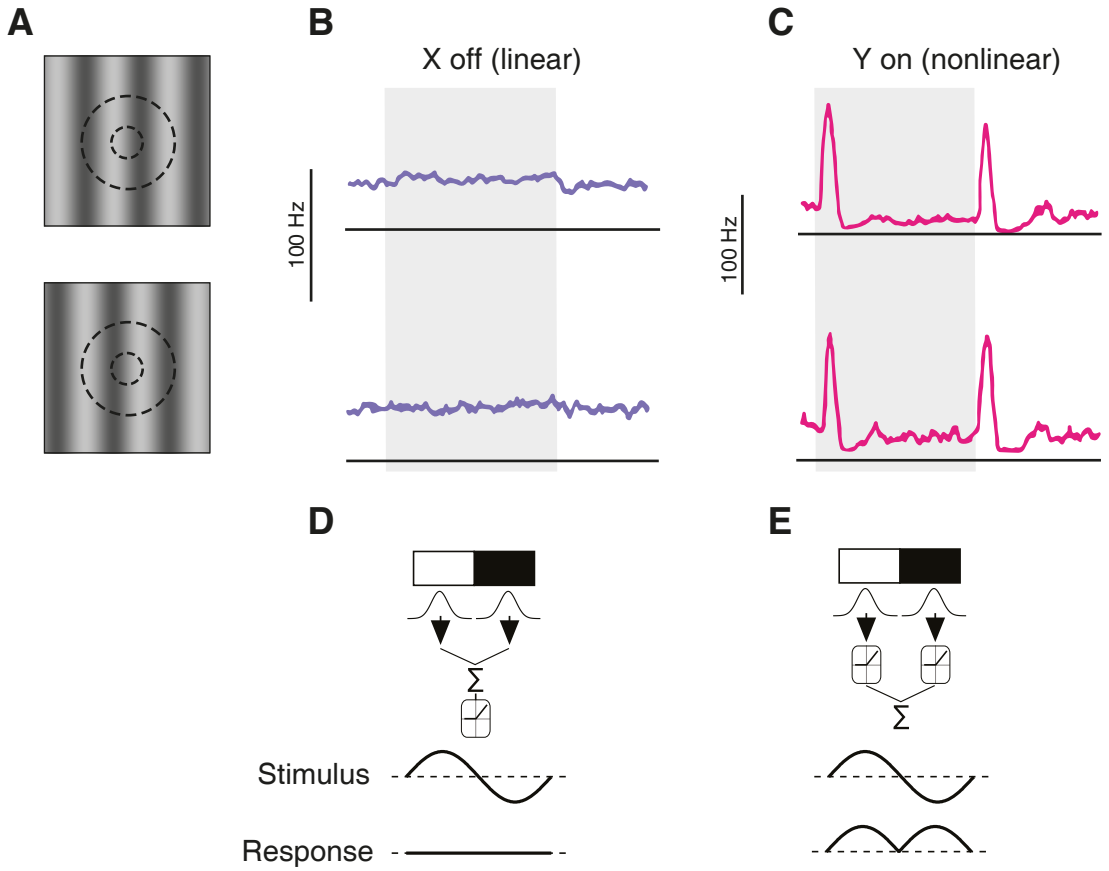


Figure 1.2: **X-cell and Y-cell response to reversing-grating stimulus.** Recordings from the cat retina (Enroth-Cugell and Robson 1966, 1984). **(A)** Two different positions of the stimulus pattern with respect to the center-surround receptive field for the duration of the presentation. **(B)** Firing rate of an OFF-center X cell over time. Shaded gray area indicates the presentation of the grated stimulus pattern. The cell exhibits linear behavior as it does not respond to the cancelled out dark and bright segments. **(C)** Firing rate of an ON-center Y cell over time. The cell nonlinearly sums the receptive field and shows a peak in firing rate at onset and offset of the stimulus presentation. Scale bars correspond to the firing rate frequency. **(D)** Schematic representation of linear stimulus integration. The dark-and-white patch corresponds to the structured stimulus over the RGC receptive field center. Subunits are depicted as one-dimensional bell curves filtering the input at different spatial locations. Below, the activation is summed and passed through a final nonlinearity. The stimulus of varying contrast over time never elicits a response, because the stimulus always sums to zero. **(E)** Similar, but for nonlinear stimulus integration. Each subunit activation is rectified (subunit nonlinearity) before the summation, removing negative activation. The model sums the stimulus nonlinearly and produces two responses, a so-called "frequency-doubling" response. Note, that a final nonlinearity, typically part of the LNLN model, is not included here.

and A. W. Freeman 1987; Hochstein and Shapley 1976a; G. W. Schwartz et al. 2012; Victor and Shapley 1979a,b). As the experiments of Enroth-Cugell and Robson (1966) demonstrate, subunits are involved in the response to stimulation at fine spatial resolution. Furthermore, the sensitivity to high spatial frequency facilitates recognizing consecutive activation of subunits, effectively establishing motion detector neurons already in the early stages of sensory processing (Demb et al. 2001b). Applying the concept of subunits to the surround of the receptive field, allows explaining its involvement in nonlinear signal integration and motion sensitivity. This extends the retina's capabilities of disinhibitory signaling and equips it with sensitivity to local motion (Jadzinsky and Baccus 2015; Ölveczky et al. 2003, 2007; Zhang et al. 2012). Related to that may be the detection of looming objects over the receptive field enabling a single RGC type to trigger defensive behavior in response to an approaching predator (Münch et al. 2009; F. Wang et al. 2021). These examples establish the important role of subunits in local spatial operations of ganglion cell types. The investigation of how the subunits contribute to these computations is not concluded. In particular, the number of subunits in the receptive field, their excitatory and inhibitory interactions, and how their arrangement may give rise to and shape the computations of different ganglion cell types is yet to be examined. To answer these open questions it is imperative to investigate the functional decomposition of the receptive field of RGCs. To gain valuable insight for medical application in humans, the analysis of subunits in the primate retina is of particular interest.

1.4 PRIMATE MODEL

The anatomical organization of the mammalian nervous system follows a common set of developmental rules across species (Krubitzer 2007). Non-primate animal models therefore pose a suitable surrogate for invasive experimental studies to drive advancements in neuroscience and human medicine. Nevertheless, substantial evolutionary differences set the visual system of primates apart from non-primates. Common animal models like the mouse (*Mus musculus*) have lower visual acuity (Prusky and R. M. Douglas 2004) and are subjected to a different visual environment resulting in developmental differences in the distribution of cell types across the retina (Qiu et al. 2021). Non-human primates, on the contrary, share similarities with humans (*Homo sapiens*), like well-segregated parallel streams in the thalamus and orientation columns in the visual cortex (see Preuss 2007; Nassi and Callaway 2009). Some non-human primates also have a fovea and perceive high acuity trichormatic color (Jacobs 2008). Specifically in the retina, investigations of RGCs in humans matched the description of non-human primate RGCs more so than those in non-primates (Reinhard and Münch 2021). Consequently, the gap of inference is smaller from non-human primates to humans than from other mammals (Kling et al. 2020). From the basic-research perspective, it is to note that recent developments suggest that optogenetic studies in primates may soon be within reach (J. E. Park and Silva 2019; Sasaki et al. 2009).

Among non-human primates, the common marmoset (*Callithrix jacchus*) is a New World primate that has gained popularity in visual neuroscience over the recent years (Mitchell and Leopold 2015; Solomon and Rosa 2014). Its visual system shares high similarity with that of the more commonly studied macaque (*Macaca* spp.) model (see Solomon and Rosa 2014; see Masri et al. 2019a). In particular, the visual optics and cone density in its retina show high similarity to those of macaques and humans (Troilo et al. 1993). This high degree of morphological similarity allows transferring previous insights of retinal circuitry from studies of the macaque (Chan et al. 1997; Chan and Grünert 1998; Chan et al. 2001; Ghosh et al. 1996; Jusuf et al. 2004). The recent advancements in genetically engineered models (J. E. Park and Silva 2019; Sasaki et al. 2009) and the exposed visual system on the outer surfaces of the cortex (Paxinos et al. 2012), make the marmoset an attractive animal model. Furthermore, the marmoset retina seems to show little age-related variability (Haverkamp et al. 2022), easing the comparison of scientific insights across individual animals.

Most ganglion cells and bipolar cells in the marmoset retina match the description of ON-center or OFF-center cells (Protti et al. 2014). The most numerous ganglion cells in the marmoset retina are the midget cells that extend their axons into the parvocellular layers of the LGN (Gomes et al. 2005; Goodchild et al. 1996; Jusuf et al. 2006a). The next most common type is the parasol ganglion cell that receives input from multiple diffuse bipolar cells (Chan et al. 2001; Eriköz et al. 2008; Gomes et al. 2005) and connects to the magnocellular layers of the LGN (Szmajda et al. 2008).

These two RGC types receive a lot of attention due to their reliable occurrence across non-human primates (Goodchild et al. 1996; Watanabe and Rodieck 1989) and humans (Dacey and Petersen 1992; Kolb et al. 1992; Rodieck et al. 1985; Soto et al. 2020), where they make up around 80% of ganglion cells (Dacey 1993, 2004; Dacey and Petersen 1992; G. D. Field and Chichilnisky 2007; Masri et al. 2019b; Yan et al. 2020b). They were first described by Polyak, who extended the efforts of Cajal (1893) with Golgi staining in the primate retina (Poljak 1935; Polyak 1941).

Midget ganglion cells have small receptive fields, encoding high acuity signals and predominantly chromatic information (Kolb and Marshak 2003; Livingstone and Hubel 1988). With comparably larger receptive fields at any given retinal eccentricity, parasol ganglion cells exhibit lower spatial resolution and specialize in edge detection and motion detection (Croner and Kaplan 1995; Livingstone and Hubel 1988). Cells of both types are found throughout retina (Dacey 1993; Dacey and Petersen 1992; Kolb et al. 1992; Polyak 1941; Rodieck et al. 1985), each tiling the retina in a regularly spaced mosaic to maximize their coverage of the visual field (G. D. Field and Chichilnisky 2007). Selective loss of each type results in distinct visual impairment (Merigan and Maunsell 1993).

Aside from their morphological classification, their light responses have been extensively studied (Dacey 2004; G. D. Field and Chichilnisky 2007; Grimes et al. 2018; B. B. Lee 1996). After a long history of efforts (Benardete et al. 1992; Monasterio 1978; White et al. 2002), the relation to Y cells observed in cat was finally drawn recently, identifying the physiological characteristics of Y cells in smooth and large radiate cells, as well as among parasol cells in primate (Crook et al. 2008a,b; Petrusca et al. 2007). Specifically OFF parasol cells have been reported to respond to finely structured natural

stimuli not captured with the LN model (Turner and Rieke 2016). Similar observations have been made in midget cells that were described better with a nonlinear receptive field (Freedland and Rieke 2022; J. Freeman et al. 2015). The investigation of nonlinear operations in the primate retina are ongoing. To analyze functional receptive field characteristics, neuronal responses to stimulation are recorded from the retina.

1.5 MULTI-ELECTRODE RECORDINGS OF GANGLION CELLS

RGC spikes, representing the output of the retina, are classically recorded from isolated retinas in *in vitro* experiments. Visual stimuli are projected onto the photoreceptor side of the retina, while extracellular electrodes record voltage modulations from the the ganglion cell layer. The prevalent use of planar multi-electrode arrays (MEAs) for extracellular recordings (Meister et al. 1994) is explained by a number of benefits. Recordings based on voltage allow for high temporal resolution, crucial for aligning the stimulus-response mapping at a neural time scale. Unlike in single cell recordings (Asari and Meister 2012; Vries et al. 2011), extracellular recordings do not require probing and identifying cell classes at time of recording. On the contrary, in a setup with multiple electrodes, responses from many cells are recorded simultaneously. Cell types are differentiated post-hoc, based on their response properties. Furthermore, the use of a planar recording array provides a suitable arrangement for non-obscured light stimulation, real-time inspection with a microscope, and constant perfusion with nutrient-rich solution to keep the cells alive. Because the recorded cells are not penetrated with an electrode during extracellular recordings, ganglion cells stay functional for a longer duration. These advantages present a versatile method for statistical analysis of entire cell populations.

For analysis, the spikes are extracted from the voltage traces by their amplitude and ascribed to distinct neural units in a subsequent process termed spike-sorting. Using an assortment of methods, including clustering of spike amplitudes and electrode-triangulation of spike origin, responses from individual cells are segregated as best as possible (Carlson and Carin 2019). Although this process is largely automated (J. Lee et al. 2020; Pachitariu et al. 2016; Yger et al. 2018), the complex categorization usually requires manual curation afterwards. To classify the acquired RGCs responses by ganglion cell type - in primate that may be the four major types of ON parasol, OFF parasol, ON midget, and OFF midget - various techniques are used (Baden et al. 2016; Drinnenberg et al. 2018; Ravi et al. 2018). The most common method is to categorize the cells by their known physiological properties, like their spatial receptive field sizes, impulse response, and sustained or transient spiking behavior (Goetz et al. 2022; Rhoades et al. 2019). This is often complemented by verification of non-overlapping receptive fields and a regularly spaced mosaic per cell type (Gauthier et al. 2009). The receptive field is calculated using reverse correlation techniques following white-noise stimulation (Chichilnisky 2001).

Aside from the RGC receptive field, the relationship between ganglion cell and bipolar cell responses are of special interest. Their functional connection is essential to gaining more insights about the properties of receptive field subunits. However, unlike RGCs that

are exposed to MEAs in the ganglion cell layer of the retina, the somas of bipolar cells are buried in the inner nuclear layer that is difficult to access experimentally (Asari and Meister 2012; Bohl et al. 2022). Additionally, these interneurons are largely non-spiking, further complicating their identification and analysis (Asari and Meister 2012; Vries et al. 2011). Especially in combined recordings with RGCs, it is currently not feasible to obtain a complete picture of bipolar-to-ganglion cell circuitry (J. K. Liu et al. 2017), let alone on a population level.

1.6 INFERENCE OF RECEPTIVE FIELD SUBSTRUCTURE

Understanding how bipolar cells contribute to the response of a ganglion cell without direct access to its presynaptic circuitry is challenging. Due to the experimental limitations, the focus has shifted to analytical approaches. Various computational models are currently being developed to estimate the bipolar cell input to a RGC in form of subunits in its receptive field. The bipolar-cell subunits are inferred from spiking responses of the RGC recorded under controlled light stimulation.

The most intuitive approach to estimate subunits is to train a LNLN model end-to-end using an artificial stimulus as input and the RGC responses as output. The subunits arise as filters from parameter fitting in the first stage of the LN cascade. To guide the selection of subunit number, anatomical investigations have been successfully incorporated (J. Freeman et al. 2015; G. W. Schwartz et al. 2012). If no prior information on subunit layout is available, the number of subunits has to be chosen manually or cross-validated on the model response prediction performance (Maheswaranathan et al. 2018). Reducing the number of spatial dimensions of the stimulus aids in training convergence, allowing to successfully match subunits to bipolar cell receptive fields (Maheswaranathan et al. 2018; Real et al. 2017).

Using a convolutional neural network model has proven to be another intriguing technique of describing the retinal circuitry (Maheswaranathan et al. 2019; McIntosh et al. 2016; Tanaka et al. 2019). This model, as well, is trained in supervised manner on stimulus and response. But to accommodate the high degree of model complexity, the parameters are fit using the data of a population of RGCs instead. The convolution, consequently, forms abstract subunit filters invariant in position that reoccur across the RGCs. Trained on natural stimuli, the filters provide familiar characteristics of retinal cells, particularly the center-surround structure of bipolar cell receptive fields (Maheswaranathan et al. 2019; Tanaka et al. 2019). Nevertheless, the model complexity obscures an interpretable mapping to the retinal circuitry.

On the other end of the spectrum are subunit recovery approaches based on statistical inference. At the core of these methods sits the collection of stimuli corresponding to points in time around ganglion cell spiking. To enable unbiased statistical analysis, the stimulus is often based on white-noise sequences at cost of potentially long recording duration due to its weak stimulation (Meister et al. 1994).

Spike-triggered covariance (STC) offers a fast way to decompose the receptive field into separate filters based on principal component analysis (Samengo and Gollisch 2013; O.

Schwartz et al. 2006). These filters are, however, holistic representations of the receptive field that do not resemble localized subunits (Gollisch 2013; Kaardal et al. 2013).

An extension to the decomposition approach is an added non-negativity constraint. Spike-triggered non-negative matrix factorization (STNMF) has been demonstrated to yield localized subunits that match the receptive fields of bipolar cells (J. K. Liu et al. 2017). The non-negativity constraint on the components aids in recovering localized structures (D. D. Lee and Seung 1999). However, this does not allow the recovery of center-surround subunits. Both STC and STNMF do not require an a priori choice of the number of subunits. Instead, subunits are selected among the decomposition by their spatial localization in STNMF (J. K. Liu et al. 2017).

The method of spike-triggered clustering seeks out distinct patterns among the stimulus collection corresponding to subunits. Localized subunits have been shown to arise from hierarchical clustering interleaved with continuously updating a trained LNLN model (Shah et al. 2020). Cross-validation on the response prediction performance guides in the selection of the number of subunits.

This brief assessment of existing methods reveals that they are diverse in their approaches and in their advantages. In order to place subunits into the context of the computational tasks of the retina, subunit recovery should integrate well as a small part of larger analyses of broader focus. However, the current approaches are implemented as extensive studies exclusive to the recovery of subunits. They require long or stimulus-specific recordings leaving little room for further investigations. Moreover, all of them share a time-intensive analysis demanding powerful computing hardware. A desirable goal, on the contrary, is a compact analysis that seamlessly integrates into a routine analysis pipeline. Naturally, subunit recovery should verifiably match bipolar cell inputs and additionally reflect their center-surround structure. For ease of application, extensive cross-validation and hyperparameter tuning should be avoided. Instead of invariant or abstract filters, such a method should recover distinct subunits in size, shape, location, and input contribution to allow investigation of overlap and interaction between subunits within and across receptive fields. And finally, the implementation ideally accepts data of individual RGCs as input, instead of requiring large cell populations that may not always be available in a recording.

Although the existing methods cover many aspects, none of them fulfill all of these criteria in their current state. This leaves a gap between theoretical subunit recovery and practical application in conjunction with broader research questions of nonlinear computations in the retina. Filling this gap is the essential key to answering how RGCs encode lateral (Barlow et al. 1964; Wei 2018), looming (Münch et al. 2009), and local motion (Ölveczky et al. 2003; Zhang et al. 2012), what aspects of natural image processing can be attributed to excitatory bipolar cell inputs (Yu et al. 2022), what are the major drivers behind the antagonistic surround of the RGC receptive field (Merwine et al. 1995; Takeshita and Gollisch 2014), how are local changes of contrast extracted (Kuffler 1953), how are chromatic signals integrated (Baden and Osorio 2019), how do morphologically classified RGCs differ in their functional encoding (Yan et al. 2020b), and how is the detection of other nonlinear features implemented in the retina (Gollisch and Meister

2010). Answering these and more questions requires appropriate tools at hand and the development of a suitable subunit recovery method is the fundamental next step.

1.7 THESIS OUTLINE

In this thesis, I investigate the inference of receptive field subunits of RGCs. I develop new methods of subunit recovery in an attempt to move answers to open questions in current research closer into reach.

I first give an account of the history and current efforts in the investigation of nonlinear spatial integration in the retina (Chapter 2). To establish the functional relevance of nonlinear subunits, I contrast examples of how linear approximations fail to describe responses of RGCs with known phenomena of nonlinear computations in the retina. I review state-of-the-art statistical and model-based methods of subunit inference in more detail. These are put into context with natural stimulation, temporal dynamics, and their correspondence to bipolar cell input. Open questions emphasize the need for further efforts to tackle prevailing challenges in the estimation of the receptive field substructure, mentioning matrix factorization as promising candidate.

To pave the way towards answering these questions, I develop a more compact method of subunit recovery (Chapter 3). To address the prevalent shortcomings of the current methods, the aim is to provide the receptive field decomposition in a matter of seconds to facilitate population analyses of RGC recordings. On the basis of STNMF, the method is implemented by employing modern algorithms capitalizing on speed through iterative approximation. To make an effort on another drawback of current methods, I introduce techniques of hyperparameter tuning that provide shared configurations across RGCs of the same functional type. I record data from the primate retina to demonstrate the speed and reliability of the algorithm and use salamander data to match subunits with recorded bipolar cell receptive fields. Subunits are inferrable with only 40 minutes of recording time at an analysis time of a few sconds. Recovered subunit layouts of midget and parasol ganglion cells in the primate retina reveal new insights relevant for current investigations of how parallel signaling channels, such as ON and OFF pathways, are spatially coordinated.

Investigating the potential of dimensionality reduction with matrix factorization, I devise two algorithms that address further obstacles (Chapter 4). I dissect the mathematics involved in the constraints of non-negative matrix factorization to eliminate the need for explicit regularization. I demonstrate how different constraints in the matrix factorization are exchangeable to lift the non-negativity constraint. Applied to a dataset of facial images, the algorithm succeeds in recovering facial elements of positive and negative pixel values co-existing in one factorized component. This shows the potential of application to recover the center-surround structure of bipolar-cell-like subunits. The method delivers on the major shortcomings and has, in theory, the potential to drive forward the investigation of nonlinear spatial integration in the retina.

Finally, I discuss the proposed methods and the implications of the findings in the primate retina and put the thesis into broader scientific context (Chapter 5).

2 RETINAL RECEPTIVE-FIELD SUBSTRUCTURE: SCAFFOLDING FOR CODING AND COMPUTATION

Attached manuscript. Invited review published in *Trends in Neuroscience*. Reprinted below is the published manuscript, licensed under CC BY 4.0.

Authors:

Sören J. Zapp[1,2,†], Steffen Nitsche[1,2,†], Tim Gollisch[1,2,3]

1. University Medical Center Göttingen, Department of Ophthalmology, Göttingen, Germany
2. Bernstein Center for Computational Neuroscience, Göttingen, Germany
3. Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, Göttingen, Germany

† These authors contributed equally

Author contributions:

I reviewed the literature about past and current methods of subunit inference and about the involvement of subunits in direction selectivity, I contributed to writing the manuscript, and prepared the figures Figure 1, Figure 2, and Figure 3. SN reviewed the literature about subunit function and adaptation, contributed to writing the manuscript, and prepared the figures Figure 4, Figure 5, and Figure 6. TG supervised the literature research and prepared the initial draft. All authors helped revise the manuscript.

Review

Retinal receptive-field substructure: scaffolding for coding and computation

Sören J. Zapp [1,2,4], Steffen Nitsche [1,2,4] and Tim Gollisch [1,2,3,*,@]

The center-surround receptive field of retinal ganglion cells represents a fundamental concept for how the retina processes and encodes visual information. Yet, traditional approaches of using the receptive field as a linear filter to integrate light intensity over space often do not capture the responses of a ganglion cell to complex visual stimuli. Thus, models with local nonlinearities in subunits of the receptive field or with local temporal dynamics are emerging to better reflect relevant aspects of retinal circuitry and capture stimulus encoding. Here, we review recent efforts to identify such receptive-field substructure and evaluate its role in visual stimulus encoding. The concomitant development of new computational tools may pave the way toward a model-based, functional approach to retinal circuit analysis.

Highlights

Visual stimulus encoding by the retina is not fully captured by the center-surround receptive fields of retinal ganglion cells. Mounting evidence of nonlinear spatial signal integration under natural stimuli and of specific visual functions solved by distinct ganglion cell types indicates the need to better understand receptive-field substructure.

Nonlinear spatial integration can be captured by subdividing ganglion cell receptive fields into subunits, which are thought to correspond to presynaptic bipolar cells. Several statistical and model-based methods have recently been developed to identify the subunit layout from spiking responses of ganglion cells to visual stimuli.

The subunits provide scaffolding for retinal computations, which may act through local adaptation and inhibition to shape responses to dynamic stimulation and to extract specific visual features.

Beyond receptive-field filtering: aligning circuit complexity and functional diversity

The sense of vision in vertebrates relies on a range of different computations. These distill relevant information from the incoming light patterns while discarding irrelevant components and dynamically adapting to the recently encountered structure of inputs. It is now well established that the underlying computations begin in the retina [1]. In this neural network, a multitude of different neuron types (more than 100 in the mouse, [2]), form dozens of parallel information channels [3–5] to extract various visual features, such as lateral [6,7], looming [8], and local motion [9,10] as well as local contrast [11] and color [12].

Understanding what the diversity of neural hardware is for, and how it implements different computations, requires going beyond the often-held view that the main function of the retina is to filter incoming images by the center-surround receptive fields of ganglion cells. Instead, nonlinear interactions, local dynamics, and signal gating within the receptive field have become a focus of interest, requiring new approaches and techniques to dissect the substructure of receptive fields and the signal processing within [13]. These developments occur in a tight interaction of experimental investigations with advances in computational data analysis and retinal circuit models. Among the primary overarching goals are to: (i) determine what is needed to describe and predict retinal responses to natural stimuli; and (ii) infer the structure and operations of the presynaptic circuitry that shape the responses of the different output channels of the retina.

In this review, we highlight recent progress in analyzing the substructure of retinal ganglion cell receptive fields. We focus on the characteristics and functional roles of nonlinear signal integration over space and of local temporal dynamics within the receptive field. In doing so, we emphasize new computational techniques to infer retinal circuit structure from recordings of the spiking activity of ganglion cells. The new methodology resonates well with model-based analyses in other sensory areas and with current developments in the field of artificial neural networks. Given the preserved general structure of the retina across vertebrate species and the many similar

[1]Department of Ophthalmology, University Medical Center Göttingen, 37073 Göttingen, Germany
[2]Bernstein Center for Computational Neuroscience Göttingen, 37073 Göttingen, Germany
[3]Cluster of Excellence 'Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells' (MBExC), University of Göttingen, 37075 Göttingen, Germany
[4]These authors contributed equally

*Correspondence: tim.gollisch@med.uni-goettingen.de (T. Gollisch).
@Twitter: @TimGollisch (T. Gollisch).

basic operations, the presented examples are drawn freely from retinas of different vertebrates, mostly monkey, mouse, and salamander.

The center-surround receptive field in the retina and its challenges

The retina transduces incoming light into neuronal signals in its photoreceptors and processes these signals through its network of interneurons (bipolar, horizontal, and amacrine cells) before the retinal ganglion cells send the messages from the retina about the visual world as spike patterns to various brain regions (Figure 1A). The function of this network is often summarized by referring to the famous center-surround structure of ganglion cell receptive fields (Figure 1B), in particular in the simplified views of general neuroscience textbooks and in research that considers processing in downstream brain regions. This is perhaps not surprising, given the long and successful history of research on receptive fields, from their application as spatial filters to predict responses to moving light spots [14] to their utility for efficient stimulus coding [15].

However, many aspects of retinal stimulus encoding cannot be explained by stimulus filtering with center-surround receptive fields. The perhaps best-known example are the so-called 'Y-type

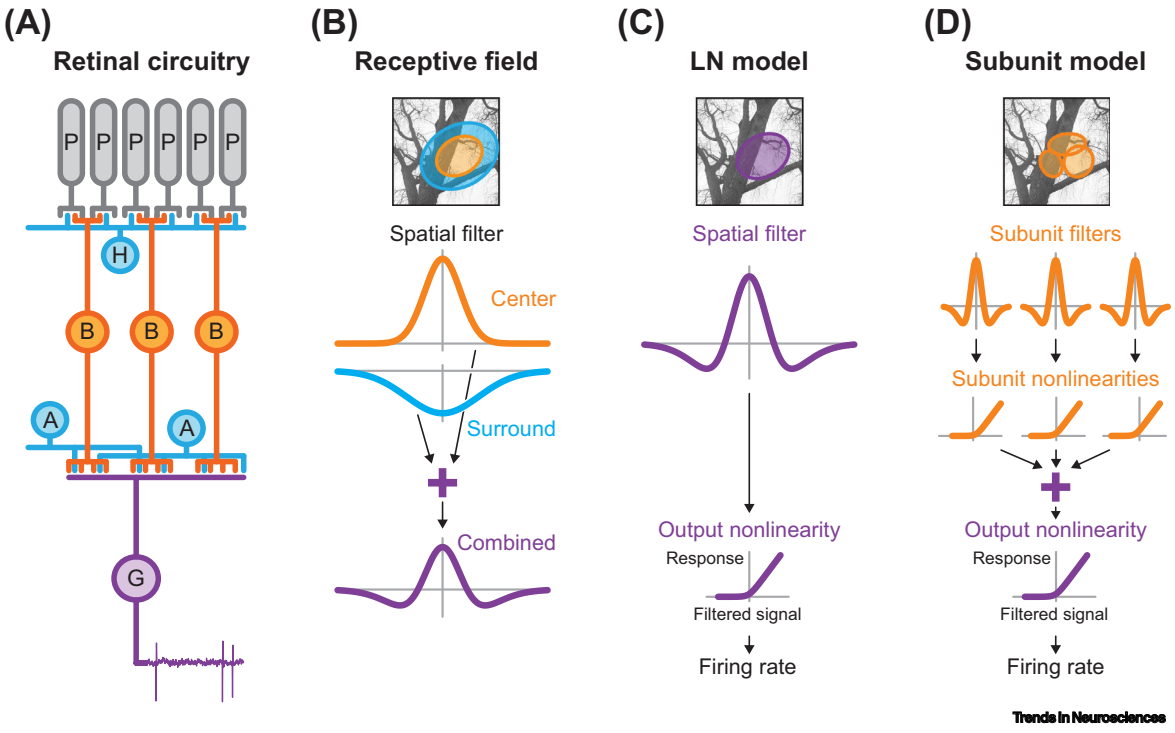


Figure 1. Retinal circuitry and standard models of spatial processing. (A) Schematic retinal circuitry, comprising photoreceptors (P), bipolar cells (B), two types of inhibitory interneurons [horizontal (H) and amacrine (A) cells], and a ganglion cell (G). Visual information propagates from top to bottom. The output is the spiking response of the ganglion cell, as indicated here by an extracellularly recorded voltage trace. (B) Ganglion cell center-surround receptive field, depicted here as a difference-of-Gaussians model. The spatial filter that comprises the receptive field contains a positive 2D Gaussian function for the center (orange, shown here as a 1D Gaussian cross-section) and a concentric, but broader negative Gaussian function for the surround (blue). The filter signals, obtained as weighted sums of pixel values, are added, yielding the combined receptive-field activation. Center and surround are attributed to excitatory bipolar cells and inhibitory interneurons, respectively, as indicated by the corresponding colors. Top: illustration of center and surround as 2D outlines overlaid on a visual stimulus. (C) Spatial linear–nonlinear (LN) model. First, a spatial filter, representing the receptive field of the ganglion cell, is applied to the visual stimulus, corresponding to a weighted summation of stimulus pixel values. The filtered signal is subsequently passed through a nonlinear transformation to produce the response of the ganglion cell, measured, for example, as the evoked firing rate. (D) Subunit model. Multiple subunits act as spatial filters with subsequent nonlinearities. Their responses are summed in a weighted manner and passed through a final nonlinearity to produce the response of the ganglion cell (e.g., firing rate).

ganglion cells', first found in the cat retina [16], which respond strongly (and with frequency doubling) to contrast-reversing spatial gratings with no net change in illumination over the receptive-field center or surround. This demonstrates that activation and deactivation from brightening and darkening do not cancel out for these cells but are combined nonlinearly over space.

More general analyses of whether receptive-field substructure beyond the center-surround shape is relevant under different stimulus scenarios typically require a model-based approach. The receptive field, often extended to contain a temporal dimension, is then used to filter the applied visual stimulus to predict responses for comparison with actual recorded data. In its most prominent model version, receptive field-based filtering is followed by a nonlinear transformation of the filtered signal, which can prevent negative predicted firing rates as well as implement thresholding and saturation of responses. This is known as the linear–nonlinear (LN) model (Figure 1C), which has become, in some respects, a standard model for retinal ganglion cell responses and forms the backbone of many data analysis approaches and extended computational models.

Yet, the receptive field-based LN model is often not a good predictor of responses for many ganglion cells, in particular when natural stimuli are considered. Under flashed natural images, responses of many ganglion cells in both mouse [17] and salamander [18] deviate from LN model predictions. Furthermore, these deviations systematically depend on the spatial contrast of the images inside the receptive field. Similarly, in the macaque retina, models based on receptive-field filtering fail to accurately predict responses of parasol cells to natural movies [19]. OFF parasol cells, in particular, are sensitive to the fine spatial structure of natural stimuli, which is not captured by the LN model [20].

Receptive-field subunits and their functional relevance

The primary model extension to go beyond receptive-field filtering and incorporate sensitivity to spatial structure within the receptive-field center is to subdivide the center into smaller subunits, giving rise to a 'subunit model'. Each subunit acts as an independent spatial filter, the signal of which is nonlinearly transformed (e.g., rectified) before summation into the integrated ganglion cell activation (Figure 1D). The subunits are generally believed to correspond to the bipolar cells that provide the excitatory input to ganglion cells, because this excitation alone already displays nonlinear spatial integration [21]. Signal transmission between bipolar and certain ganglion cells in the mouse retina has indeed been found to be nonlinear [22,23]. However, this does not exclude that, for other types of ganglion cells or under different illumination conditions, subunits might reflect other (e.g., amacrine) cells or correspond to larger groups of bipolar cells, perhaps electrically coupled [24]. In addition, nonlinearities at the input stage to bipolar cells [25] and in photoreceptor signaling [26,27] may contribute to nonlinear spatial integration upstream of bipolar-cell subunits.

Sensitivity to finely structured stimuli and motion signals

The subunit structure of receptive fields helps explain the different functional properties and specific computations of retinal ganglion cells [28]. Most prominently perhaps, rectified subunit signals mediate sensitivity to high spatial frequencies and to small objects below the scale of ganglion cell receptive-field centers [16,23,29]. This is achieved by communicating local changes in illumination (Figure 2) even without net luminance changes across the receptive field. The sensitivity to fine spatial patterns also leads to a joint encoding of luminance and spatial-contrast information [17,18,30] and may contribute to monitoring the proper focusing of images onto the retina [31].

Furthermore, subunits are particularly important for motion processing. Subtle object or texture motion that changes the illumination pattern over individual subunits can then trigger responses even when a linear receptive field would not be activated. This may contribute to the perception

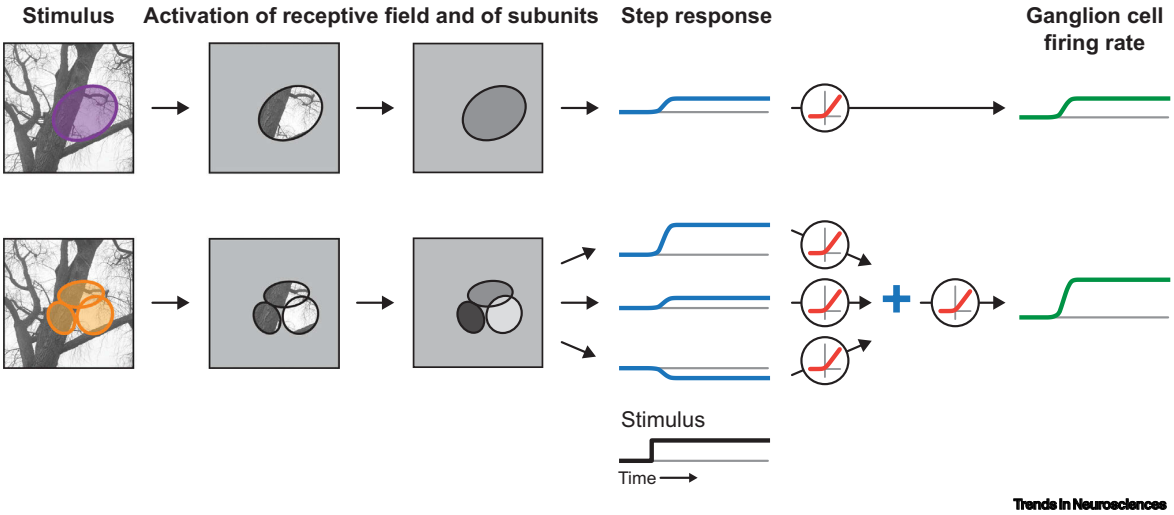


Figure 2. Schematic of ganglion cell processing with and without subunits. Top: linear–nonlinear (LN) model, with receptive field as a spatial filter. Bottom: subunit model, with multiple subunits as filters. The receptive field and the subunits evaluate an image in confined regions (left) to extract mean light-intensity signals. In the subunit model, the small size of the subunits can resolve spatial structure of the stimulus within the receptive field of the ganglion cell. The filtered signals generate activation signals, depicted here as step-response curves (blue, center), corresponding to image presentation after homogeneous background light intensity. Positive activation here corresponds to darkening in the spatial-filter region (OFF-cell responses). The subunit responses are rectified individually before summation, which here leads to a stronger ganglion cell response (green, right) than in the LN model.

of contrast-mediated motion signals ('second-order motion') [32] and to detecting subtle fixational eye movements [33,34]. Combining rectification of bipolar-cell subunit signals with gap-junction coupling between bipolar cells can further enhance sensitivity to specific motion signals [24,35], which is hypothesized to enhance the information about future locations of continuously moving objects [36]. Even for ON–OFF direction-selective ganglion cells, where much of the computation of directionality occurs in the presynaptic starburst amacrine cells, the crucial integration of excitatory and inhibitory inputs occurs locally along the dendritic tree of the ganglion cell, providing a subunit architecture to the motion processing [37,38].

Spatial nonlinearities of inhibitory signals

For some ganglion cells, the subunit-mediated sensitivity to texture motion also pertains to inhibitory signals from the receptive-field periphery. This can lead to response suppression under global image motion and thereby specific sensitivity to local object motion in object motion-sensitive ganglion cells described in the salamander and rabbit retina [9] and in the so-called 'W3 ganglion cells' of the mouse [10]. For other cells, the motion sensitivity in the receptive-field periphery results in disinhibition and increased responsiveness after a global image shift [39]. This exemplifies that nonlinear spatial integration and subunit models also relate to inhibitory signaling in the retina and thereby have a role in gating visual information.

Spatial integration in the suppressive surround of salamander ganglion cells appears to be similarly nonlinear as in the receptive-field center, albeit on a larger spatial scale [40], suggesting that amacrine cells might provide relevant nonlinearities. In mouse retina, the possibility to optogenetically stimulate specific amacrine cell types was recently used to directly probe signal transmission to ganglion cells. This revealed strong rectification of the functionally crucial inhibition that direction-selective ganglion cells receive from starburst amacrine cells [41,42], whereas other amacrine cells were found to transmit signals linearly [43,44].

With inhibition being prevalent not only in the receptive-field surround, but also in the center, one may wonder why the subunit models discussed above are often considered without inhibitory elements. One answer is that local inhibitory effects can be part of the subunits, so that the latter correspond to bipolar cells together with local modulations by inhibitory amacrine and horizontal cells. These may add a surround component to the subunits, modulate their temporal signaling, affect the nonlinearity of signal transmission to the ganglion cell, or shape the adaptation characteristics. The view of subunits as 'effective bipolar cells', the processing of which also comprises inhibitory effects, is reinforced by the finding that many local inhibitory interactions occur directly at the bipolar cell synaptic terminal [45,46], where they fundamentally shape the temporal and spatial characteristics of the subunit signal received by the ganglion cell [47].

An example of local inhibition fundamentally shaping the functional characteristics of a ganglion cell is provided by the widely studied transient OFF alpha ganglion cell of the mouse retina. Even though flashed natural images expose little deviation from linear spatial integration for these cells [17], stimuli with more complex spatiotemporal structure can reveal nonlinear local processing by crossover inhibition, which couples the ON and OFF signaling pathways in the retina [48,49]. Here, it supplies ON-pathway inhibition under light increments to an OFF-type ganglion cell. This is thought to let the cell differentiate between looming and translatory motion of a dark object, given that responses are suppressed for the latter because of inhibition triggered by the brightening at the trailing edge of the object [8]. For saccade-like image transitions, the local crossover inhibition, in conjunction with serial inhibition, has also been suggested to underlie a particular sensitivity to recurring image patterns, which has been observed in these cells [50].

Accessing the subunits for studying retinal computations

The examples discussed above show how inhibition can suppress or gate subunit signals in a selective manner, implementing computations that are not captured by the shape of the center-surround receptive field. This suggests viewing the layout of bipolar-cell subunits as scaffolding for local signal processing, setting up the spatial layout for operating on the visual input. The types and dynamics of the operations, from simple nonlinear transformations to complex inhibitory interactions, then determine the computation performed by the ganglion cell and, thus, its potential function for visual processing. Therefore, access to the subunit layout could be an important step for analyzing the details of retinal computations either through model-based data analysis or by targeting individual subunits for stimulation in experiments.

However, experimental complications arise from the relative inaccessibility of bipolar cells. These cells are buried in the middle layers of the retina; they are mostly nonspiking and receive much of their functionally relevant inputs at their synaptic terminals. Some progress has been made by applying fluorescent reporters to monitor vesicle fusion [51,52] or glutamate release [53] at the bipolar cell terminal. This has allowed observations of rectified transmitter release from certain bipolar cells [22] and characterizations of the functional diversity of bipolar cell types and of the role of inhibition in shaping this diversity [47,54].

Nonetheless, for analyzing how ganglion cells integrate bipolar cell input, alternatives to direct measurements of bipolar cell signals are desirable. A promising approach arises from currently developed methods for model-based inference of bipolar-cell subunits via the relatively easy-to-record ganglion cell activity. By aiming at identifying bipolar cell receptive fields through their effects on the output of the retina, this approach may provide a 'virtual microscope' for studying the layout of bipolar cells and their functional connections to ganglion cells.

Inference of subunits with multilayered models

LNLN models

Perhaps the most straightforward approach to identifying and characterizing receptive-field subunits is to set up a subunit model and fit the parameters to recorded data. The model typically comprises two stacked layers of LN model-like components, one for the subunits and one for the subsequent signal summation and final transformation (Figure 1D), and, therefore, is also called 'LNLN model'. Although the general idea of subunit models dates back to the original observation of nonlinear spatial integration [16], methods for identifying concrete subunit layouts from experimental data emerged only relatively recently. One approach is to use anatomical knowledge to structure the model, for example, by determining subunit positions and weights according to the reconstructed dendritic tree of a ganglion cell. This has been used to demonstrate that the concrete subunit layout matters for accurate response predictions [23]. Similarly, prior assessment of the spatial layout of the relevant photoreceptors, as is possible in the peripheral primate retina [55], has been exploited to constrain LNLN models [56].

However, without knowledge of the anatomical details for a given cell, the difficulty of obtaining a fitted LNLN model lies in finding a workable parameterization, particularly since the number and locations of subunits are unknown. To reduce model complexity, some studies have focused on a single spatial dimension by applying stripe-like visual stimuli and thereby successfully fitted LNLN models to ganglion cells recorded from salamander retina [57,58]. The extracted subunits (Figure 3A) displayed response characteristics and receptive fields similar to bipolar cells that had been independently recorded [57,58]. These findings support the notion that fitting subunit models can be used to infer bipolar cell properties. If the number of parameters is further reduced by assuming regular spacing and identical shapes of subunits, the LNLN model can even be extended to include adaptation-like feedback components [58]. A different application of LNLN model fits is to separate the temporal characteristics of ON-type and OFF-type inputs or of excitatory and suppressive receptive-field components, as has been shown for mouse retinal ganglion cells [59]. Combining the modeling of such different input types with a spatial layout of subunits will be an exciting, yet challenging future direction.

Convolutional neural networks

Fitting complex parameterized models received a push with the advent of deep learning-inspired artificial neural networks. In particular, convolutional neural networks enjoy current popularity, owing to their versatility and computational efficiency. The convolutional operations in these networks (applying the same filtering at all locations of the input space) are reminiscent of how an individual bipolar cell type (or another retinal neuron type) covers visual space with its receptive fields, applying nearly the same signal processing at different spatial locations. Thus, not surprisingly, retinal ganglion cells were among the first applications of convolutional neural networks to perform data-driven circuit modeling [60–62].

Using two convolutional layers, these networks can outperform standard receptive-field models, such as the LN model, in predicting responses of salamander ganglion cells to new stimuli, because the multilayered structure captures essential nonlinear substructure of ganglion cell receptive fields. Moreover, hidden units can match certain properties of bipolar and amacrine cells, such as localized center-surround receptive fields [60,61] (Figure 3B). When trained on salamander ganglion cell responses to natural stimuli, the models are also able to reproduce specific response features beyond what can be explained by spatiotemporal filtering [60,62]. However, despite these agreements, the derived networks have so far remained somewhat abstract, with no direct correspondence to circuit structure or subunit layout.

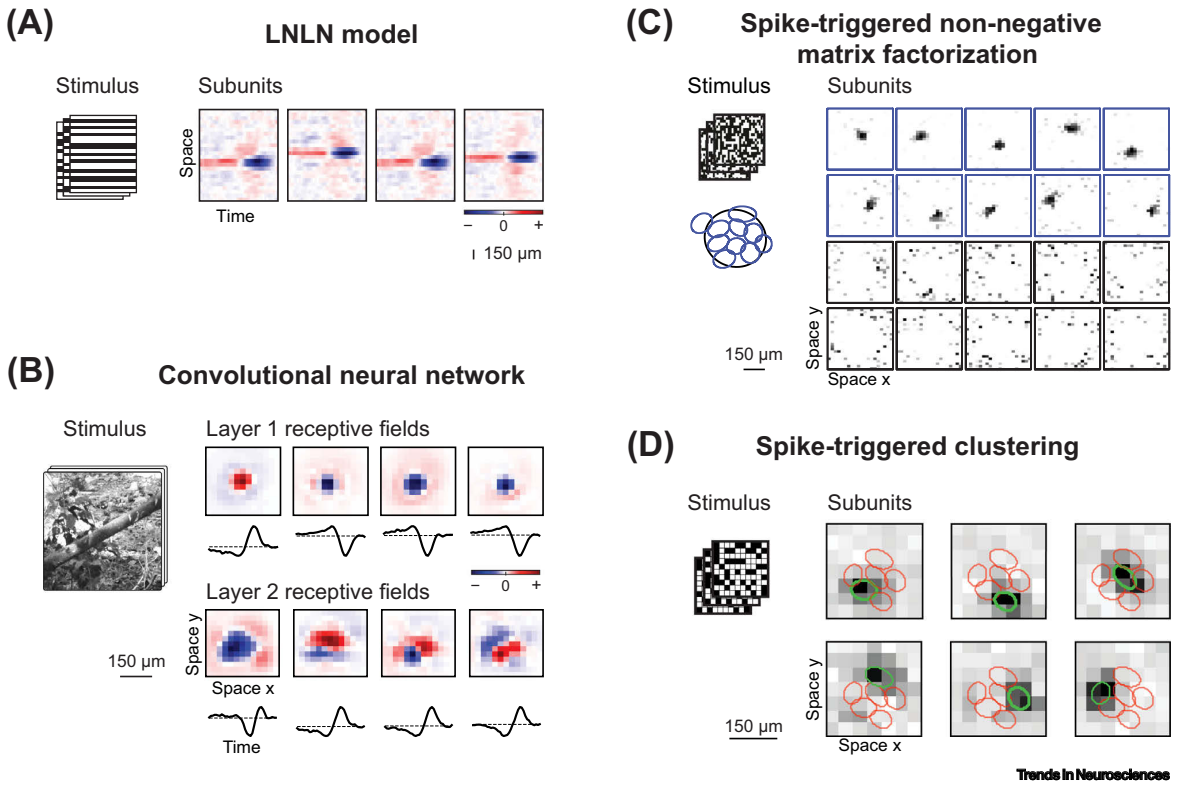


Figure 3. Examples of computationally inferred subunits from different methods. (A) Spatiotemporal subunits with one spatial dimension obtained from fitting a linear–nonlinear–linear–nonlinear (LNLN) model to recorded salamander ganglion cell responses under stimulation with flickering stripes. Subunits in the example are OFF-type, with localized preference for negative contrast (blue), preceded and flanked by preference for positive contrast (red), corresponding to biphasic temporal filtering and a spatial antagonistic surround, respectively. (B) Spatiotemporal subunits, corresponding to receptive fields of convolutional filters from a two-layer convolutional neural network, fitted to recorded salamander ganglion cell responses under natural stimulation. Receptive fields are separated into a spatial component (red/blue: positive/negative contrast preference) and a temporal sensitivity trace below. (C) Spatial subunits obtained by analyzing salamander ganglion cell responses to spatiotemporal white noise with spike-triggered non-negative matrix factorization. The obtained spatial modules (white indicates zero; darker shades indicate more positive entries) can be separated into true subunits (blue module frames) and noise modules (black frames) according to the spatial autocorrelation of the modules. The recovered subunits are fitted by 2D Gaussian profiles (blue ellipses, corresponding to 1.5-sigma contours), which roughly tile the Gaussian fit to the receptive field (black ellipse). (D) Spatial subunits obtained by spike-triggered clustering, here with six clusters, with data recorded from a macaque ganglion cell under spatiotemporal white-noise stimulation. Adapted from [57] (A), [60] (B), [68] (C), and [70] (D).

Inference of subunits by statistical analyses of successful stimuli

Structure of spike-triggered stimulus ensembles

An alternative to fitting a complete subunit model is to extract subunits from statistical analyses of the stimulus–response relationship of a ganglion cell. The basic idea is to record ganglion cell responses to spatially structured stimuli and investigate which spatial patterns make the cell fire. The set of these successful stimuli is often called the spike-triggered stimulus ensemble (STE), and the search for subunits then amounts to identifying the underlying structure of the STE (Figure 4). This often takes the form of a dimensionality-reduction analysis [63], for which a wide arsenal of general techniques is already available. However, to avoid confounding structure in the STE from prior correlations in the applied visual stimuli, these techniques typically rely on stimuli with white-noise statistics, which may not drive cells strongly if a finely structured layout is used, thus requiring long recordings.

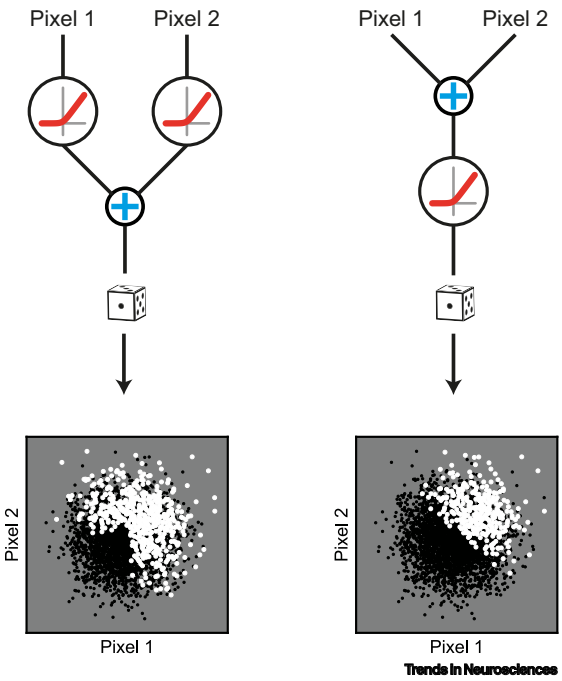


Figure 4. Effect of subunits on the spike-triggered stimulus ensemble (STE). Whether two stimulus pixels are in the same subunit influences the joint distribution of their values in the STE, as exemplified here by two simple models (top). Stimuli are normally distributed contrast values of two pixels. In the first model (left), the values are individually rectified and then summed into a spike probability, representing pixels in distinct subunits. In the second model (right), the two pixel values are summed before rectification, representing pixels within a subunit. Spikes are generated by a binary random process. For each model, scatter plots at the bottom display the set of applied stimuli (black dots) and stimuli that generated spikes, that is, the STE (white dots). The two STEs exhibit different distributions, which forms the basis for the inference of subunits with statistical methods. In particular, the STE of the first model contains a wider range of stimuli with opposing contrast values.

One common technique for extracting relevant dimensions of the STE is spike-triggered covariance analysis, which is essentially a form of principal component analysis of the STE [64,65]. However, when applied to data recorded from the salamander retina under spatiotemporal white noise, the resulting filters are generally not localized subunits, but rather span the entire receptive field, reminiscent of a Fourier decomposition [57,66–68]. Appropriate mixing of these filters might help recover localized components [67], but may be limited by noise in the originally identified filters and, to our knowledge, this has not yet been thoroughly tested on experimental data.

Spike-triggered non-negative matrix factorization and clustering

A more direct identification of subunits can be achieved by analyzing the STE with spike-triggered non-negative matrix factorization (STNMF) [68]. Here, multiple spatial filters are extracted under the constraint of having no negative elements. Non-negative matrix factorization is known to provide parts-based representations of complex data [69], which helps here to uncover local structure in receptive fields. In fact, subunit filters obtained for salamander ganglion cells were found to be localized (Figure 3C) and to match the receptive fields of simultaneously recorded bipolar cells [68], supporting the idea that bipolar cell layouts can be inferred from ganglion cell recordings. Note, however, that subunits are derived with no negative entries, and any structure with opposing sign, such as an antagonistic surround, is not recovered by this method.

A related approach is to find the elementary stimulus patterns that contribute to generating spikes by clustering analyses of the STE. The idea is that activation of one or potentially a few subunits must have been sufficiently strong to trigger a spike, while other subunits remained subthreshold. The centroids of a soft clustering of the STE from macaque OFF parasol cells indeed display

spatially localized structure as expected for receptive-field subunits [70] (Figure 3D). Moreover, the spike clustering can be viewed as a step in the maximum likelihood fit of an LNLN model and, thus, directly yields a subunit model for predicting responses to spatially structured stimuli.

Regardless of the method, inferred subunit layouts so far typically show subunits that more or less tile the receptive field with little overlap. This is consistent with excitatory inputs dominated by a single type of bipolar cell, as appears to be the case for some ganglion cells [71]. Yet, it raises the question whether layouts with two (or more) bipolar cell types, each arranged in its own mosaic pattern, could also be handled. Analyses of model simulations suggest that STNMF may, in principle, be able to extract subunits with considerable overlap [68], but evidence for this from experimental data is still lacking. One possibility is that extracted subunits reflect the bipolar cell type that provides the strongest input. It might then be interesting to explore whether stimuli can be modified (e.g., via ambient light intensity or spatial, temporal, or chromatic characteristics) to switch the dominant input. This might allow reconstructing different bipolar cell mosaics in different stimulation contexts.

Temporal dynamics of subunits

Functions and mechanisms of local and global adaptation

The fact that ganglion cells integrate over many bipolar cells not only breaks up spatial integration into subunits, but is also important for understanding the functional consequences of adaptation. For example, retinal ganglion cells adapt their sensitivity to ambient light intensity and to visual contrast. Conceptually, adaptation could occur locally, at the level of individual subunits before their signals are integrated by the ganglion cell, or globally after stimulus integration, at the level of the entire ganglion cell receptive field (Figure 5). The functional relevance becomes clear, for example, when considering a small object somewhere in the receptive field. Global, but not local, adaptation then reduces sensitivity for a second object irrespective of its location. Furthermore, for a moving object, global adaptation is already triggered when entering the receptive field and thereby reduces responses to the subsequent path through the receptive field, a mechanism that has been suggested to counteract temporal delays of signal transduction for moving objects ('motion anticipation'; Figure 5A) [72]. By contrast, local adaptation of individual subunits maintains sensitivity at other locations within the receptive field or for novel objects moving along different paths (Figure 5B,C). Thus, whether adaptation occurs globally on the spatial scale of the entire receptive field or locally for individual subunits is crucial for understanding the function of a ganglion cell in response to dynamic spatial stimuli.

For retinal contrast adaptation, mechanisms that could support local or global adaptation have both been identified. A primary contribution to contrast adaptation is thought to occur at the bipolar-to-ganglion cell synapse, where synaptic gain is altered by synaptic depression [73,74] and inhibition onto the synaptic terminal [75,76]. Computational models of vesicle depletion or presynaptic inhibition have successfully captured much of the adaptive dynamics of excitation in certain ganglion cells [76,77]. Owing to the local origin at the synaptic terminal, this adaptation might be thought to be local and restricted to individual subunits. Yet, presynaptic inhibition need not be locally confined and could act on multiple subunits. In addition, even local synaptic depression may influence sensitivity throughout the receptive field of the ganglion cell if the reduction in vesicle release leads to a postsynaptic hyperpolarization that affects the entire ganglion cell. Thus, it remains an open question whether gain changes at the bipolar cell terminal contribute to local or global adaptation. A clearer contribution to local adaptation comes from contrast-induced gain changes in the membrane potential of bipolar cells themselves. This has been observed in the salamander retina [78], although not for all bipolar cells [79]. Conversely, unambiguous global adaptation may follow from mechanisms triggered by the activity of the

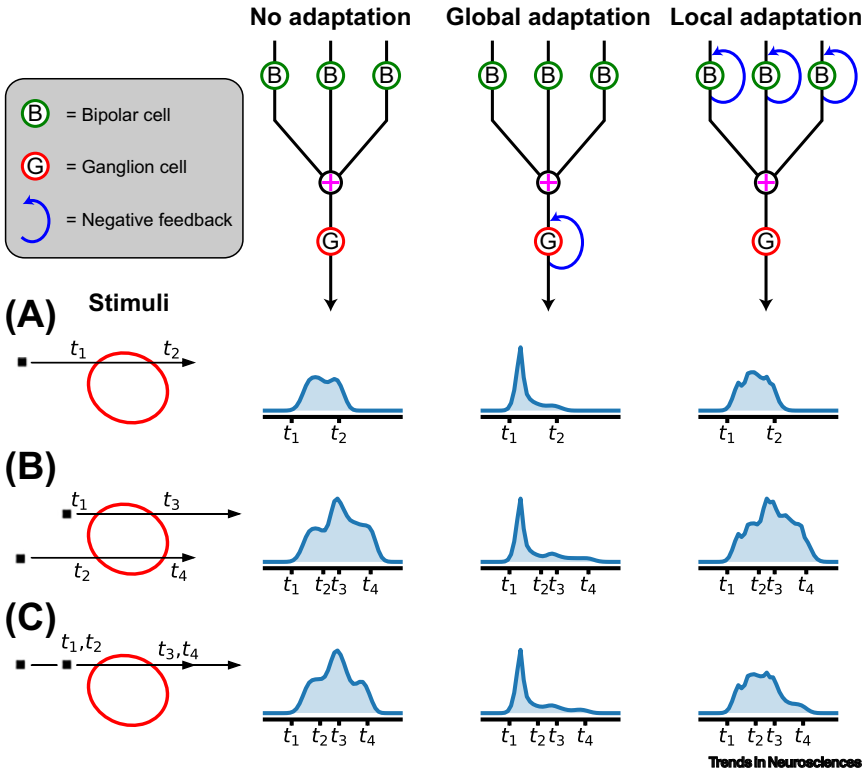


Figure 5. Functional consequences of local and global adaptation. Simulated responses to different motion stimuli (left) of three subunit models with no adaptation, global adaptation, and local adaptation (top). All models simulate multiple bipolar cells (only three are depicted) connected to a ganglion cell. Bipolar cells are simulated by applying a spatial Gaussian filter and a monophasic temporal filter, followed by a rectification. Ganglion cells sum the results and apply another rectification yielding a firing rate as output. Global adaptation is implemented as multiplicative feedback by applying an exponentially decaying filter to the model output, followed by a sigmoidal nonlinearity. Local adaptation occurs through analogous multiplicative feedback on the bipolar-cell level. Simulated firing rates (blue) are shown in response to objects moving through the receptive field, with the actual configurations shown on the left. Black squares and arrows show objects and their trajectories. Red ellipses correspond to the 1.5-sigma contour of the receptive field. Times t_1 to t_4 indicate when the objects enter the first or leave the last subunit, respectively, as indicated in the plots. When a single object moves through the receptive field (A), global adaptation leads to a rapid response decay, causing the peak firing rate to occur well before the object is halfway through the receptive field. This has been associated with the phenomenon of motion anticipation [72] and is not apparent for the models with no adaptation or with local adaptation. By contrast, when two objects successively move through the receptive field, local adaptation leads to distinct responses, depending on whether the two objects move along different trajectories (B) or along the same trajectory (C), yielding reduced response amplitude for the latter. By contrast, the models with no or global adaptation each display similar responses for these two scenarios.

ganglion cell itself, such as inactivation of sodium channels [80] or recruitment of potassium currents [81]. Which mechanisms dominate for a given cell and whether this depends on stimulus context, such as light level, is largely unknown.

Experimental tests of local and global contrast adaptation

Functionally, only a few studies have probed the spatial scope of contrast adaptation so far. A study of cross-adaptation in rabbit retina, which used switching stimulation from one small location to another within the receptive field of a recorded ganglion cell, showed transient increases in

firing rate after each switch [82]. This can be interpreted as being caused by local adaptation, since the new stimulus location after the switch is not yet adapted. However, the fast timescale of the transient firing rate increases might also be explained without any adaptation. Instead, nonlinear spatial integration can boost activity during the brief transition time when activation from one location is ramping up while activation from the other has not yet fully decayed [83].

Local adaptation can also be observed in the object motion-sensitive cells discussed above, which adapt to continuous object motion, but transiently regain sensitivity when the object motion is detected by different regions inside the receptive-field center [84]. Other studies of ganglion cells in salamander [83] and mouse retina [85] have observed that changes in temporal filtering and sensitivity following locally restricted switches in visual contrast can, for different cells, occur either locally, confined to the region of the contrast switch, or globally over the entire receptive field (Figure 6). Thus, as with linear and nonlinear spatial integration, the occurrence of local and global contrast adaptation likely depends on cell type.

A different example of spatially structured adaptation has been found in the context of contrast sensitization [52,86,87], which describes the transient sensitivity increase after exposure to high contrast (as opposed to the more classical adaptive sensitivity decrease). For some sensitizing ganglion cells, the occurrence of either sensitization or classical adaptation depends on stimulus location. Stimuli near the midpoint of the receptive-field center cause adaptation and more peripheral stimuli elicit sensitization, which is thought to result from adaptation in excitatory as well as inhibitory subunits within the receptive-field center [88].

Concluding remarks

Retinal function goes well beyond stimulus filtering by the center-surround receptive fields of ganglion cells. Nonlinear stimulus integration within the receptive field, local signal gating by inhibitory interactions, and local adaptation endow the receptive field with a substructure that shapes visual stimulus encoding. The receptive-field substructure also provides the substrate for different specific computations in the retina, which rely on local analyses of light patterns. Thus, investigating the substructure of ganglion cell receptive fields is important for capturing how the retina encodes natural stimuli and for understanding the circuit implementation of specific retinal functions.

Reliable inference of ganglion cell subunits may prove a valuable tool for circuit analyses in the retina. It provides indirect access to the properties of retinal bipolar cells, supplying information about their receptive-field attributes and nonlinear transformations as well as about their functional connectivity to ganglion cells. For example, obtaining the same subunit contour from two or more ganglion cells indicates shared input from the same bipolar cells [68]. Furthermore, identified subunit layouts will be useful in experiments to optimally place probe stimuli to further study the nonlinearities and dynamics of stimulus encoding and in computational analyses to construct and constrain encoding models.

As methods for analyzing receptive-field substructure mature, it will be interesting to compare their applicability to retinas from different animal models and to evaluate whether they reveal differences in retinal circuit organization. Given that many aspects of the basic retinal layout are similar across vertebrates, one should expect that the general methodology discussed here is not restricted to particular species. For example, nonlinear spatial integration is a general observation, and evidence that receptive-field subunits correspond to bipolar cells has also come from various vertebrates, including guinea pig [21], salamander [68], mouse [23], and macaque [56]. Nonetheless, details differ. For example, in the primate retina, the fovea is a region of specialized connectivity and signal processing characteristics [89,90]. Beyond the fovea, nonlinear spatial

Outstanding questions

How do the different methods developed for identifying subunits of retinal ganglion cells compare? Can different methods be combined to arrive at a consensus of inferred subunit layouts? How robust are the different methods across different ganglion cell types, different animal species, and different illumination conditions?

What is missing in subunit models with nonlinear spatial integration to fully capture ganglion cell encoding of flashed natural images, for which temporal dynamics may be ignored?

The implementation of specific computations by retinal circuits, such as the detection of different types of motion signals, is often explained through conceptual, proof-of-principle models. How can such conceptual models be combined with concrete subunit layouts of individual cells to explore the circuit–function connection in more detail?

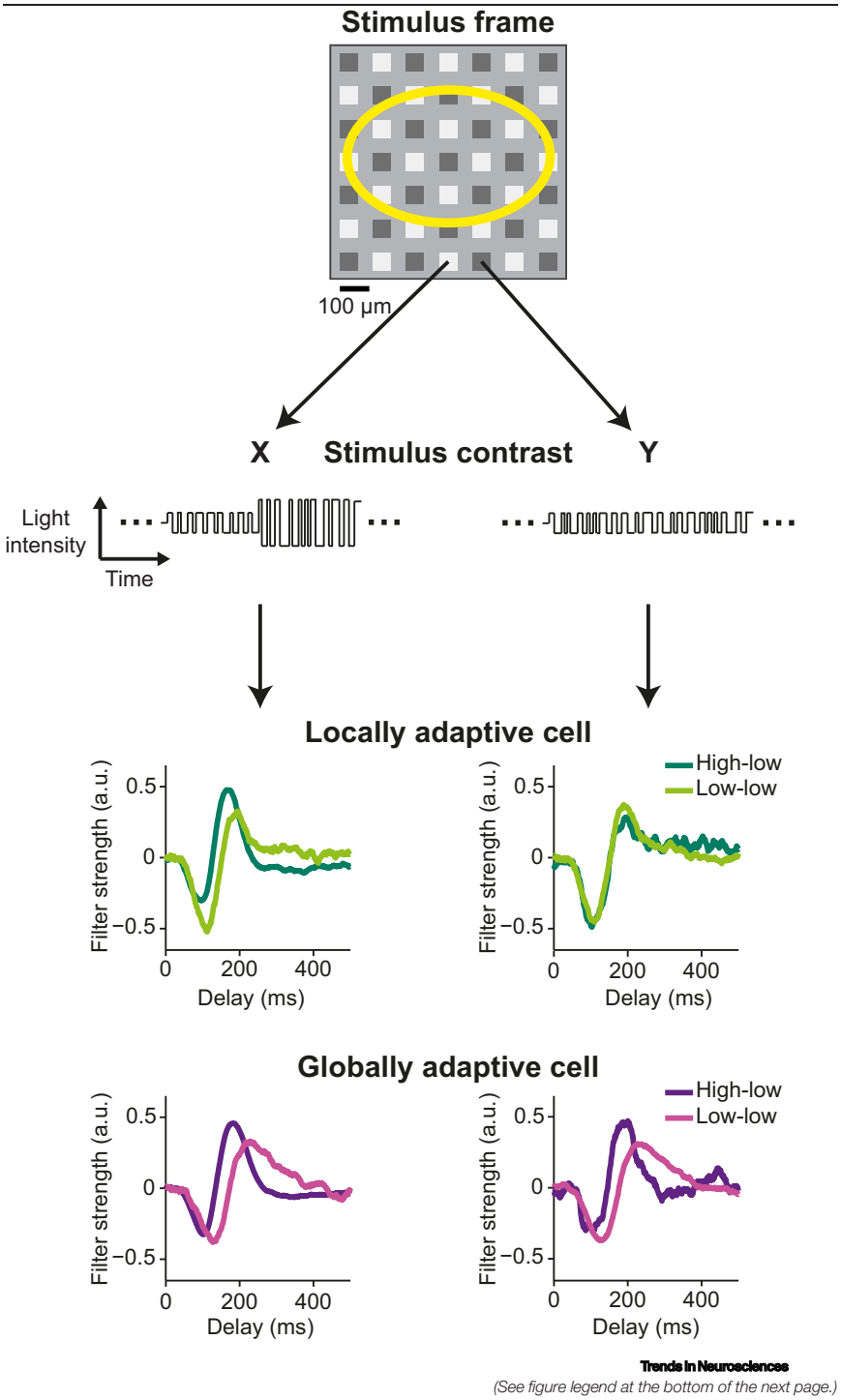
Nonlinear signal transformations may already occur in photoreceptors or at the inputs to bipolar cells. How important are these early nonlinearities for stimulus encoding by ganglion cells? How could they be incorporated into stimulus–response models?

Are subunits a relevant concept for the ganglion cell receptive-field surround? How could one extract information about surround subunits from experimental data? Would these subunits correspond to bipolar cells or to amacrine cells?

What makes some ganglion cells adapt locally to visual contrast and others globally? What mechanisms are involved and what specific functional consequences arise?

How can inhibitory interactions be integrated into subunit models and what would be the ways to obtain interaction parameters from experimental data?

After fitting abstract network models, such as convolutional neural networks, to retinal data, can the obtained network structure be related to actual circuit components of the retina?



(See figure legend at the bottom of the next page.)

integration appears to be the norm in the salamander retina [91], whereas cell types with more linear characteristics can be found in mammalian retinas, as shown in cat [16], mouse [17], and macaque [20,92]. Similarly, the spatial scope of contrast adaptation displays greater diversity in mouse [85] than in salamander [83]. Furthermore, it seems likely that the basic scaffolding of bipolar-cell subunits supports different species-specific functions through different nonlinear transformations and inhibitory interactions.

Moreover, questions of receptive-field substructure also pertain to other visual areas or sensory systems. For example, subunit models have been applied successfully to primary visual cortex [93–95] and to motion-processing areas of the primate visual pathway [96,97]. In the auditory system, complex dependencies of neural responses on the spectrotemporal structure of acoustic stimuli have led to related models with nonlinear substructure in the spectrotemporal receptive fields of the neurons [98–100]. Subunit identification here might be a promising endeavor to relate this substructure to neural circuitry.

Despite the recent advances discussed here, the challenge of inferring receptive-field substructure in the retina is still far from solved (see Outstanding questions). Future developments may profit from the conceptual overlap that exists with current efforts to analyze large neural ensembles. Finding structure in high-dimensional population activity mirrors finding structure in spike-generating visual stimuli. Thus, similar methods for identifying a manageable number of latent variables may apply, such as dimensionality-reduction techniques [63,101]. For example, matrix factorization approaches are used equivalently for both types of problem [68,96,102–104].

In addition, the rapid developments in machine learning and artificial neural networks promise new approaches to study neural signal processing [105,106]. An enticing question here is how these approaches can be harnessed to go beyond subunits of ganglion cells and study inhibitory interactions in the retina. This could help tackle one of the biggest remaining mysteries about the retina, namely the functions and interactions of the vast set of inhibitory amacrine cells [2,107]. Incorporating such complexity into generic ganglion cell models, amenable for data-driven parameter optimization, will be a considerable challenge. However, recent computational advances may help deal with the resulting model intricacies.

Incorporating large populations of recorded ganglion cells into a single model with shared presynaptic circuitry, for example, will help constrain parameter estimation [61,108]. Further constraints can come from detailed anatomical knowledge, such as the morphology of individual cells [23] or the statistics of neuronal connectivity [109–111]. Derived neural network models can then be used for *in silico* analyses, for example, by extracting maximally effective stimuli [112] or by generating mechanistic hypotheses for specific computations through model reduction techniques [62]. For models of arbitrary structure, combining model simulations with machine-learning algorithms can help with parameter optimization [113]. These and other ongoing developments of computational approaches to infer structural and functional components of neural networks promise new insights into the nonlinear and dynamic substructure of retinal receptive fields.

Figure 6. Measurements of local and global contrast adaptation. Shown are the temporal stimulus filters of a locally adaptive and a globally adaptive ganglion cell, recorded in mouse retina under local switches in temporal contrast. The stimulus comprised locations X and Y, each stimulated homogeneously and independently by binary white noise (top: sample stimulus frame, yellow ellipse sketches an exemplary 3-sigma receptive-field contour of a ganglion cell). At locations Y, contrast remained constant, whereas at locations X, contrast changed every 40 s between a low (low-low condition) and a high (high-low condition) value. For the sample cell with local adaptation, the temporal filter for locations X changed with the contrast at X (left), but the filter for Y remained unaffected (right). By contrast, for the sample cell with global adaptation, the filters for both sets of locations displayed similar adaptive changes. Adapted from [85].

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft [DFG, German Research Foundation; project IDs 154113120 (SFB 889, project C01) and 432680300 (SFB 1456, project B05)] and by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement number 724822).

Declaration of interests

The authors declare no competing interests.

References

1. Schwartz, G. (2021) *Retinal Computation*, Academic Press
2. Yan, W. *et al.* (2020) Mouse retinal cell atlas: molecular identification of over sixty amacrine cell types. *J. Neurosci.* 40, 5177–5195
3. Baden, T. *et al.* (2016) The functional diversity of retinal ganglion cells in the mouse. *Nature* 529, 345–350
4. Grünert, U. and Martin, P.R. (2020) Cell types and cell circuits in human and non-human primate retina. *Prog. Retin. Eye Res.* 78, 100844
5. Goetz, J. *et al.* (2021) Unified classification of mouse retinal ganglion cells using function, morphology, and gene expression. *bioRxiv* Published online June 11, 2021. https://doi.org/10.1101/2021.06.10.447922
6. Barlow, H.B. *et al.* (1964) Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *J. Physiol.* 173, 377–407
7. Wei, W. (2018) Neural mechanisms of motion processing in the mammalian retina. *Annu. Rev. Vis. Sci.* 4, 165–192
8. Münch, T.A. *et al.* (2009) Approach sensitivity in the retina processed by a multifunctional neural circuit. *Nat. Neurosci.* 12, 1308–1316
9. Ölveczky, B.P. *et al.* (2003) Segregation of object and background motion in the retina. *Nature* 423, 401–408
10. Zhang, Y. *et al.* (2012) The most numerous ganglion cell type of the mouse retina is a selective feature detector. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2391–E2398
11. Kuffler, S.W. (1953) Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.* 16, 37–68
12. Baden, T. and Osorio, D. (2019) The retinal basis of vertebrate color vision. *Annu. Rev. Vis. Sci.* 5, 177–200
13. Demb, J.B. and Singer, J.H. (2015) Functional circuitry of the retina. *Annu. Rev. Vis. Sci.* 1, 263–289
14. Rodieck, R.W. (1965) Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vis. Res.* 5, 583–601
15. Atick, J.J. and Redlich, A.N. (1992) What does the retina know about natural scenes? *Neural Comput.* 4, 196–210
16. Enroth-Cugell, C. and Robson, J.G. (1966) The contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol.* 187, 517–552
17. Karamanlis, D. and Gollisch, T. (2021) Nonlinear spatial integration underlies the diversity of retinal ganglion cell responses to natural images. *J. Neurosci.* 41, 3479–3498
18. Liu, J.K. *et al.* (2022) Simple model for encoding natural images by retinal ganglion cells with nonlinear spatial integration. *PLoS Comput. Biol.* 18, e1009925
19. Heitman, A. *et al.* (2016) Testing pseudo-linear models of responses to natural scenes in primate retina. *bioRxiv* Published online December 7, 2016. https://doi.org/10.1101/045336
20. Turner, M.H. and Rieke, F. (2016) Synaptic rectification controls nonlinear spatial integration of natural visual inputs. *Neuron* 90, 1257–1271
21. Demb, J.B. *et al.* (2001) Bipolar cells contribute to nonlinear spatial summation in the brisk-transient (Y) ganglion cell in mammalian retina. *J. Neurosci.* 21, 7447–7454
22. Borghuis, B.G. *et al.* (2013) Two-photon imaging of nonlinear glutamate release dynamics at bipolar cell synapses in the mouse retina. *J. Neurosci.* 33, 10972–10985
23. Schwartz, G.W. *et al.* (2012) The spatial structure of a nonlinear receptive field. *Nat. Neurosci.* 15, 1572–1580
24. Kuo, S.P. *et al.* (2016) Nonlinear spatiotemporal integration by electrical and chemical synapses in the retina. *Neuron* 90, 320–332
25. Schreyer, H.M. and Gollisch, T. (2021) Nonlinear spatial integration in retinal bipolar cells shapes the encoding of artificial and natural stimuli. *Neuron* 109, 1692–1706
26. Endeman, D. and Kamermans, M. (2010) Cones perform a non-linear transformation on natural stimuli. *J. Physiol.* 588, 435–446
27. Angueyra, J.M. *et al.* (2022) Predicting and manipulating cone responses to naturalistic inputs. *J. Neurosci.* 42, 1254–1274
28. Gollisch, T. and Meister, M. (2010) Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* 65, 150–164
29. Victor, J.D. and Shapley, R.M. (1979) The nonlinear pathway of Y ganglion cells in the cat retina. *J. Gen. Physiol.* 74, 671–689
30. Yu, Z. *et al.* (2022) Adaptation in cone photoreceptors contributes to an unexpected insensitivity of primate On parasol retinal ganglion cells to spatial structure in natural images. *Elife* 11, e70611
31. Mani, A. and Schwartz, G.W. (2017) Circuit mechanisms of a retinal ganglion cell with stimulus-dependent response latency and activation beyond its dendrites. *Curr. Biol.* 27, 471–482
32. Demb, J.B. *et al.* (2001) Cellular basis for the response to second-order motion cues in Y retinal ganglion cells. *Neuron* 32, 711–721
33. Greene, G. *et al.* (2016) Non-linear retinal processing supports invariance during fixational eye movements. *Vis. Res.* 118, 158–170
34. Hennig, M.H. and Wörgötter, F. (2007) Effects of fixational eye movements on retinal ganglion cell responses: a modelling study. *Front. Comput. Neurosci.* 1, 69–84
35. Manookin, M.B. *et al.* (2018) Neural mechanisms mediating motion sensitivity in parasol ganglion cells of the primate retina. *Neuron* 97, 1327–1340
36. Liu, B. *et al.* (2021) Predictive encoding of motion begins in the primate retina. *Nat. Neurosci.* 24, 1280–1291
37. Jain, V. *et al.* (2020) The functional organization of excitation and inhibition in the dendrites of mouse direction-selective ganglion cells. *Elife* 9, e52949
38. Murphy-Baum, B.L. and Awatramani, G.B. (2022) Parallel processing in active dendrites during periods of intense spiking activity. *Cell Rep.* 38, 110412
39. Jadzinsky, P.D. and Baccus, S.A. (2015) Synchronized amplification of local information transmission by peripheral retinal input. *Elife* 4, e09266
40. Takeshita, D. and Gollisch, T. (2014) Nonlinear spatial integration in the receptive field surround of retinal ganglion cells. *J. Neurosci.* 34, 7548–7561
41. Pottackal, J. *et al.* (2020) Receptoral mechanisms for fast cholinergic transmission in direction-selective retinal circuitry. *Front. Cell. Neurosci.* 14, 604163
42. Pottackal, J. *et al.* (2021) Computational and molecular properties of starburst amacrine cell synapses differ with postsynaptic cell type. *Front. Cell. Neurosci.* 15, 660773
43. Pottackal, J. *et al.* (2021) Photoreceptive ganglion cells drive circuits for local inhibition in the mouse retina. *J. Neurosci.* 41, 1489–1504
44. Park, S.J.H. *et al.* (2018) Convergence and divergence of CRH amacrine cells in mouse retinal circuitry. *J. Neurosci.* 38, 3753–3766
45. Asari, H. and Meister, M. (2012) Divergence of visual channels in the inner retina. *Nat. Neurosci.* 15, 1581–1589
46. Zaghloul, K.A. *et al.* (2007) Functional circuitry for peripheral suppression in Mammalian Y-type retinal ganglion cells. *J. Neurophysiol.* 97, 4327–4340

47. Franke, K. *et al.* (2017) Inhibition decorrelates visual feature representations in the inner retina. *Nature* 542, 439–444
48. Werblin, F.S. (2010) Six different roles for crossover inhibition in the retina: correcting the nonlinearities of synaptic transmission. *Vis. Neurosci.* 27, 1–8
49. Manookin, M.B. *et al.* (2008) Disinhibition combines with excitation to extend the operating range of the OFF visual pathway in daylight. *J. Neurosci.* 28, 4136–4150
50. Krishnamoorthy, V. *et al.* (2017) Sensitivity to image recurrence across eye-movement-like image transitions through local serial inhibition in the retina. *Elife* 6, e22431
51. Odermatt, B. *et al.* (2012) Encoding of luminance and contrast by linear and nonlinear synapses in the retina. *Neuron* 73, 758–773
52. Nikolaev, A. *et al.* (2013) Synaptic mechanisms of adaptation and sensitization in the retina. *Nat. Neurosci.* 16, 934–941
53. Marvin, J.S. *et al.* (2013) An optimized fluorescent probe for visualizing glutamate neurotransmission. *Nat. Methods* 10, 162–170
54. Zhao, Z. *et al.* (2020) The temporal structure of the inner retina at a single glance. *Sci. Rep.* 10, 4399
55. Field, G.D. *et al.* (2010) Functional connectivity in the retina at the resolution of photoreceptors. *Nature* 467, 673–677
56. Freeman, J. *et al.* (2015) Mapping nonlinear receptive field structure in primate retina at single cone resolution. *Elife* 4, e05241
57. Maheswaranathan, N. *et al.* (2018) Inferring hidden structure in multilayered neural circuits. *PLoS Comput. Biol.* 14, e1006291
58. Real, E. *et al.* (2017) Neural circuit inference from function to structure. *Curr. Biol.* 27, 189–198
59. Shi, Q. *et al.* (2019) Functional characterization of retinal ganglion cells using tailored nonlinear modeling. *Sci. Rep.* 9, 8713
60. Maheswaranathan, N. *et al.* (2019) The dynamic neural code of the retina for natural scenes. *bioRxiv* Published online December 17, 2019. https://doi.org/10.1101/340943
61. McIntosh, L.T. *et al.* (2016) Deep learning models of the retinal response to natural scenes. *Adv. Neural Inf. Proces. Syst.* 29, 1369–1377
62. Tanaka, H. *et al.* (2019) From deep learning to mechanistic understanding in neuroscience: the structure of retinal prediction. *Adv. Neural Inf. Proces. Syst.* 32, 8535–8545
63. Pang, R. *et al.* (2016) Dimensionality reduction in neuroscience. *Curr. Biol.* 26, R656–R660
64. Samengo, I. and Gollisch, T. (2013) Spike-triggered covariance: geometric proof, symmetry properties, and extension beyond Gaussian stimuli. *J. Comput. Neurosci.* 34, 137–161
65. Schwartz, O. *et al.* (2006) Spike-triggered neural characterization. *J. Vis.* 6, 484–507
66. Gollisch, T. (2013) Features and functions of nonlinear spatial integration by retinal ganglion cells. *J. Physiol. Paris* 107, 338–348
67. Kaardal, J. *et al.* (2013) Identifying functional bases for multidimensional neural computations. *Neural Comput.* 25, 1870–1890
68. Liu, J.K. *et al.* (2017) Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nat. Commun.* 8, 149
69. Lee, D.D. and Seung, H.S. (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791
70. Shah, N.P. *et al.* (2020) Inference of nonlinear receptive field subunits with spike-triggered clustering. *Elife* 9, e45743
71. Yu, W.-Q. *et al.* (2018) Synaptic convergence patterns onto retinal ganglion cells are preserved despite topographic variation in pre- and postsynaptic territories. *Cell Rep.* 25, 2017–2026
72. Berry, M.J. *et al.* (1999) Anticipation of moving stimuli by the retina. *Nature* 398, 334–338
73. Jarsky, T. *et al.* (2011) A synaptic mechanism for retinal adaptation to luminance and contrast. *J. Neurosci.* 31, 11003–11015
74. Manookin, M.B. and Demb, J.B. (2006) Presynaptic mechanism for slow contrast adaptation in mammalian retinal ganglion cells. *Neuron* 50, 453–464
75. Nagy, J. *et al.* (2021) GABA-A presynaptic inhibition regulates the gain and kinetics of retinal output neurons. *Elife* 10, e60994
76. Cui, Y. *et al.* (2016) Divisive suppression explains high-precision firing and contrast adaptation in retinal ganglion cells. *Elife* 5, e19460
77. Ozuysal, Y. and Baccus, S.A. (2012) Linking the computational structure of variance adaptation to biophysical mechanisms. *Neuron* 73, 1002–1015
78. Rieke, F. (2001) Temporal contrast adaptation in salamander bipolar cells. *J. Neurosci.* 21, 9445–9454
79. Baccus, S.A. and Meister, M. (2002) Fast and slow contrast adaptation in retinal circuitry. *Neuron* 36, 909–919
80. Kim, K.J. and Rieke, F. (2003) Slow Na+ inactivation and variance adaptation in salamander retinal ganglion cells. *J. Neurosci.* 23, 1506–1516
81. Weick, M. and Demb, J.B. (2011) Delayed-rectifier K channels contribute to contrast adaptation in mammalian retinal ganglion cells. *Neuron* 71, 166–179
82. Brown, S.P. and Masland, R.H. (2001) Spatial scale and cellular substrate of contrast adaptation by retinal ganglion cells. *Nat. Neurosci.* 4, 44–51
83. Garvert, M.M. and Gollisch, T. (2013) Local and global contrast adaptation in retinal ganglion cells. *Neuron* 77, 915–928
84. Ölveczky, B.P. *et al.* (2007) Retinal adaptation to object motion. *Neuron* 56, 689–700
85. Khani, M.H. and Gollisch, T. (2017) Diversity in spatial scope of contrast adaptation among mouse retinal ganglion cells. *J. Neurophysiol.* 118, 3024–3043
86. Appleby, T.R. and Manookin, M.B. (2019) Neural sensitization improves encoding fidelity in the primate retina. *Nat. Commun.* 10, 4017
87. Kastner, D.B. and Baccus, S.A. (2011) Coordinated dynamic encoding in the retina using opposing forms of plasticity. *Nat. Neurosci.* 14, 1317–1322
88. Kastner, D.B. and Baccus, S.A. (2013) Spatial segregation of adaptation and predictive sensitization in retinal ganglion cells. *Neuron* 79, 541–554
89. Wässle, H. *et al.* (1989) Cortical magnification factor and the ganglion cell density of the primate retina. *Nature* 341, 643–646
90. Sinha, R. *et al.* (2017) Cellular and circuit mechanisms shaping the perceptual properties of the primate fovea. *Cell* 168, 413–426
91. Bölinger, D. and Gollisch, T. (2012) Closed-loop measurements of iso-response stimuli reveal dynamic nonlinear stimulus integration in the retina. *Neuron* 73, 333–346
92. Petrusca, D. *et al.* (2007) Identification and characterization of a Y-like primate retinal ganglion cell type. *J. Neurosci.* 27, 11019–11027
93. Almasi, A. *et al.* (2020) Mechanisms of feature selectivity and invariance in primary visual cortex. *Cereb. Cortex* 30, 5067–5087
94. Bartsch, F. *et al.* (2021) Model-based characterization of the selectivity of neurons in primary visual cortex. *bioRxiv* Published online September 15, 2021. https://doi.org/10.1101/2021.09.13.460153
95. Vintch, B. *et al.* (2015) A convolutional subunit model for neuronal responses in macaque V1. *J. Neurosci.* 35, 14829–14841
96. Beyeler, M. *et al.* (2016) 3D visual response properties of MSTd emerge from an efficient, sparse population code. *J. Neurosci.* 36, 8399–8415
97. Mineault, P.J. *et al.* (2012) Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proc. Natl. Acad. Sci. U. S. A.* 109, E972–E980
98. Ahrens, M.B. *et al.* (2008) Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *J. Neurosci.* 28, 1929–1942
99. McFarland, J.M. *et al.* (2013) Inferring nonlinear neuronal computation based on physiologically plausible inputs. *PLoS Comput. Biol.* 9, e1003143
100. Keshishian, M. *et al.* (2020) Estimating and interpreting nonlinear receptive field of sensory neural responses with deep neural network models. *Elife* 9, e53445
101. Cunningham, J.P. and Yu, B.M. (2014) Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* 17, 1500–1509
102. Onken, A. *et al.* (2016) Using matrix and tensor factorizations for the single-trial analysis of population spike trains. *PLoS Comput. Biol.* 12, e1005189
103. Soulat, H. *et al.* (2021) Probabilistic tensor decomposition of neural population spiking activity. *Adv. Neural Inf. Proces. Syst.* 34, 15969–15980

104. Williams, A.H. *et al.* (2018) Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron* 98, 1099–1115
105. Barrett, D.G. *et al.* (2019) Analyzing biological and artificial neural networks: challenges with opportunities for synergy? *Curr. Opin. Neurobiol.* 55, 55–64
106. Glaser, J.I. *et al.* (2019) The roles of supervised machine learning in systems neuroscience. *Prog. Neurobiol.* 175, 126–137
107. Diamond, J.S. (2017) Inhibitory interneurons in the retina: types, circuitry, and function. *Annu. Rev. Vis. Sci.* 3, 1–24
108. Klindt, D. *et al.* (2017) Neural system identification for large populations separating 'what' and 'where.'. *Adv. Neural Inf. Proces. Syst.* 30, 3506–3516
109. Helmstaedter, M. *et al.* (2013) Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* 500, 168–174
110. Marc, R.E. *et al.* (2014) The AII amacrine cell connectome: a dense network hub. *Front. Neural Circuits* 8, 104
111. Park, S.J.H. *et al.* (2020) Connectomic analysis reveals an interneuron with an integral role in the retinal circuit for night vision. *Elife* 9, e56077
112. Walker, E.Y. *et al.* (2019) Inception loops discover what excites neurons most using deep predictive models. *Nat. Neurosci.* 22, 2060–2065
113. Gonçalves, P.J. *et al.* (2020) Training deep neural density estimators to identify mechanistic models of neural dynamics. *Elife* 9, e56261

ACCELERATED SPIKE-TRIGGERED NON-NEGATIVE MATRIX FACTORIZATION ALLOWS HYPERPARAMETER TUNING AND REVEALS DIFFERENCES AND ALIGNMENTS AMONG GANGLION CELL SUBUNIT MOSAICS IN THE PRIMATE RETINA

3

Attached manuscript. Research article, currently in preparation. Reprinted below is the manuscript at its current version.

Authors:

Sören J. Zapp[1,2], Dimokratis Karamanlis[1,2,3], Mohammad H. Khani[1,2,3], Helene M. Schreyer[1,2], Matthias Mietsch[4,5], Dario A. Protti[6], Tim Gollisch[1,2,7]

1. University Medical Center Göttingen, Department of Ophthalmology, Göttingen, Germany
2. Bernstein Center for Computational Neuroscience, Göttingen, Germany
3. International Max Planck Research School for Neurosciences, Göttingen, Germany
4. Laboratory Animal Science Unit, German Primate Center, Göttingen, Germany
5. German Center for Cardiovascular Research, Partner Site Göttingen, Göttingen, Germany
6. School of Medical Sciences (Neuroscience), The University of Sydney, Sydney, NSW, Australia
7. Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, Göttingen, Germany

Author contributions:

I developed and implemented the novel variant of STNMF, collected and curated the marmoset retina data, analyzed the data, and prepared figures and the initial draft of the manuscript. DK, MHK, HMS, MM, and DAP all contributed to the preparation of the marmoset retina and data collection. TG supervised the project and contributed to writing the manuscript. All authors helped revise the manuscript. The salamander data stem from a publically available repository.

Gollisch, T., and Liu, J. K. (2018). Data: Salamander retinal ganglion cells under finely structured spatio-temporal white noise. *G-Node*. https://doi.org/10.12751/g-node.62b65b.

Accelerated spike-triggered non-negative matrix factorization allows hyperparameter tuning and reveals differences and alignments among ganglion cell subunit mosaics in the primate retina

Sören J. Zapp[1,2], Dimokratis Karamanlis[1,2,3], Mohammad H. Khani[1,2,3], Helene M. Schreyer[1,2], Matthias Mietsch[4,5], Dario A. Protti[6], Tim Gollisch[1,2,7]

1. University Medical Center Göttingen, Department of Ophthalmology, Göttingen, Germany
2. Bernstein Center for Computational Neuroscience, Göttingen, Germany
3. International Max Planck Research School for Neurosciences, Göttingen, Germany
4. Laboratory Animal Science Unit, German Primate Center, Göttingen, Germany
5. German Center for Cardiovascular Research, Partner Site Göttingen, Göttingen, Germany
6. School of Medical Sciences (Neuroscience), The University of Sydney, Sydney, NSW, Australia
7. Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, Göttingen, Germany

ABSTRACT

A standard circuit motif in sensory systems is the pooling of sensory signals from a neuronal layer whose cells carry information about different locations in (stimulus) space. While the joint collection pooled signals define the receptive field of the downstream neuron, nonlinear transformations of the signal transfer give rise to functional subunits inside the receptive field. For ganglion cells in the vertebrate retina, for example, receptive field subunits are thought to correspond to presynaptic bipolar cells. Identifying the number and locations of subunits from the stimulus–response relationship of a recorded ganglion cell has been an ongoing challenge in order to characterize the retina's functional circuitry and to build computational models that capture nonlinear signal pooling. Here we present a novel version of spike-triggered non-negative matrix factorization (STNMF), which can extract localized subunits in ganglion-cell receptive fields from recorded spiking responses under spatiotemporal white-noise stimulation. The method provides a more than 100-fold speed increase compared to a previous implementation, which can be harnessed for systematic screening of hyperparameters, such as sparsity regularization. We demonstrate the power and flexibility of this new STNMF variant by analyzing populations of ganglion cells from salamander as well as primate retina. We find that subunits of midget as well as parasol ganglion cells in the marmoset retina form separate mosaics that tile visual space. Moreover, subunit mosaics are aligned with each other for ON and OFF midget as well as for ON and OFF parasol cells, indicating a spatial coordination of ON and OFF signals at the bipolar-cell level. Thus, STNMF can reveal organizational principles of signal transmission between successive neural layers, which are not easily accessible by other means.

INTRODUCTION

In sensory pathways, signals typically carry spatial, temporal or spatiotemporal information of the outside world. The information is transmitted by a cascade of neurons, encoding a constrained region of the stimulus as their receptive field. Along the pathway, post-synaptic neurons pool signals from their pre-synaptic layer into their receptive field. Stimuli may be processed uniformly across the receptive field or are integrated nonlinearly, giving rise to substructure within the receptive field. Nonlinear substructure may be ascribed to the receptive fields of pre-synaptic inputs that form functional subunits within the receptive field of the post-synaptic neuron. While linear convergence may serve to convey signals, nonlinear integration of subunit inputs in the receptive field enables complex computations. Through uneven weighting and nonlinear transformation of their signal, subunits give rise to diverse responses to stimuli. These effects have been observed as early as in low-level sensory organs like the retina. Bipolar cell inputs may be integrated by retinal ganglion cells as nonlinear subunits in their receptive fields to play a substantial role in sensitivity to high spatial contrast (Enroth-Cugell and Robson 1966; Schwartz et al. 2012; Victor and Shapley 1979), and local dynamics like contrast adaptation (Jarsky et al. 2011; Manookin and Demb 2006; Rieke 2001). Understanding how subunits are integrated in the receptive field is essential to, on the one hand, describing functional connectivity and the propagation of information in sensory networks. Insights may aid in explaining local dynamics of receptive field computation. On the other hand, creating a spatial map of subunit arrangement by neuronal cell type may guide morphological studies. A reoccurring implementation of receptive field layouts is a mosaic-like arrangement of feature detectors. In the retina it was reported across species for photoreceptors (Marc and Sperling 1977), bipolar cells (Cohen and Sterling 1990a, 1990b; Sterling et al. 1988), and ganglion cells (Anishchenko et al. 2010; DeVries and Baylor 1997). It is unclear how functional subunits may fit into that picture, whether subunits arrange in mosaics, and how their arrangement differs across ganglion cell types.

The retina subjected to light stimulation has proved to be a suitable model organism for investigating the functional connectivity in sensory systems. Nevertheless, revealing subunits experimentally by stimulating and recording pre- and post-synaptic layers of neurons simultaneously remains challenging. In light of the experimental limitations, several methods have been proposed to infer subunits computationally. Spiking responses of the retinal ganglion cell under careful stimulation are examined to extract subunits of its receptive field. The methods include cascade networks modelling subunits as linear filters with a nonlinear transfer function converging into linear summation and another nonlinearity (Maheswaranathan et al. 2018; McFarland et al. 2013). Convolutional neural networks offer a similar model of convergence with potentially deeper layering and more expressive power (Maheswaranathan et al. 2019; McIntosh et al. 2016; Tanaka et al. 2019). On the other side of the spectrum, there are the statistical analyses, spike-triggered clustering (Shah et al. 2020) and spike-triggered non-negative matrix factorization (STNMF) (Liu et al. 2017). STNMF recovers subunits from pixel correlations in the spike-triggering stimuli under a sparsity constraint. Subunits extracted with STNMF have been shown to correspond to receptive fields of simultaneously recorded bipolar cells. However, the proposed implementation is time-intensive in the order of tens of hours per ganglion cell. Furthermore, difficulty to control the effects of hyperparameters does not allow the transfer to datasets of different organisms or to different cell types.

Here, we develop a new implementation of STNMF that addresses the shortcomings of previous implementations. We introduce a novel combination of state-of-the-art algorithms of non-negative matrix factorization (NMF) and complement it with guided initialization procedures to improve speed and reliability. The resulting method allows subunit recovery in a matter of seconds for one cell. We demonstrate how clustering consensus methods can be applied for hyperparameter selection and offer tools that reliably provide suitable values for sparsity regularization. This makes STNMF more versatile across cells of different functional properties and across datasets of different species. In this work, we firstly demonstrate its superior speed and flexibility in analyses of cell populations of salamander and primate retinas. Secondly, the implementation of STNMF allows to match subunit properties to anatomical findings in parasol and midget ganglion cells of primates. Thirdly, we use STNMF to investigate subunit mosaic arrangements of ON- and OFF-type pathways and compare them to ganglion cell receptive field mosaics.

RESULTS

STNMF extracts spatial subunits from pixel correlations in the stimulus eliciting spiking response of a ganglion cell. For detecting these correlations, STNMF requires a spatiotemporally uncorrelated stimulus, preferably spatiotemporal binary white noise, with a spatial resolution finer than the expected subunit size. Local changes in contrast stimulate different parts of the substructure in the receptive field. Without prior knowledge about the location and size of receptive field and subunits, white noise makes a powerful stimulus to drive the cell and probe its receptive field substructure in its different pixel configurations. Recurring, localized pixel patterns in the stimulus that elicit a spiking response of the cell are extracted as spatial subunits by STNMF.

In order to do so, STNMF performs its inference on the spike-triggered stimuli. These are the stimulus frames preceding each spike within a window of the temporal sensitivity of the receptive field. Since subunits typically share temporal dynamics (Maheswaranathan et al. 2018) with the receptive field, focusing on the spatial profile has been a viable option to reduce input dimensions (Liu et al. 2017; Shah et al. 2020). To do so, the temporal component of the spike-triggered stimuli is integrated out. They are compressed to one effective spike-triggered stimulus frame each, by taking the average of the stimulus frames weighted with the temporal profile of the receptive field, the spike-triggered average (STA). Cropped to the spatial region of the receptive field, the collection of these effective stimuli is called the effective spike-triggered stimulus ensemble (STE) (**Figure 1A**). With its dense stimulus-response mapping, we use it as the input to STNMF to find subunits using semi-non-negative matrix factorization (semi-NMF) (Ding et al. 2010).

STNMF decomposes the STE matrix into two smaller matrices, reoccurring spatial modules and their corresponding weights (**Figure 1B**). Any stimulus that triggered a given spike is the linear combination of all modules and the weights ascribed to that spike. With a non-negativity constraint on the module matrix, governed by the use of semi-NMF, there are no negative pixel values to cancel out in the linear combination. As a result, the modules emerge as additive building blocks. By limiting the number of modules while preserving the linear combinations, the additive components are forced to share modules according to the pixel correlations in the STE. This grouping forms spatially localized modules that correspond to reoccurring pixel patterns that trigger the spiking

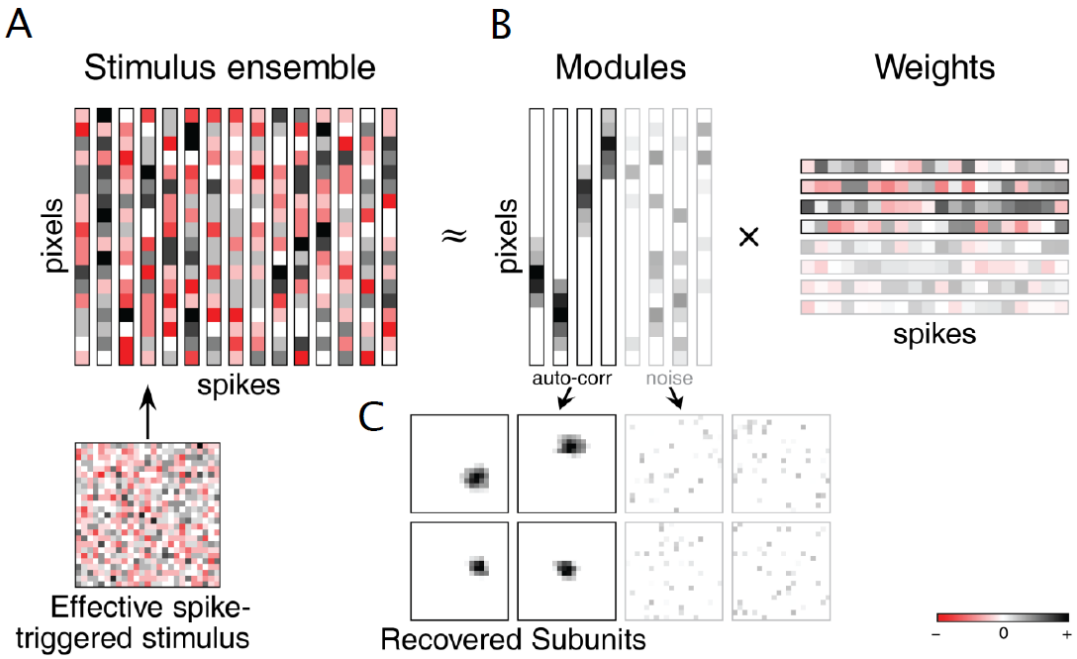


Figure 1 Spike-triggered non-negative matrix factorization. (A) The effective spike-triggered stimulus ensemble (STE) is a matrix of pixels by spikes. Each column is the effective spike-triggered stimulus, the stimulus that triggered a given spike reduced to a spatial frame. The STE serves as the input to the matrix factorization. (B) Using semi-NMF the STE is decomposed into two smaller matrices, the spatial modules and their corresponding spike weights. By design some modules exhibit a localized structure while others (here dimmed) capture noise. (C) The modules can be distinguished by their autocorrelation as recovered subunits that represent the structure of the receptive field.

response. They are the subunits within the receptive field of the ganglion cell as recovered by STNMF (**Figure 1C**).

The optimal number of modules is unknown, and its selection poses a common hurdle in the field of NMF. Much like for clustering algorithms, clusters merge or oversplit if their number is not chosen correctly. STNMF, however, is largely robust to the selection of the number of modules. In the rectangular shape of the stimulus frames, pixels outside the receptive field generally carry no information contributing to the spiking of the ganglion cell. Consequently, they show little correlation with other pixels across the spikes and may appear alongside any other pixel in any module and may be distributed across fewer or more modules. Any excess modules will be populated by such noisy pixels making the localized subunits in shape and number robust to varying number of modules. Their selected number consequently serves as an upper bound only and does not require any further assumptions on the data. As previously suggested, we find 20 modules to be a good number for our application on retinal ganglion cells (Liu et al. 2017). The recovered modules of interest are visibly distinguishable from the noisy excess modules by their spatially localized structure. We automate this differentiation by defining those modules as recovered subunits that have a sufficiently high autocorrelation (**Figure 1C**).

We investigated different algorithms underlying and surrounding NMF to provide the most suitable implementation in context of subunit recovery. We support our findings with analyses of neural spiking data. We demonstrate our methods on multielectrode recordings from retinal ganglion cells of salamander recorded previously (Gollisch and Liu 2018) and comparison to previous analyses (Liu et al. 2017). We complement these analyses with new findings in multielectrode recordings from ganglion cells in marmoset retinas. All ganglion cells were subjected to visual white-noise stimulation.

Accelerated fast hierarchical alternating least squares speeds up subunit recovery

Due to its non-negativity constraint, NMF is NP-hard (Vavasis 2010). To obtain a good decomposition NMF is generally performed by iterative improvements. Various algorithms facilitate these iterations under the non-negativity constraint. Most of them are accurate but costly algorithms (J. Kim and Park 2011). They seek exact solutions for the modules and weights at each iteration by considering all data at once. The pixel values of all modules are updated from all spike weights and all stimuli. This poses a high computational demand. In recent years, hierarchical alternating least squares (HALS) (Cichocki et al. 2007; Ho 2008; L. Li and Zhang 2009) has gained popularity. Instead of calculating accurate solutions at each iteration of NMF, it relies on superior speed to make up for approximated solutions at each step. Over more but comparably faster iterations it outperforms previous methods in many applications of NMF (Cichocki et al. 2007; Cichocki and Phan 2009; Gillis and Glineur 2012) while ensuring convergence (Gillis and Glineur 2008). Instead of inferring all modules from their weights at once, HALS updates the modules sequentially. For each module, the STE is reconstructed from the remaining modules and their weights. The pixel values of the left-out module are then inferred from the data unaccounted for in the incomplete reconstruction. By iterating over the modules sequentially, the missing data caused by leaving out individual modules becomes more concentrated and the approximation improves by each module updated. Therefore, efficiency is improved, by replacing large matrix multiplications with local vector operations. The speed of this approach compensates for this approximation, because more iterations are possible in shorter amount of time. Various modifications of the HALS algorithm have been proposed recently to further increase its speed or its accuracy.

Here, we combine two HALS algorithms into what we call Accelerated Fast HALS (AF-HALS). Fast HALS (Cichocki and Phan 2009) achieves a speed-up through computational simplification by employing vector normalization to canceling out terms in the update equation. Accelerated HALS (Gillis and Glineur 2012) improves factorization accuracy per iteration by performing multiple consecutive HALS updates as long as the number floating point operations stays below the heavy matrix operations preceding the next iteration. With Fast HALS each iteration in NMF becomes faster, requiring less time in total. With Accelerated HALS each iteration becomes more accurate, requiring fewer iterations. To demonstrate a direct comparison in speed and accuracy in the context of STNMF, we selected the popular active-set method (Lawson and Hanson 1974). It is representative for the state-of-the-art NMF algorithms and prevailing for its high accuracy at each iteration. Previous implementations of STNMF were based on the active-set method. We demonstrate the differences exemplary on a retinal ganglion cell from the salamander retina. Spiking activity was recorded under white-noise stimulation and previously analyzed in the context of subunit recovery (Gollisch and Liu 2018; Liu et al. 2017). We find that under the same circumstances and

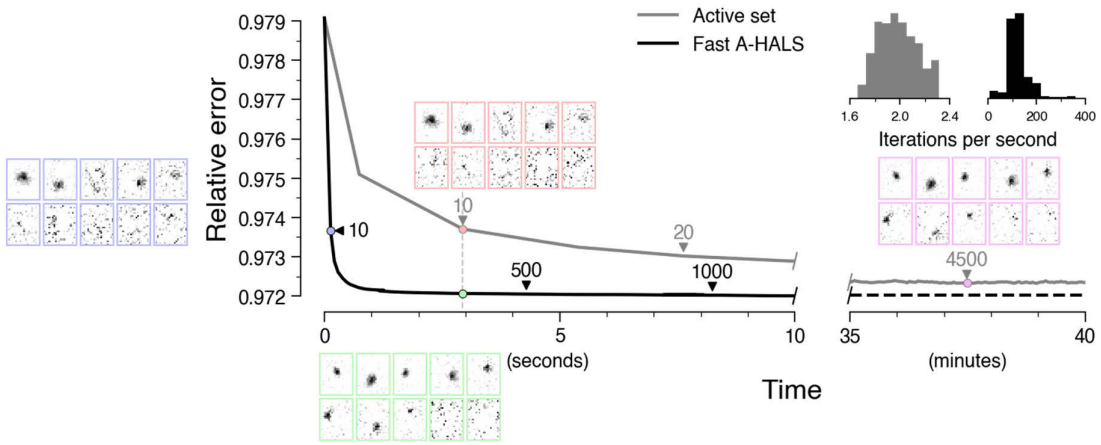


Figure 2 Reconstruction error of STNMF implemented with Accelerated Fast HALS (black) and STNMF based on the active-set method (gray) for an example ganglion cell from salamander data. Arrows denote number of iterations. Insets allow visual inspection of the recovered modules at different states during the iterations. Iterations of AF-HALS are faster than those of the active-set method (top right). Although the active set method calculates more accurately at each step, the reconstruction process is similar after ten iterations in both methods (blue and red insets). This is due to refined updates in AF-HALS. The active set method does not reach the error that AF-HALS accomplishes within a few seconds. As reconstruction error does not describe the subunit decomposition, the visualized subunits of the active-set method after 45 minutes (purple inset) arrive at the state that AF-HALS reached within five seconds (green inset).

initial conditions, AF-HALS converges on localized modules both faster and within fewer iterations. (**Figure 2**).

Guided initialization replaces repetitions

As NMF is a non-convex optimization problem (Vavasis 2010), there are no guarantees to uniquely identify the globally optimal solution. In typical iterative implementations, solutions are guaranteed to only be local optima and are highly dependent on the starting point of the iterative search (D. D. Lee and Seung 1999). The starting point is given by the initial values of the modules from which the iterative improvements originate. When there are no countermeasures against getting stuck in local optima, the initialization locks the algorithm into a fixed region of the solution space. A popular way to compensate for that is choosing the best solution among multiple repetitions of NMF starting from modules that are initialized with different instantiations of random pixel values (M. W. Berry et al. 2007; Cichocki et al. 2009).

However, random initialization does not offer a strong starting point in general (Wild 2003), many repetitions pose a high demand on computational time, and a measure is necessary for determining the best solution. We eliminate these obstacles by substituting the time-consuming repetitions by implementing guided initialization. We followed an approach that is more favorable to random initialization with techniques based on singular value decomposition (SVD) (Boutsidis and Gallopoulos 2008). Components of SVD provide spatial structure that would otherwise only arise throughout the iterations of NMF when starting from random initialization. Furthermore, SVD-based initialization yields lower error in solutions and earlier convergence of NMF (Qiao 2015). To

accommodate for the non-negativity constraint of NMF, methods of SVD-based initialization typically set negative values in the SVD components to zero. Instead of discarding this valuable information, we retain it by using non-negative singular value decomposition with low-rank correction (NNSVD-LRC) (Atif et al. 2019). NNSVD-LRC takes advantages of the sign ambiguity of SVD and duplicates selected components with opposite sign before setting the negative values in all components to zero. In comparison, this offers potentially twice the amount of information from the components of high singular values. Furthermore, this procedure produces initial modules with around half of their values zero. That offers around 50% sparsity before even starting NMF. NNSVD-LRC subsequently performs a few initial NMF iterations on these low-rank extractions. At comparably low computational cost, this provides a head start for the full NMF. We modified NNSVD-LRC to be suitable for semi-NMF and improved its numerical stability. This includes duplicating the previously skipped component of the highest singular value and replacing zero-columns to avoid convergence issues in HALS-based NMF.

We find that a single run of STNMF initialized with NNSVD-LRC tends to recover all recurring subunits of 100 repetitions of randomly initialized STNMF. To show this, we compared initialization and STNMF of retinal ganglion cells of salamander (Gollisch and Liu 2018). Spatially localized subunits recovered across randomly initialized STNMF runs were marked as reoccurring by their relative spatial overlap. These match the recovered subunits from NNSVD-LRC-based STNMF. These findings demonstrate the viability of NNSVD-LRC and allows omitting the 100 repetitions offering a 100-fold speed increase (**Figure 3**).

Effects of sparsity regularization on subunit layout

While the non-negativity constraint is placed on all matrices in NMF, only the modules are non-negative in semi-NMF. Due to the loosened constraint, semi-NMF fails to retain additive, parts-based solutions without additional constraints (Ding et al. 2010). This crucial aspect of NMF can be corrected in semi-NMF through sparsity regularization on one of the factor matrices (Hoyer 2002, 2004; H. Kim and Park 2007; J. Kim and Park 2008). Unlike other methods of subunit recovery or machine learning in general, where regularization serves as an optional step to counteract overfitting or to alleviate the lack of sufficient data, sparsity regularization in STNMF is therefore an essential part of its functionality. To obtain localized subunits, we enforce the modules to be both non-negative and sparse. We achieve this by adding an additional term to the objective function that penalizes large number of non-zero pixel values in the modules.

Increasing sparsity regularization drives more pixel values in the modules to zero. Once regularization becomes too strong, subunits are forced to shrink in size or will be split into multiple smaller subunits until they eventually vanish. The optimal amount of sparsity depends on the dimensionality of the data. Accessible properties, like the number of pixels and number of spikes, as well as unknown characteristics prior to subunit recovery, like the true number of subunits and their sizes, play a role. Consequently, the regularization is difficult to control and has to be adjusted for cells that differ in their properties. Likewise, cells that share characteristics will share similarity in suitable amount of regularization.

We examined the impact of sparsity on the subunit decomposition on data recorded from ganglion cells of a salamander retina (Gollisch and Liu 2018). We ran STNMF on the STEs of selected cells

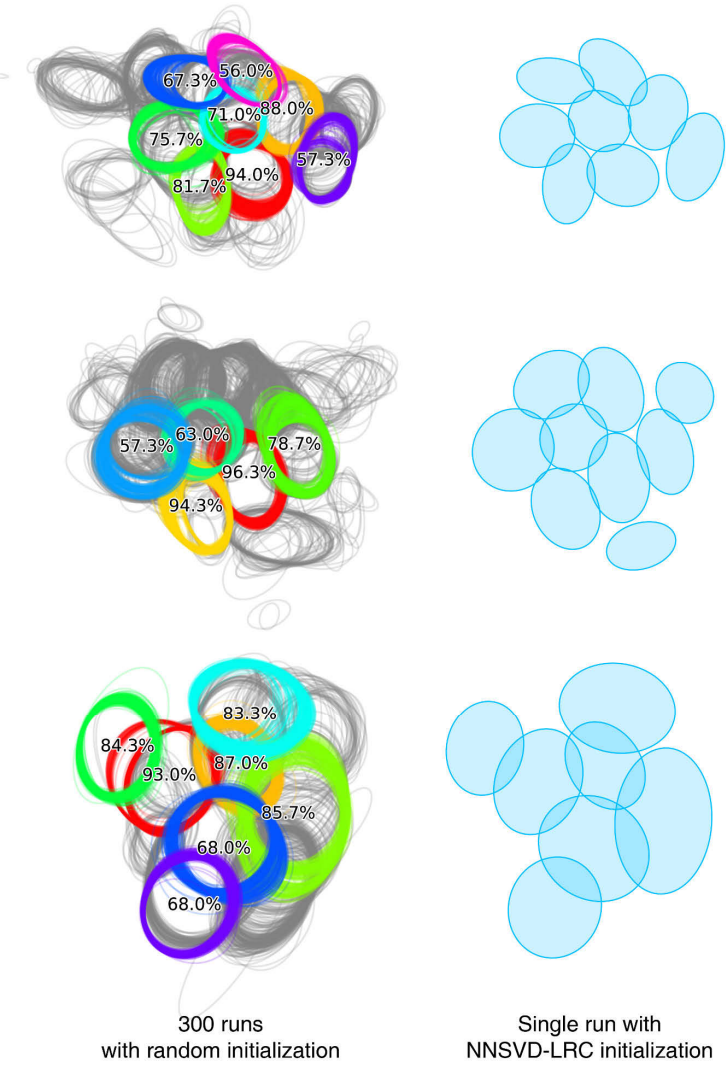


Figure 3 A single run of NNSVD-LRC initialized STNMF (right) finds all subunits that reoccur among 300 randomly initialized runs (left) for three example cells (marmoset OFF parasol cells top and middle, salamander Fast Off cell bottom). Subunits are shown as 2-dimensional Gaussian elliptical fits of 1.5 standard deviations. Reoccurring subunits are matched by more than 50% area overlap of the ellipses. Subunits that reoccur in more than half of the repetitions are colored and assigned with a percentage of reoccurrence. The NNSVD-LRC based initialization finds all subunits that appear across the randomly initialized runs for all cells.

at varying sparsity regularization weights (**Figure 4C inset**). Without sparsity regularization, we obtain holistic representations. These break up into large subunits that split into multiple smaller subunits with increasing regularization. For a certain range, the subunit layout becomes more reliable despite changes in regularization, with only smaller variations. These modulations relate to the mean STNMF weight of each subunit. While we find high-weight subunits to be robust to deviations in the

sparsity regularization parameter, less prominent subunits are more sensitive and may differ in shape and size or vanish from the decompositions. The STNMF weight of a subunit has been shown to be related to the amount of spiking contribution of the ganglion cell (Jia et al. 2022; Liu et al. 2017). A subunit less reliable to appear in the decomposition is indicative of less involvement in the spike generation of the ganglion cell and the formation of its receptive field. Consequently, its importance in describing the receptive field or in modelling the spiking response of the ganglion cell proportional to its STNMF weight. Although prominent subunits are robust to variations, controlling for the right amount of sparsity regularization is beneficial to provide a subunit layout that is most likely to represent the underlying structure of the receptive field.

Consensus analyses aid in selection of regularization parameter

The regularization requires the choice of a hyperparameter value to control the amount of sparsity. With our tractable implementation of STNMF, we here provide a method to do so in reasonable amount of time. A common approach to finding the most suitable weight for regularization is through cross-validation. However, techniques that rely on held-out data are problematic in applications of unsupervised learning like NMF (Bro et al. 2008; Owen and Perry 2009). The weight matrix in STNMF is highly dependent on the spikes from the specific input and cannot be used to compare training and testing data sets. Likewise, comparing the pixel values of the module matrix is not trivial and we find that subunits recovered from data with more spikes are more refined compared to those obtained from less data, if emerging at all. Although adjusted cross-validation-like procedures for matrix factorization have been proposed (Fu and Perry 2017; Owen and Perry 2009), inputs with missing data does not generalize well to accelerated NMF algorithms like Accelerated HALS that our AF-HALS is based on. This is due to optimizations regarding vectorized calculations and matrix operations that are factored out of the column-wise updates. Furthermore, there is no intuitive quality measure for recovered subunits in the context of sparsity regularization. The reconstruction error of the factorization, for instance, simply increases with increasing regularization – for both training and testing set.

Instead of cross-validation, we turn to the solution of a similar parameter selection problem. A common hurdle for clustering techniques is the choice of the optimal number of clusters. This problem also translates to NMF and may be addressed using so-called consensus methods (Monti et al. 2003). For the course of this process, we temporarily neglect the aforementioned SVD-based initialization and revert to initialize NMF by random sampling. By the stochasticity of random initializations, NMF may vary in its solution on each run. This seemingly disadvantageous aspect is exploited here to find the best number of clusters. The more appropriate the choice of the number of clusters, the more the clustering assignments become similar across solutions, that is, the clustering consensus increases and the clustering becomes more stable (Brunet et al. 2004). Although we do not face this issue in STNMF, as the number of modules only serves as an upper bound, we extended the consensus approach to find a suitable regularization parameter for sparsity instead.

In STNMF, the weight factor matrix does not contain hard cluster assignments, but a soft weighting of how much each subunit contributed to eliciting a given spike. To build the consensus across repeated runs, we define the subunit most affiliation of each spike instead. The affiliated subunit is the module with the maximum weight for a given spike. The consensus matrix holds, for all pairs of

spikes, the fraction of runs in which the pair is affiliated with the same subunit. For our purposes, we add a slight adjustment. We define any pair with either spike being most affiliated with a non-localized module as not equal, such that only spikes attributed to localized subunits contribute positively to the consensus. The ordered consensus matrix visualizes the subunit affiliation (**Figure 4B**). As proposed previously (Brunet et al. 2004), we quantify the dispersion of values in the consensus matrix by its cophenetic correlation coefficient (CPCC). The cophenetic correlation measures the degree of clustering in the consensus matrix in a hierarchical clustering distance manner. This provides a scalar metric between zero and one, with higher values indicating increased clustering and thus increased stability of the decomposition. As we curate the consensus matrix to consider localized subunits only, a high coefficient can also arise when strong regularization prevents the formation of any localized subunits causing perfect consensus.

For a given sparsity parameter, we run STNMF repeatedly under different random initialization. The consensus analysis reflects how robust the parameter is to the initializations and describes the stability of the subunit decomposition to be compared across different parameters. We visualize the stability with the CPCCs forming a stability curve over all parameters (**Figure 4C**). Stability increases with stronger sparsity regularization, as it increasingly constrains the solution and decreases variability among the decompositions. Once the regularization strength outweighs the solution, stability plateaus or decreases. This behavior exhibits a distinctive bend on the stability curve after a sharp incline. We choose the parameter at the bend as the optimal sparsity value, that is, just before the stability plateaus. It reflects the desired amount of regularization: As stable as possible, while affecting the solution as little as possible. This procedure is a viable option of identifying appropriate hyperparameters. After selection of a suitable parameter, we run STNMF in that configuration initialized with NNSVD-LRC to obtain the final set of localized subunits.

Regularization parameter is inferable from a subset of cells

We applied the consensus analysis on data recorded from salamander retina (Gollisch and Liu 2018). To obtain a suitable sparsity regularization parameter for subsequently recovering the subunits for all cells, we proceed as follows. Among the cells, we select a representative subset that resides near the population average for properties like receptive field diameter, number of spikes, and amount of noise in the STA by visual estimation. Here, with our computationally fast implementation of STNMF at hand, we are able to probe the effect of sparsity regularization quickly as a first step to narrow down the range of interest. We run STNMF with NNSVD-LRC with increasing sparsity regularization to find an upper bound after which all values in the modules approach zero (**Figure 4A**). As STNMF imposes sparsity at each iteration, we find 200 iterations to suffice for estimating the sparsity as the fraction of zero values in the modules. We observe that sparsity regularization typically starts to dominate the solution within one order of magnitude of the parameter value. Within the determined range, we investigate the solution stability for a coarse selection of sparsity regularization values using the consensus across repeated STNMF runs of random initialization. As this serves to further narrow down suitable bounds of sparsity, we only perform five iterations at each sparsity value to save time. For cells with a nonlinear receptive field that does in fact exhibit subunits, the stability curve rises sharply and shows one initial peak before decreasing towards the determined bound of sparsity (**Figure 4A**). The area around the peak is the range of interest as it captures both ends of the impact of sparsity. Within this window, we then perform the full consensus

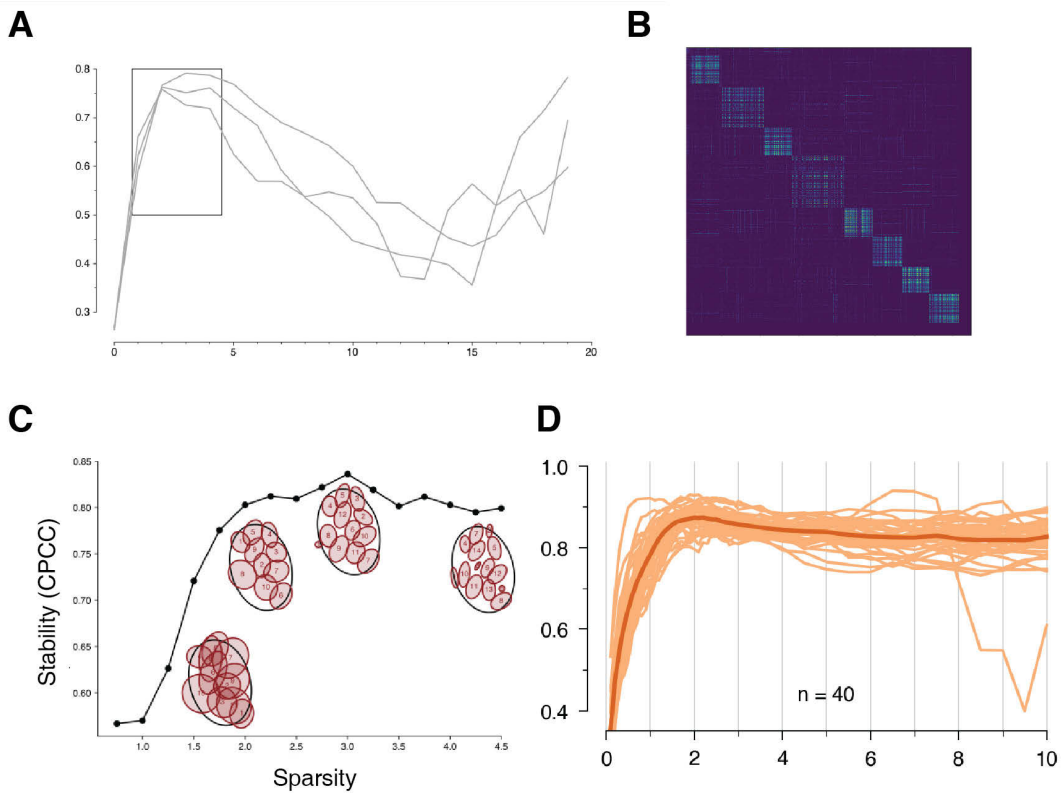


Figure 4 Consensus analysis for determining sparsity regularization. (A) The first step in the analysis is finding the threshold where regularization dominates the decomposition evident by the level of sparsity in the modules (blue). Within that range, coarse values of cophenetic correlation are probed to find the window of interest (dashed box). (B) Exemplary consensus matrix showing the pairwise agreement between all spikes. A good consensus is indicated by clear clusters of most affiliated subunits (high values) on the diagonal. (C) Stability curve for an example cell from salamander data. The diagonal clustering structure in (B) is measured with a scalar value, the CPCC. The correlation increases with increasing sparsity regularization, supported by refined subunit outlines (inset). After a peek in stability, subunits become too sparse until they eventually vanish with too high sparsity. (D) Stability curves of all Fast Off cells from a salamander retina.

analysis at high resolution of parameter values on the subset of cells. Here propose between 20 to 30 repetitions for each parameter value with 1000 STNMF iterations to obtain a reasonable estimate of stability. Finally, we estimate the suitable sparsity regularization parameter from the bend in the average of the stability curves by visual inspection. We then apply that parameter to NNSVD-LRC-based STNMF to recover the subunits of all cells.

We confirmed that inferring the parameter from a subset of cells is representative for all cells by performing the consensus analysis on the full population on a wide range of sparsity regularization. For the sake of thoroughness, we performed almost twice as many repetitions (50) of each of the 44 parameter values ranged between zero and ten. Note, that this extensive analysis takes substantially more time and was performed here for comparison purposes only. We suggest using the proposed values and ranges and following the procedure described above. At stronger regularization, the stability of some cells abruptly approaches zero. In these cases, the regularization is too strong and

breaks subunits into smaller structures arbitrarily, preventing consensus across repeated decompositions. An appropriate amount of regularization most likely is lower in these cases. Nevertheless, all stability curves are well in agreement such that an arbitrary subset of cells is representative of the population here (**Figure 4D**).

Subunits match previous analyses of bipolar cell receptive fields

We recover subunits of Fast Off ganglion cells in salamander data (Gollisch and Liu 2018) given a the procedure described above. To estimate the subunits of all cells, STNMF took a total of five minutes on a conventional office computer averaging at around 9 seconds per ganglion cell. Although that surpasses well beyond a 100-fold reduction in inference time in comparison, the obtained subunits match previous analyses (Liu et al. 2017). We observe a close resemblance in the set of subunits of individual cells (**Figure 5A**) and find the subunit outlines to align for the cell population (**Figure 5B**). For some of the receptive fields, we recover additional subunits to the otherwise matching subunit layout that have not been detected before. Furthermore, we were able to reproduce previously reported overlaps between subunits across cell pairs (**Figure 5C**). The differences we observe concern subunits with lower magnitude weights in the STNMF decomposition. These do not match in comparison with the previous analyses. One possible explanation maybe the differing procedure to determine the final subunit layout. Here, the subunits correspond to the localized modules from STNMF without further curation. In the previous analysis (Liu et al. 2017), the final subunits were obtained from the average of reoccurring subunits over multiple repetitions. Lower-weight subunits that occur less frequently across the repetitions are more likely to have been excluded or merged to adjacent subunits.

Experiments combining multielectrode recordings of ganglion cells with single cell recordings of bipolar cells had been performed previously in salamander retinas to investigate the relationship between computationally inferred subunits and measured bipolar cell receptive fields (Liu et al. 2017). Here, we confirm the results using the data of the respective ganglion cells. We find similar overlaps between subunits and bipolar cell receptive fields. These findings suggesting that our implementation can uncover bipolar cell receptive fields as subunits (**Figure 5D**).

Sparsity regularization parameter is shared within cell type

Typically, in machine learning applications, parameters are tuned to individual inputs. However, here the required amount of sparsity regularization depends on the dimensionality of the data, that is, the number of pixels and the number of spikes. These are directly related to receptive field size and firing rate of the cell, respectively. As cells of the same functional type share these properties more so than cells of different types, the required sparsity may differ across cell types. To analyze whether this is reflected in the optimal sparsity regularization parameter, we recorded spiking activity from ON midget and OFF midget ganglion cells of a marmoset retina under white-noise stimulation (**Figure 6A**). We examined the regularization consensus and compared the resulting stability for a wide range of parameter values between zero and ten. We found the stability curves as expected with a clear bend to determine the optimal regularization parameter value for cells of both types (**Figure 6B**). When arranged by cell type, the shape of the stability curves exhibit similarities for each. However, this cannot be stated for curves across cell type (**Figure 6C**). The ON midget cells show a sharper

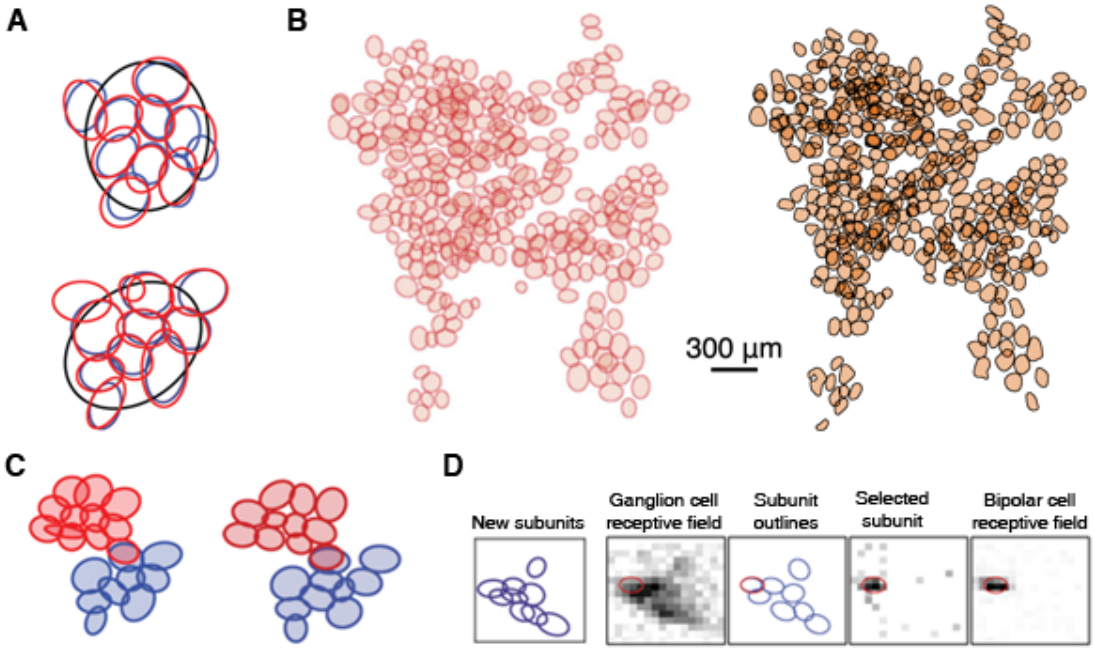


Figure 5 Subunit comparison to previous analysis of salamander data. (A) Subunit decompositions for two example cells. The decompositions resemble the previously estimated subunits (blue (Liu et al. 2017), red new). (B) Subunit mosaics of the previous analysis as 1.5 SD Gaussian outlines (left) and with the new implementation as 1 SD contours (right). We find mostly all subunits that have been identified previously and a few more. (C) Subunit overlap across adjacent Fast Off ganglion cells is reproduced. (D) We find subunits that resemble the receptive field of recorded bipolar cells. The subunits we recover (left) resemble the bipolar cell receptive field (right). Panels B (left), C (right), and D (right) are adapted from (Liu et al. 2017) licensed under CC BY 4.0.

incline in stability, while the OFF midget curve starts at lower stability. This pronounces the agreement for cells of same type, which is better visualized by principal component analysis (**Figure 6D**). Although not linearly separable for the first two principle components, the distances between pairs of curves are significantly larger across cell types than for curves of the same type (independent t-test, $p < 0.01$) (**Figure 6E**). This does not indicate that stability always differs across cell type, but rather suggests agreement within cell type. Along with the tight alignment of the stability curves observed in the salamander dataset (**Figure 4D**), this provides evidence that the optimal sparsity regularization parameter may be inferred from a representative subset of cells for each functional cell type.

Sparsity regularization parameter may translate across experiments

Similarity in consensus analysis among cell types extends across experiments. For three datasets recorded from marmoset retinas, we find the stability curves of OFF midget cells to align (**Figure 7**). Although experimental conditions differed and temporal resolution of the stimulation and recording length varied, the same optimal sparsity regularization parameter is deductible. This indicates that the sparsity regularization parameter may also translate across experiments for cells of the same type.

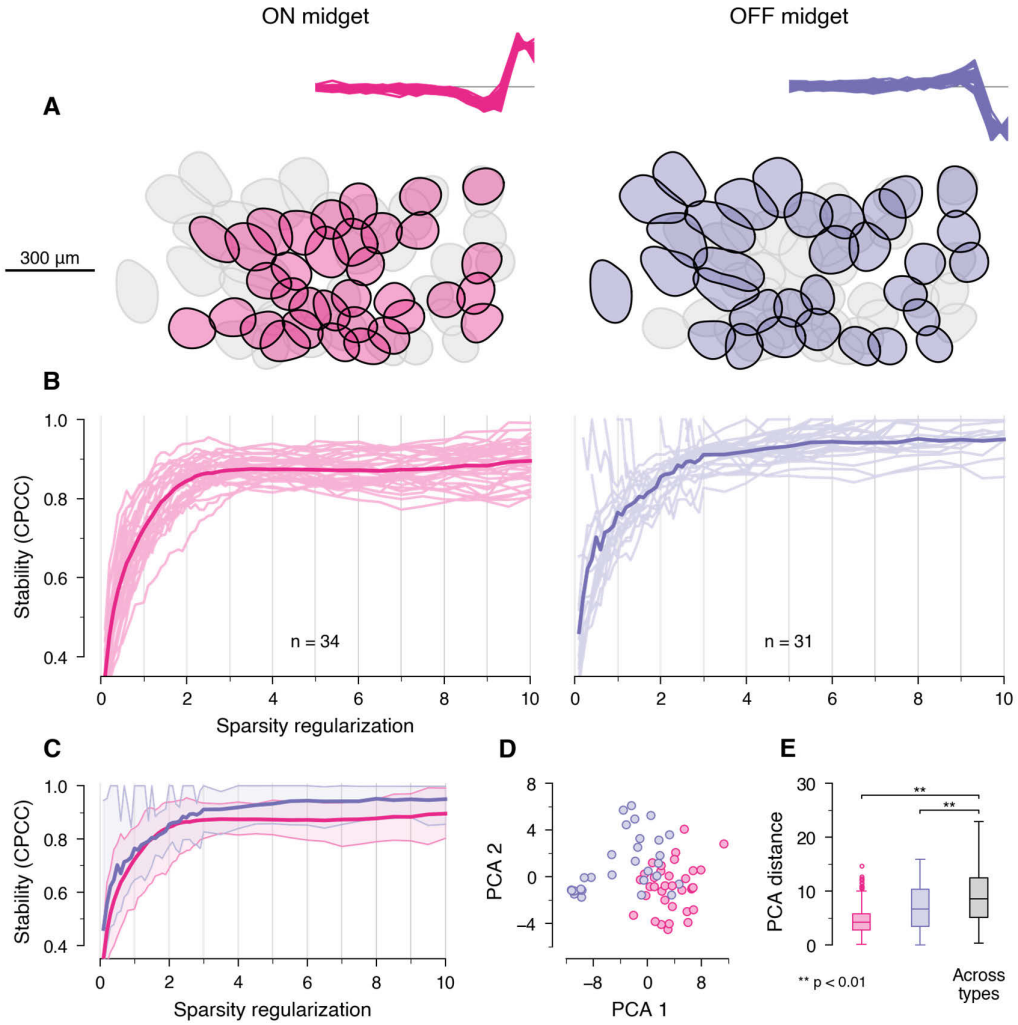


Figure 6 Consensus comparison of ON and OFF midget cells in marmoset retina. (A) Receptive field mosaic and temporal profiles of ON and OFF midget ganglion cells. (B) Stability curves show the expected increase and plateau (C) Superimposed stability curves of different cell types show slight differences (D) Stability curves projected on the first two principal components (E) Inter- and across-type distances in PCA space show significant differences across and within cell types.

Consequently, we suggest estimating the parameter value once on a subset of cells of one experiment, and applying it to subsequent analyses of recordings of similar conditions. This procedure speeds up the parameter selection process further and allows subunit recovery on population level in a matter of seconds. Nevertheless, there is no evidence that the similarity across experiments holds in general. We recommend at least verifying or fine-tuning the parameter selection briefly when moving to another dataset. Specifically so, should experiment conditions vary.

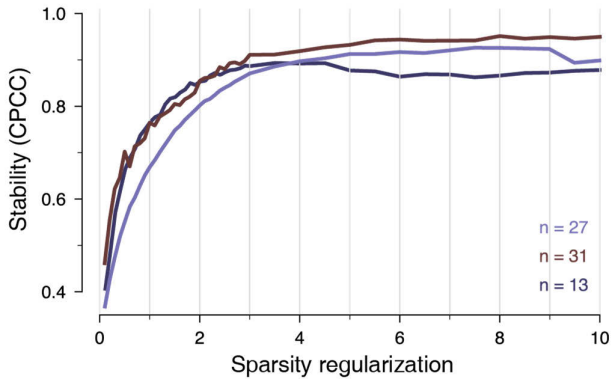
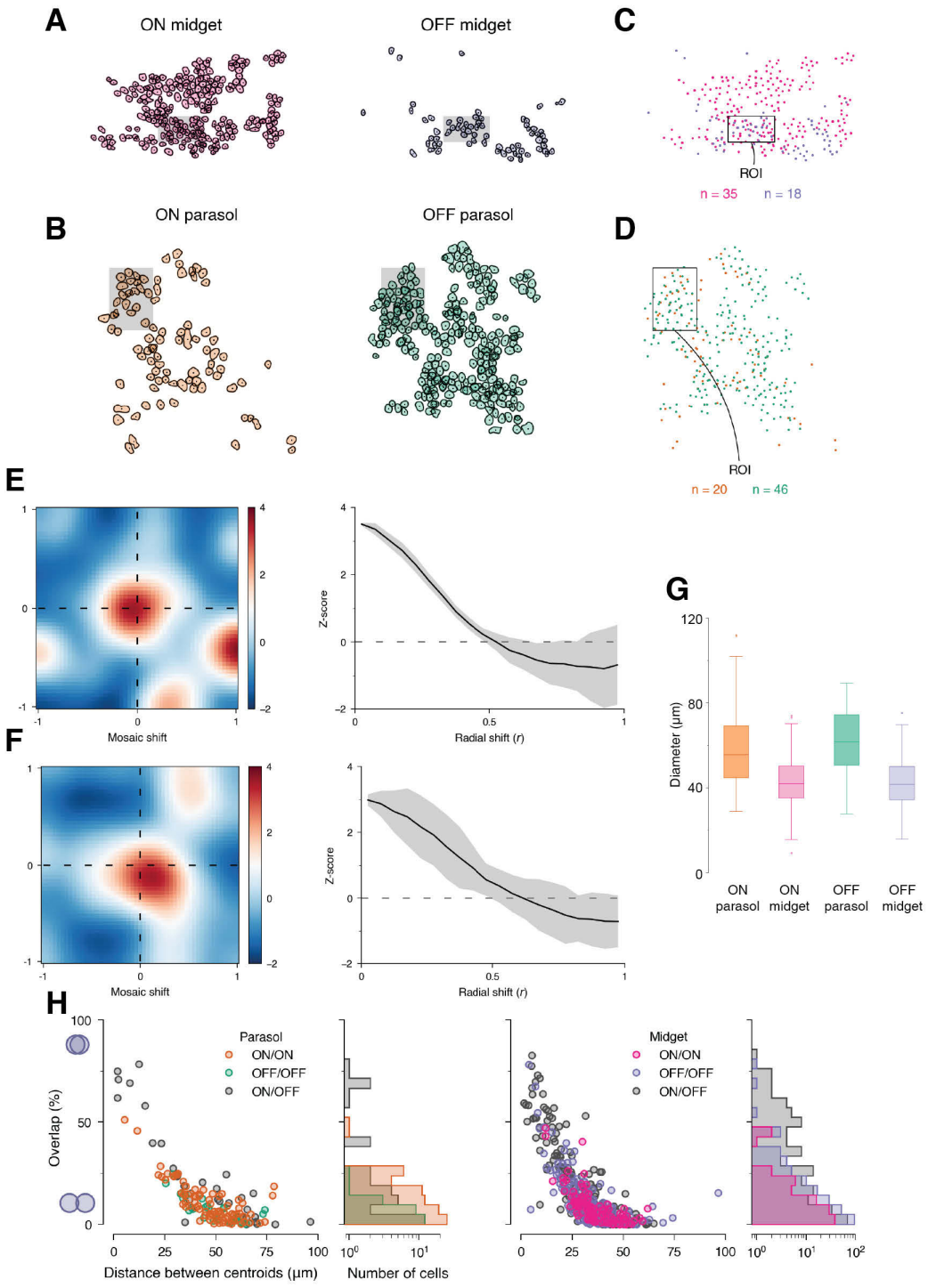


Figure 7 Superimposed stability curves of the OFF midget cells of three marmoset retinas. The curve and the number of cells from each recording is color-coded. The curves show a high similarity suggesting that the subunit decomposition is similar with respect to the applied sparsity.

Subunits arrange in mosaics by ganglion cell type

With the means to determine suitable amount of sparsity regularization per ganglion cell type and a computationally fast method of subunit inference, we were able to investigate subunit decomposition on cell population level. We ran STNMF on ON and OFF parasol cells and on ON and OFF midget cells from several marmoset retinas. We find that each set of subunits tile the receptive field by regularly spaced arrangement. Furthermore, this tiling extends beyond one receptive field. Subunits associated with ganglion cells of the same functional type tile the retina in a mosaic-like fashion (**Figure 8A**). The extent of individual receptive fields is visually not easily distinguishable from the subunits alone. Comparable sizes and distances are prerequisites for a uniform distribution. We find subunit diameters to be similar within cell types (**Figure 8G**). For uniform distances, we expect pairwise overlap of subunits to remain small. To verify, we visualize the relative overlap for each pair of subunits against the spatial distance of the subunit centers for parasol cells (**Figure 8H**) and for midget cells (**Figure 8I**). We observe few violations beyond 25% of overlap within ON-type pairs (ON/ON) and within OFF-type pairs (OFF/OFF). Unlike within type, subunits do not arrange regularly across ganglion cell types. Across parasol and midget types, the average subunit sizes differs (**Figure 8G**). Within parasol and midget types, we observe more overlap violations (overlap larger than 25%) in subunit pairs between ON- and OFF-type ganglion cells (**Figure 8H**, **I**). As the violations occur much more often across than within ON and OFF types, they suggest the existence of distinct ON and OFF subunit mosaics.

Figure 8 Mosaic analyses. (A), (B) subunit mosaics of ON and OFF midget and parasol ganglion cells from marmoset retinas. Center of mass (black markers) and regions of interest (shaded area) are indicated (C), (D) Center of mass mosaics with indicated region of interest, n = contained numbers of subunits (E), (F) IMCE topographical maps and radial average curves are z-scored and shift distance is normalized. (G) Subunit diameters for one marmoset experiment (H) Relative subunit overlap as a function of distance for parasol and midget mosaics. There are much more ON/OFF pairs overlapping beyond 25 %, (right) log-histograms emphasize the difference in overlap amount.


A
ON midget
OFF midget
C
ROI
n = 35
n = 18
B
ON parasol
OFF parasol
D
ROI
n = 20
n = 46
E
Mosaic shift
Z-score
Radial shift (r)
F
Mosaic shift
Z-score
Radial shift (r)
G
Diameter (µm)
ON parasol
ON midget
OFF parasol
OFF midget
H
Parasol
ON/ON
OFF/OFF
ON/OFF
Overlap (%)
Distance between centroids (µm)
Number of cells
Midget
ON/ON
OFF/OFF
ON/OFF

Subunit mosaics confirm anatomical differences between parasol and midget cells

Average diameters indicate differences across subunits of parasol and midget cells suggesting distinct presynaptic inputs. Parasol cells and midget cells in primates receive excitatory inputs from diffuse bipolar cells (Calkins and Sterling 2007; R. A. Jacoby et al. 2000; Wässle 1999) and midget bipolar cells (Jusuf et al. 2006a, 2006b; Kolb and Dekorver 1991), respectively. We find the average midget subunit diameter of around 44 µm to match the receptive field size of midget bipolar cells (Dacey et al. 2000). The average parasol subunit diameter of around 58 µm lies between reported anatomical dendritic tree sizes (30-50 µm) (Boycott and Wässle 1991) and measured receptive field sizes (74 – 114 µm) (Dacey et al. 2000) of defuse bipolar cells. Midget cells contain 5.5±2.1 subunits and parasol cells have 6.2±2.8 subunits (mean and standard deviation). As visible from the mosaic overlap (**Figure 8H**, **I**), we observe high relative overlap more often between parasol subunits than midget subunits within one mosaic. As the mosaic is otherwise regularly spaced, the high relative overlap (greater than 50%) between subunits of adjacent ganglion cell receptive fields appear as near identical duplicates on top of each other. As illustrated in an example of subunits of three parasol ganglion cell receptive fields (**Figure 9**), the overlapping subunits integrate well into the mosaic tiling. A possible interpretation is that parasol ganglion cells exhibit the same subunits suggesting shared presynaptic input. The fact that we find these overlaps among parasol mosaics but not for midget cells, matches anatomical findings. Dendritic trees of midget ganglion cells do not overlap independent of eccentricity, while parasol cells do towards the periphery (Dacey 1993; B. B. Lee et al. 2010).

ON- and OFF-type subunit mosaics may be aligned

We investigated the interdependence of the subunit mosaics between ON and OFF types. We followed the classification into three groups of coding regimes as proposed previously (Roy et al. 2021). There, pairs of ON- and OFF-type mosaics of retinal ganglion cells were classified as being either aligned, anti-aligned, or independent. This revealed insights into the optimization of encoding of visual scenes on a population level. We applied the method to our subunit mosaics, specifically to the pair of ON- and OFF-type mosaics of midget cells (one recording) and parasol cells (one recording), following the proposed procedure closely. We reduced the contour of each subunit to a single coordinate point, defined by its center-of-mass (**Figure 8C**, **D**). The inter-mosaic coordination energy (IMCE) describes the pairwise spacing between centroids across mosaics (Roy et al. 2021). The IMCE decreases with distance between pairs. The class of alignment of the pair of mosaics is determined by measuring the IMCE as one mosaic is shifted vertically and horizontal on top of the other. For aligned mosaics, subunits are close to one another and the energy is initially high and decreases with shift. For anti-aligned mosaics, the energy rises with subunits shifting closer to each other. There is no distinct increase or decrease for independent mosaics. We visualize the IMCE for different mosaic offsets as a topographic map (**Figure 8E**, **F**). For both midget and parasol mosaics the energy is highest at zero shift. The z-scored radial average of the IMCE is decreasing with mosaic shift in both cases (**Figure 8E**, **F**). This indicates that the ON- and OFF-type subunit mosaics are aligned. The mosaics are partially depleted due to missing receptive fields possibly related to lesions on the retina during the experiment. Although the procedure is robust to subsampling to certain extend (Roy et al. 2021), we manually chose a region of interest (ROI) in both cases, in which we find few gaps in the tiling for both ON- and OFF-type mosaics (**Figure 8C**, **D**). The tested portion

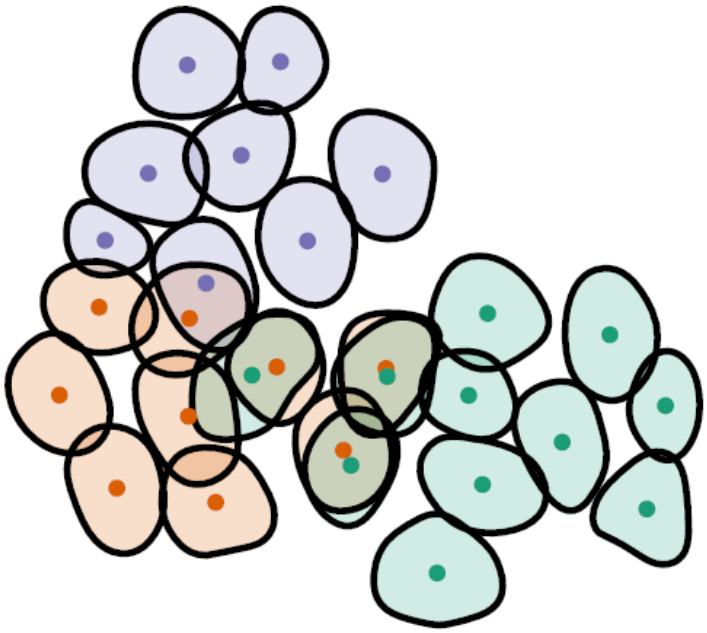


Figure 9 Overlapping subunits across three adjacent OFF parasol cells (blue, orange, green) in marmoset retina. Markers are center of mass. Three subunits are tightly overlapping with the center of mass being very close to each other. This can be interpreted as identical subunits recovered from different ganglion cells.

of the mosaic pairs include around 20-40 subunits each. To rule out effects of the reduced mosaics, we confirmed the course of the IMCE with a larger ROI (**Figure 8E**, **F**) defined by the convex hull of subunit centroid pairs excluding outliers, as proposed (Roy et al. 2021). Although not as pronounced (as expected from depleted mosaics), the energy remains sharply decreasing.

STNMF is available as software-package for Python

We provide the implemented software as Python-package. It includes the NMF algorithm and initialization procedures optimized with numpy for multi-threaded CPU computation. Tools for visualization are also available, as well as, routines for the consensus analysis and custom extensions via callback functions. The software, along with an extensive documentation with example code, will be made available upon the publication of this manuscript.

DISCUSSION

We propose a method of subunit recovery based on STNMF that finds subunits within a receptive field in a matter of seconds. Through purpose-built initialization it is steered towards a suitable solution and by means of a tailored NMF algorithm, it is scalable to population recordings with respect to reasonable analysis time and computational demand. By adjusting consensus methods to our needs, we demonstrate a viable method for hyperparameter selection. Our findings suggest that the selection of the sparsity regularization parameter based on a subset of neuronal cells is transferable to cells of the same functional type and may serve as estimate across datasets. Subsequent analyses for cells of the same type reduce to a parameter-free procedure of plug-and-play. We have shown that this may extend to datasets recorded under the same experimental conditions.

Recovered subunits may correspond to bipolar cells

We apply the method to data previously analyzed (Liu et al. 2017) and provide the same findings under a more than 100-fold increase in speed. Furthermore, we demonstrate the method's relevance for biological application with selected subunits matching the receptive fields of bipolar cells in data of combined recordings of ganglion cells and bipolar cells (Liu et al. 2017).

The differences we found in subunits of parasol and midget cells in recordings from marmoset retinas agree with the anatomy. Most strikingly, we found overlapping subunits among adjacent parasol cells but not in midget cells. Unlike midget cells, parasol cells have overlapping dendritic trees and are more likely to share presynaptic inputs (Dacey 1993; B. B. Lee et al. 2010).

There are suggestions that functional subunits of primate parasol cell receptive fields correspond to diffuse bipolar cell inputs (Crook et al. 2008, 2014). Since our subunits from marmoset do not exceed receptive field and dendritic tree size estimates of macaque diffuse bipolar cells (Boycott and Wässle 1991; Dacey et al. 2000), it is unlikely that the subunits are combinations of multiple bipolar cell inputs. To the extent of our comparison of midget and parasol cells, the retina of marmoset and macaque are reported to be functionally identical (B. B. Lee et al. 2010). Nevertheless, we abstain from absolute size comparisons. Our findings concerning subunit density, uniform tiling of the retina, and potentially shared subunits across parasol cell dendritic trees (B. B. Lee et al. 2010) support the notion that the subunits may correspond to individual bipolar cell inputs. Nevertheless, that input represents the receptive field of the diffuse bipolar cell, that integrates a high percentage of amacrine cell input (R. Jacoby et al. 1996; Kolb and Dekorver 1991).

We cannot make a comparable case for midget cell subunits. Midget ganglion cells receive inputs from an average of eight midget bipolar cells (mid-periphery) to on average 13 midget bipolar cells (periphery) in the marmoset retina (Jusuf et al. 2006b). We find far fewer subunits per midget ganglion cell. Yet, subunits of midget cells form a retinal mosaic, too. Furthermore, the receptive field of midget bipolar cells appears to be constructed primarily through circuitry in the outer retina (Dacey 2000; Dacey et al. 2000), which could facilitate functional decomposition, one could expect. Due to variations in absolute anatomical sizes, the subunits could still be bipolar cell receptive fields. To what extent STNMF subunits relate to midget bipolar cells is subject to further investigation.

Subunits arrange in a uniform mosaic by cell type

Our subunit analysis with STNMF revealed distinct mosaics of subunits for ON and OFF parasol, and ON and OFF midget cells in the marmoset retina, as previously demonstrated for mosaics of parasol and midget ganglion cells in marmoset (Szmajda et al. 2005). The distinct subunit mosaics are in line with ganglion cell types stratifying in different levels of the inner plexiform layer (Watanabe and Rodieck 1989) and receiving input from distinct bipolar cell types (Masland 2001; Wässle 2004). The observed uniformity of the subunit mosaics matches reports of bipolar cell input density to be independent to the size of the dendritic tree of the ganglion cell in the marmoset retina (Eriköz et al. 2008). Taking these similarities into account, it is likely from a population perspective that the recovered subunits relate to bipolar cell inputs, albeit they may be modulated by gab junctions (Kolb 1979).

There is recent evidence that the mosaics of ON and OFF parasol ganglion cell receptive fields are spatially coordinated (Roy et al. 2021). Data from macaque retinas suggest that their mosaics are anti-aligned. Models of efficient coding indicate that anti-alignment arises in the presence of noise and provides more reliable encoding. Increased resolution is an additional benefit of anti-alignment and it may have advantages for down-stream convergence (Roy et al. 2021). We find that subunits of ON- and OFF-type ganglion cells form distinct mosaics. Subunits converge in their respective ganglion cell type only, without any pooling across the mosaics. That makes the mosaic relationship twice removed. Consequently, we expected one of two forms of inter-mosaic coordination. On the one hand, the ON and OFF pathways may already organize as in the ganglion cell layer. We would then expect the subunit mosaics to be anti-aligned to share the benefits outlined for ganglion cell receptive field mosaics. On the other hand, the subunit mosaics may merely map the visual space for the ganglion cell receptive fields to form coordinated mosaics downstream. That would not require the subunit mosaics to be spatially coordinated and we would expect them to be randomly arranged and thus independent. Both options seem reasonable from an efficient coding and developmental perspective. If subunits were random and did not correspond to a biological counterpart, independent mosaics would be most likely. That would also be expected for inconclusive data and the radial average of the IMCE would be flat as a function of mosaic shift. However, the distinct decline in energy indicates that the ON- and OFF-type subunit mosaics of the tested midget and parasol cells are aligned. Data from two retinas are not conclusive to claim that alignment of subunit mosaics is the norm. Nevertheless, we demonstrate that there are instances of spatial coordination between subunit mosaics and that mosaics are not random. Conditions of stimulation like the light level could have effect on the mosaics. Future investigation is necessary to confirm these findings. If we assumed alignment to be the norm, we would speculate that the outlined constraints under which ganglion cell receptive field mosaics form as anti-aligned do not hold at the level of bipolar cells. Due to their finer mosaics, anti-alignment may offer little additional benefit in spatial resolution. The accumulated noise at bipolar cell level is lower than after spike generation in ganglion cells (M. J. Berry et al. 1997). Furthermore, anti-alignment may not provide advantages for integration in the ganglion cells, as they only pool from one of the two mosaics. Developmentally, alignment of mosaics may be energy efficient.

Accelerated Fast HALS with sparsity regularization yields accurate solutions

Solving a non-negative least squares problem (NNLS) sits at the core of non-negative matrix factorization. To solve it accurately, typically, costly algorithms are employed iteratively (J. Kim and Park 2011). These include multiplicative updates (Lee & Seung, 1999), the projected quasi-Newton method (D. Kim et al., 2007), the projected gradient method (Lin, 2007) and non-negative quadratic programming (Zdunek & Cichocki, 2008), and the active-set method (Lawson & Hanson, 1974), to name a few. The algorithms strive to provide exact solutions at each iteration by taking into account all data at once. Here, we avoid the high computational demand with Hierarchical Alternating Least Squares (HALS) (Cichocki et al. 2007; Ho 2008; L. Li and Zhang 2009) by solving smaller, non-restricted least squares problems followed by setting all negative values to zero. While Fast HALS (Cichocki and Phan 2009) simplifies computational operations for speed, A-HALS (Gillis and Glineur 2012) refines the coarse approximations of HALS by performing multiple rounds of updates within one iteration. The intended purpose of A-HALS is to accelerate the convergence by reducing the number of necessary iterations, which lends it its name. However, by incorporating sparsity regularization, we make use of another advantage of this procedure. Additional constraints like sparsity regularization are typically integrated into the equations of the iterative update (H. Kim and Park 2007; Liu et al. 2017). As we perform it multiple times instead, that is, at each A-HALS repetition within one iteration, the continuously applied sparsity constraint refines the subsequent repetitions by giving more weight to prominent structures and progressively suppressing the effect of uncorrelated noise. The result after one iteration offers high accuracy and sparsity. With AF-HALS with sparsity regularization, we introduce a valuable tool in the estimation of subunits and for the field of NMF.

Sparsity regularization is crucial

In the field of machine learning, sparsity regularization has become an important tool to combat overfitting and to generalize from limited available data. Thus, it has earned the role of a rather non-essential, optional enhancement. However, sparsity has proven to be a central player in signal processing and learning representations (Candès et al. 2006; d'Aspremont et al. 2007; Donoho 2006; Hoyer 2004). Sparsity in NMF offers interpretable solutions (d'Aspremont et al. 2007) and even inherently reduces the impact of noise by decreasing the effects of uncorrelated patterns in inputs (Donoho 1995; Elad 2006; Hyvärinen 1999). Aside from these benefits, sparsity is an integral part of STNMF. It replaces the dropped non-negativity constraint of semi-NMF (Hoyer 2002, 2004; H. Kim and Park 2007; J. Kim and Park 2008) as the parts-based reconstruction collapses without (Ding et al. 2010). In the context of STNMF, sparsity is the vital component producing localized subunits. Neglecting sparsity regularization for simplicity or model comparison is therefore contradictory.

Beyond simple $\ell 1$-norm-based sparsity, there have been efforts towards regularization techniques that favor localized solutions in the fields of NMF and clustering. LocalNMF has demonstrated successful decomposition of widefield calcium imaging signals into localized functional brain regions (Saxena et al. 2019). The locality penalty added to semi-NMF requires localized regions to be defined a-priori. Consequently, it is not suitable for the recovery of an unknown number and shape of subunits. Locally normalized $\ell 1$ regularization was introduced with Spike-Triggered Clustering (Shah et al. 2020) to instill spatial awareness into the $\ell 1$-regularization in subunit estimation. The

authors adjust the penalty of sparsity regularization by taking into account spatially neighboring values. For our application, the forced locality is not suitable, as it makes a module either contain a localized pattern or exclusively zero values. Noise in the input cannot escape to excess modules and the choice of module number is no longer a mere upper bound, a strength of STNMF. As a result, subunits are forced to oversplit to accommodate for the number of modules. This introduces a prior bias that could possibly skew the solution. We find that the $\ell 1$-norm penalty suffices for STNMF, while it also keeps the assumptions on the data at a minimum. We attribute the success of STNMF to the refined sparsity constraint in the repetitions of AF-HALS. The original definition of alternatingly solved sparse semi-NMF applies the sparsity constraint on the pixel contributions across modules (H. Kim and Park 2007) as implemented in previous implementations of STNMF (Liu et al. 2017), whereas here, we regularize the modules. This resonates with local non-negative matrix factorization (S. Z. Li et al. 2001), that argues for a localized decomposition instead of a sparse one.

Parameter selection by consensus analysis

To find a suitable amount of sparsity regularization, we investigate how robust the decomposition is with respect to varying initialization. To measure how stable the decomposition is, we here adjust consensus methods originating from Brunet at al. (2004) to be applied to the problem of identifying regularization parameters. An alternative to using consensus analysis is L-curve fitting. The L-curve technique (Hansen 1999a, 1999b) aids in determining a suitable parameter for a regularized least squares problem. An advantage over our proposed consensus analysis is the conceptual simplicity and that it does not require multiple repetitions of NMF. It is applicable to sparse NMF, as the update of one factor is a regularized least squares problem, and has been applied to HALS previously, although for a different type of regularization (Cichocki and Phan 2009). Nevertheless, when applied to STNMF, we did not find a sharp corner with the data available and also the dedicated method of L-corner fitting (Hansen 1994) did not succeed. With the consensus analysis, we have found a good alternative.

NNSVD-LRC offers reliable initialization

We demonstrate that a single run of STNMF with SVD-based initialization can replace the selection of the most suitable solution among 100 randomly initialized runs. Specifically, NNSVD-LRC is advantageous as it converges faster to a stationary point due to low error and increased sparsity in the initialization (Atif et al. 2019). Notably, on the one hand, techniques based on SVD do not guarantee NMF to find the global optimum (Atif et al. 2019). On the other hand, however, seeking the global optimum in network structures bears the risk of overfitting (Choromanska et al. 2014). Furthermore, the alternative of manual selection of a most suitable among many randomly initialized solutions adds two degrees of arbitrariness. Firstly, the selection of a best solution is subject to human bias. Secondly, a measure for the most-suitable solution is not easily defined. These problems may be alleviated by an adapted consensus clustering procedure (Zhou et al. 2020). Instead of choosing the best decomposition among 100 randomly initialized NMF runs, their modules are concatenated to serve as the input to another NMF run that yields a final set of modules. We found that decompositions initialized with NNSVD-LRC are comparable with solutions of that approach while requiring only one run of NMF.

Limitations

We identify three limitations of the proposed method. Firstly, the method poses strict constraints on the stimulus that the spiking data is produced from. To infer subunits from pixel correlations in the spiking response, a white noise stimulus with uncorrelated pixel contrast across space and time is necessary. This demands to allocate time to a specific stimulus during the experimental recording. Nevertheless, we were able to recover subunits successfully from less than one hour of white noise, which may be available already for means of characterizing the STA. Another caveat is the pixel size of the stimulus. Although STNMF makes few assumptions on the data, the stimulus design must provide stimulus pixels of suitable size to fit bipolar cell receptive fields of the species and location on the retina. This is usually facilitated by dendritic size estimations from the literature or approximations from measured receptive field size. The third drawback is the inability to produce subunits of both positive and negative sign due to the nature of NMF. While arbitrary sign of the STNMF weights allows the recovery of positive and negative subunits within one receptive field, inhibitory structures within one subunits cannot be represented as such. We speculate that inhibitory structures could be recovered as separate subunits of their ON and OFF parts, given that the inhibitory elements are strong enough. Bearing these limitations in mind, we are convinced that STNMF in the proposed form offers reliable and fast subunit inference under suitable experimental conditions.

Comparison to other methods of subunit inference

Popular computational methods of subunit inference in recent years are linear-nonlinear cascade models (Maheswaranathan et al. 2018; Real et al. 2017), approaches of convolutional neural network models (Maheswaranathan et al. 2019; Tanaka et al. 2019), and methods of statistical inference (Shah et al. 2020). Linear-nonlinear-linear-nonlinear (LNLN) models consist of a layer of linear filters of the spatial or spatiotemporal subunits with nonlinear transfer functions that converge into another nonlinear transfer function to produce a firing rate or spiking response. Recent work demonstrates how the filters and nonlinearities are trained in a supervised manner on the stimulus input and spiking response (Maheswaranathan et al. 2018). This implementation is a generative model for predicting the spiking response of neurons using held-out data during training. It was successfully applied to data from retinal ganglion cells of salamanders. To combat complexity, sparsity regularization was applied as well as restricting the stimulus to one spatial dimension. Training may struggle with parameters for two spatial dimensions as dimensionality allows for many degrees of freedom in the learned representations. Similarly, convolutional neural networks are trained end-to-end. The models are trained on a population of neurons and provide convolutional filters of subunit types rather than individual subunits for each neuron. With data from many neurons it was demonstrated that they are able to generalize well enough to be trained with natural images instead of artificial stimuli like white noise or flickering bars (Maheswaranathan et al. 2019; Tanaka et al. 2019). The models revealed interesting properties similar to those of bipolar cells and amacrine cells of the primate retina, whereas it remains unclear to what extent the model architecture resembles the neural circuitry of the used primate retinas. Although Spike-Triggered Clustering recovers subunits on stimulus correlations much like STNMF, it was demonstrated to be capable to predict the spiking-response to natural scenes (Shah et al. 2020). Subunit number and regularization are determined though cross-validation.

A common trait of the state-of-the-art subunit inference techniques is their high demand on computational power and time. The training of parameters of a convolutional neural network require powerful GPUs and hours of training. Extensive cross-validation to find the most suitable number of subunits is a reoccurring obstacle as it is time-consuming and has to be performed for every neuron as the number of subunits may vary. In direct comparison, our implementation of STNMF offers an analysis of only a few seconds on a conventional office CPU (all analyses have been carried out on computers with an Intel Xeon CPU E3-1270 v5 @ 3.60GHz) and inherently provides the number of subunits automatically. Hyperparameter search is reduced to a subset of neurons for each cell type making a subunit population analysis considerably faster than previously available.

ACKNOWLEDGMENTS

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 724822) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 432680300 (SFB 1456, project B05).

AUTHOR CONTRIBUTIONS

SJZ developed and implemented the novel variant of STNMF, collected and curated the marmoset retina data, analyzed the data, and prepared figures and the initial draft of the manuscript. DK, MHK, HMS, MM, and DAP all contributed to the preparation of the marmoset retina and data collection. TG supervised the project and contributed to writing the manuscript. All authors helped revise the manuscript.

DISCLOSURE STATEMENT

The authors declare no competing interests.

SOFTWARE AVAILABILITY

The STNMF software will be made available upon the publication of this manuscript.

METHODS

Ethics statement

All experimental procedures were performed in strict accordance with national and institutional guidelines. The salamander data stem from a publically available repository (Gollisch and Liu 2018) and had been described and analyzed in a previous publication (Liu et al. 2017). Recordings from marmoset retinas were performed on tissue obtained from animals used by other researchers, as approved by the institutional animal care committee of the German Primate Center and by the responsible regional government office (Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit, permit number 33.19-42502-04-17/2496).

Electrophysiology

The salamander data stem from recordings as described previously (Liu et al. 2017) and made available in a public repository (Gollisch and Liu 2018). Isolated axolotl (*Ambystoma mexicanum*) retinas were provided with oxygenated Ringer's solution at around 22°C. Ganglion cell spikes were recorded extracellularly with 252-channel planar multielectrode arrays (MEA) and extracted based on a sufficient refractory period. In combined recordings of salamander retinas, ganglion cell spikes were recorded with 60-channel MEA and sharp microelectrodes recorded intracellularly from bipolar cells (Liu et al. 2017).

Data from common marmoset (*Callithrix jacchus*) were recorded from three animals of different sex and age (female 2.3 years, male 18 years, and male 13 years). The eyes were removed by the German Primate Center Göttingen (DPZ) between three and twelve minutes after time of death. Left and right eye were labeled and dark adapted for varying durations of at least one hour. The retina was isolated and cut into pieces classified at time of dissection by their quadrant (nasal, superior, and dorsal) and eccentricity (central, mid-periphery, and periphery). The temporal quadrant was not considered due to its high density of cell bodies and axons at the fovea and optic disc. Around six pieces per quadrant were recorded while the remaining quadrants were left in oxygenated Ames' solution for later recording. The pieces (one mid-periphery, two periphery) were placed on MEAs (MultiChannel Systems, Reutlingen, Germany) of different configurations. On the 60-channel MEA (10-µm electrode diameter, 100-µm electrode distance) the piece was held fixed with 50 µL of poly-D-lysine (PDL) (Millipore, 1 mg/mL) that coated the array for one or two hours prior to mounting. The 60-channel perforated MEA (30-µm electrode diameter, 100-µm electrode distance) contains holes between the electrodes which held the piece fixed by slight suction from underneath. Both techniques were used to maximize the diversity of recorded function cell types. Two experiments were conducted with the perforated MEA. On the MEA, the piece was perfused with oxygenated Ames and kept at a temperature of around 34°C. Bandpass-filtered voltage signals (300 Hz to 5 kHz) from each electrode were sampled at 25 kHz. Spikes were extracted using Kilosort (Pachitariu et al. 2016) adjusted to accommodate for MEA data (https://github.com/dimokaramanlis/KiloSortMEA). The automated extraction was followed by visual inspection and curation using Phy2 (https://github.com/cortex-lab/phy). Clusters that displayed good separation from other clusters as well as a uniform spike waveform and a clear refractory period were considered as representing individual ganglion cells.

Visual stimulation

Spiking activity was recorded in response to visual stimulation. Stimuli were projected onto the photoreceptor side of the retina using a gamma-corrected white-light OLED projector (eMagin, USA) at 800 × 600 square pixels with a refresh rate of 60 Hz (salamander and marmoset) or 85 Hz (marmoset). The pixel size (width and height) translated to 2.5 µm or 7.5 µm on the retina depending on the configuration. The projector was controlled by a custom-built software based on C++ and OpenGL to generate the stimulus frames. The spatiotemporal white-noise stimulus was presented with a pixel size of 15 µm on the retina (30 µm for salamander) and temporally updated at 30 Hz (salamander), 20 Hz, or 21.25 Hz (marmoset). The light intensity of each stimulus pixel was drawn independently from a binary distribution with 100% contrast and a mean light level of about 2.5 mW/m^2 (salamander), 3.33 mW/m^2, 4.38 mW/m^2, or 5.45 mW/m^2 (marmoset). For analysis, the stimulus was described in contrast values of +1 (bright) or –1 (dark) stimulus pixels. In some experiments, the non-repeating stimulus was interleaved with portions of identical stimuli intended as held-out data for model validation. The held-out data was not considered in this work. The recording durations range between 43 minutes and three and a half hours of non-repeated stimuli (3:28, 0:43, and 1:42 hours).

Receptive field analysis

The receptive field of a given ganglion cell was estimated using reverse correlation to obtain the spike-triggered average (STA) (Chichilnisky 2001). The stimulus frames within the 700 ms time window preceding a spike were collected and averaged across all spikes. The STA summarizes the spatiotemporal dynamics of the receptive field. To describe the temporal filter of the STA, we selected the pixel that exhibits the maximum absolute value across all STA frames. We defined the temporal filter as the value of that pixel across the STA frames normalized to unit Euclidean norm. The spatial profile was the stimulus frame of the STA with the highest intensity pixel value.

Spatial parameterization

Receptive field diameter was estimated by fitting a two-dimensional Gaussian function to the spatial STA. The effective diameter $d = \sqrt{ab}$ was obtained from the major axis a and minor axis b of the ellipse (Liu et al. 2017) at two standard deviations. Elliptical outlines to the localized subunits were defined by the same procedure.

For inspection of subunit shape and tiling, contour outlines of the receptive field and subunits were produced. The image of each spatial filter was first up-sampled in nearest neighbor fashion to micron resolution and then smoothed with a Gaussian filter with a standard deviation appropriate to the spatial size (30 µm for receptive fields and 8 µm for subunits in marmoset and 20 µm for subunits in salamander). The contour was taken at a level of 1 sigma of the peak.

Functional cell type classification

Ganglion cells from marmoset recordings were clustered by their responses to a moving bar stimulus based on a method proposed previously (Drinnenberg et al. 2018). The barcode stimulus presented vertical bars consisting of superimposed sine waves with a period between 12750 and 120 µm at

phase shifts sampled at random from a uniform distribution. The pattern was moved horizontally from left to right at a constant speed of 1275 µm/s with a length of 12750 µm. The screen filled with mean-intensity gray behind the pattern before it was repeated 22 times. The repeated presentations formed trials to construct peri-stimulus time histograms (PSTH) of the cells' spiking responses. Correlations in the PSTHs across cells over time revealed similarities among cells irrespective of the spatial location of their receptive field. Clustering the correlations into five groups was complemented with the spatial diameter and the temporal dynamics of the STA of each cell as obtained from spatiotemporal white noise. The clusters were manually labeled as ON and OFF parasol and ON and OFF midget cells based on the sign and shape of their temporal profiles, and the diameter of the receptive field with an additional fifth unlabeled cluster of unclassified cells excluded from the analysis. The barcode analysis was not available for one experiment. There, the ganglion cells were clustered using K-means clustering on the projection of the first principal component of the temporal filters and the normalized receptive field diameter. Manual curation followed as described above to classify ON and OFF parasol and ON and OFF midget cells.

Spike-triggered stimulus ensemble

The spatial region of the stimulus relevant to a given ganglion cell's response was defined as the smallest rectangular window, $x \times y$, that encloses three standard deviations of the Gaussian fit of the receptive field (see above). The stimulus frames within this spatial window form the stimulus ensemble of that cell. The subset of stimulus frames that lie within the 700 ms time window preceding a spike is the spike-triggered stimulus ensemble. Frames in each spike-triggered stimulus were weighted by the temporal filter of the STA and then summed to one effective stimulus frame per spike. The effective stimuli of all spikes m construct the effective spike-triggered stimulus ensemble (STE) such that its average across spikes yields the spatial profile of the STA. The stimulus frames of the STE with $n = x \times y$ pixels were flattened to column vectors to form the STE matrix with the dimensions $n \times m$, with n pixels by m spikes.

Semi-non-negative matrix factorization

Semi-non-negative matrix factorization (semi-NMF) decomposes a matrix $\mathbf{V} \in \mathbb{R}^{n \times m}$ into two factor matrices $\mathbf{W} \in \mathbb{R}_{+}^{n \times r}$ and $\mathbf{H} \in \mathbb{R}^{r \times m}$ with $r < m, n$, such that

$$\mathbf{V} \approx \mathbf{WH}.$$

To relate these variable names common to NMF terminology to STNMF, $\mathbf{V}$ is the STE, where each column $v_i \in \mathbb{R}^n$ is the effective stimulus frame of spike $i \in 1, \ldots, m$. The columns of $\mathbf{W}$, $w_k \in \mathbb{R}^n$, are the $k \in 1, \ldots, r$ spatial modules, and each row of $\mathbf{H}$, $\underline{h_i} \in \mathbb{R}^r$, contains the corresponding r weights for each spike $i \in 1, \ldots, m$. The stimulus v_i that triggered spike i is approximated by the linear combination of the modules $\mathbf{W}$, as basis vectors, and the corresponding weights, $\underline{h_i}$, as coefficients or encodings. The product of $\mathbf{W}$ and $\mathbf{H}$ forms a lower dimensional representation of $\mathbf{V}$ based on recurring structures within $\mathbf{V}$. This gives rise to the basis spatial modules $\mathbf{W}$ among which we find the resulting subunits.

Semi-NMF is an NMF variant that relaxes the non-negativity constraint of $\mathbf{V} \in \mathbb{R}_{\pm}$ and $\mathbf{H} \in \mathbb{R}_{\pm}$, while $\mathbf{W} \in \mathbb{R}_{+}$ remains non-negative. $\mathbf{W}$ thereby consists of additive terms only and individual pixel

values do not cancel out in the linear combination of modules. To retain parts-based factorizations in semi-NMF, the loosened non-negativity constraint has to be replaced (Ding et al. 2010). Typically this is achieved by sparsity regularization on **W** (Hoyer 2002, 2004; H. Kim and Park 2007; J. Kim and Park 2008). Here, this resulted in the use of sparse semi-NMF.

NMF generally iteratively improves the approximation **WH** by minimizing an objective function. As it is a common choice for NMF, we chose the Euclidean distance between **V** and the reconstruction **WH**, because it is simple to differentiate (D. D. Lee and Seung 1999). Sparsity was implemented with an $\ell 1$-norm penalty on each module as an additional term in the objective function, weighted by a regularization parameter λ

$$\frac{1}{2}\|\mathbf{V}-\mathbf{WH}\|_F^2+\lambda\sum_{k=1}^{r}\|w_k\|_1, \text{ subject to } \mathbf{W}\geq 0,$$

where w_k for $k=1,\ldots,r$ are the columns of **W**, $\|\cdot\|_F^2$ is the squared Frobenius norm and $\|\cdot\|_1$ is the $\ell 1$-norm.

To distinguish the localized subunits from noisy modules in **W**, the spatial autocorrelation was calculated using Moran's I (H. Li et al. 2007; Moran 1950). Given module w_k for $k=1,\ldots,r$ as the vector $s\in\mathbb{R}^n$, the autocorrelation was computed with

$$I=\frac{n}{\sum_i\sum_j l_{ij}}\frac{\sum_i\sum_j l_{ij}(s_i-\overline{s})(s_j-\overline{s})}{\sum_i(s_i-\overline{s})^2},$$

where $i,j=1,\ldots,n$ are pairs of pixel indices, $\overline{s}$ is the mean of vector s, and $\mathbf{L}\in\mathbb{R}^{n\times n}$ is the neighbor matrix with l_{ij} equal to unity if i and j correspond to spatially adjacent pixels in the layout $n=x\times y$ and zero otherwise. The scalar I ranges from negative to positive unity with higher positive values denoting denser localization in space. We chose an autocorrelation threshold of at least $I=0.25$ to distinguish localized subunits from noise modules.

Accelerated fast hierarchical alternating least squares

The objective function was minimized by alternating optimization of the factor matrices. **H** was updated while **W** was held fixed and **W** was subsequently updated while the newly computed **H** was held fixed. This process was repeated until a termination criterion was reached. We chose a fixed number of 1000 iterations (5000 for selected checks) as most cells showed convergence much earlier. The iterations concluded with a final update of **H** for a more accurate final reconstruction.

The alternating updates of **W** and **H** split the optimization into two sub-problems. In the context of semi-NMF, **H** is not constrained to be non-negative. Given the Euclidean distance objective function, the update of **H** reduces to a least squares problem (Ding et al. 2010) and can be solved accurately using the Moore-Penrose inverse or pseudoinverse $(\cdot)^\dagger$

$$\mathbf{H}\leftarrow\mathbf{W}^\dagger\mathbf{V}.$$

We followed it with an $\ell 2$-normalization of the rows of **H**. This normalization step is generally performed in Euclidean-distance based NMF to prevent exploding coefficients during the alternative updates of **W** and **H** (Ho 2008). The update of the non-negative matrix **W**, on the other hand, poses

a non-negative least squares (NNLS) problem, in fact a regularized NNLS in the case of sparse semi-NMF. We approximated it by updating the modules sequentially with the local update for w_k for $k = 1, \dots, r$

$$w_k \leftarrow [w_k + (\mathbf{V}\mathbf{H}^\top)_k - \mathbf{W}(\mathbf{H}\mathbf{H}^\top)_k - \lambda \mathbf{1}_n]_+,$$

where w_k and $(\cdot)_k$ are the kth column of the continuously updated $\mathbf{W}$ and the matrix multiplication $(\cdot)$, respectively, $\mathbf{1}_n \in \mathbb{R}^n$ is a column vector of length n containing all ones, and $[\cdot]_+ = \max\{\cdot, 0\}$ is the elementwise maximum with zero of a matrix. This update rule is based on Fast HALS (Cichocki and Phan 2009) with additional sparsity regularization. Fast HALS simplifies the update equation of HALS (Cichocki et al. 2007) and results in a significant speed up (Cichocki and Phan 2009). It drops terms from the equation facilitated by the rows of $\mathbf{H}$ being $\ell 2$-normalized prior to the update. Fast HALS takes advantage of this normalization step.

While Fast HALS improves the local update rule on the scale of one module, A-HALS (Gillis and Glineur 2012) optimizes the computational cost on the scale of all modules in $\mathbf{W}$. While $\mathbf{W}$ is updated continuously, A-HALS exploits the fact that the terms $\mathbf{V}\mathbf{H}^\top$ and $\mathbf{H}\mathbf{H}^\top$ in the local update remain unchanged over the course of iterating over the modules. These matrix multiplications outweigh the other terms of the local update rule in number of floating-point operations, making the update of $\mathbf{W}$ costly compared to the outcome. A-HALS wraps an outer loop around the local updates, effectively updating $\mathbf{W}$ multiple times per one update, in order to tradeoff the floating-point operations in the pre-computed matrix multiplications and in the local updates. This increases the approximation at comparably little additional cost and improves the accuracy of a single update of $\mathbf{W}$.

To determine the number of outer iterations, we found the previously suggested parameters to be suitable (Gillis and Glineur 2012). The maximum number of outer iterations is determined by its number of floating-point operations as calculated with $\alpha \cdot \rho_{\boldsymbol{W}}$, where $\rho_{\boldsymbol{W}} = 1 + \frac{mn+nr}{mr+m}$ and $\alpha = 0.5$ is a scaling factor. An additional dynamic stopping criterion checks whether the last improvement of $\mathbf{W}$ is smaller than $\epsilon = 0.1$ of first improvement, that is, the first cycle of the outer loop. The improvement of $\mathbf{W}$ is measured in the Frobenius norm of the difference to the previous iteration, $\left\|\mathbf{W} - \mathbf{W}^{(previous)}\right\|_F$. Aside from the acceleration by performing outer updates, A-HALS improves numerical stability by ensuring no zero-columns in $\mathbf{W}$ to avoid convergence issues of HALS (Gillis and Glineur 2008; Ho 2008). Columns of all zeros are replaced by a small positive constant ($1\mathrm{e}^{-16}$). In the context of semi-NMF, we observed this to benefit the convergence of an SVD-based pseudoinverse $\mathbf{W}^\dagger$ for the subsequent update of $\mathbf{H}$. We combined Fast HALS and A-HALS by replacing the local update in A-HALS with the equations from Fast HALS and complement them with sparsity regularization.

Non-negative singular value decomposition with low rank correction

The aforementioned alternating optimization starts with an update of the weights $\mathbf{H}$ inferred from the STE $\mathbf{V}$ and an initialization of $\mathbf{W}$. We initialized the modules $\mathbf{W}$ using a slightly modified Non-negative Singular Value Decomposition with Low Rank Correction (NNSVD-LRC) which was introduced for NMF (Atif et al. 2019). Here, we describe a modified version for semi-NMF with slight optimizations. Given r modules, we extracted half that number of components from a truncated

SVD of $\mathbf{V}$. Specifically, that are $p = \left\lceil \frac{r}{2} \right\rceil$ components with $\mathbf{V} = \mathbf{U\Sigma A}$. To obtain a two-factor form as in NMF, the singular values were integrated into the left and right singular vectors to $\mathbf{U} \leftarrow \mathbf{U}\sqrt{\mathbf{\Sigma}}$ and $\mathbf{A} \leftarrow \sqrt{\mathbf{\Sigma}}\mathbf{A}$, respectively. Unlike the original NNSVD-LRC, we only initialized $\mathbf{W}$. The components of $\mathbf{U}$ sorted by decreasing singular values serve as the columns of $\mathbf{W}$. Each component is used twice, with and without a sign-flip. In the end, all negative entries are set to zero to make the initialization of $\mathbf{W}$ non-negative. In case of an odd r, the component with the lowest singular value is only added once. This resulted in

$$\mathbf{W} = \left[u_1, -u_1, u_2, -u_2, \dots, u_p, -u_p\right]_+.$$

The resulting $\mathbf{W}$ served as initializations for a few NMF iterations on the low-rank approximation $\mathbf{V}^{(p)} = \mathbf{UA}$. This further improves $\mathbf{W}$ at comparably low computational cost, because $\mathbf{V}^{(p)}$ consists of the truncated SVD for which NMF iterations reduce from a complexity of $O(nmr)$ to $O((n+m)r^2)$ (Atif et al. 2019; Gillis and Glineur 2012). In the original NNSVD-LRC, the low-rank NMF was performed using A-HALS until the difference in relative reconstruction error $\left\|\mathbf{V}^{(p)} - \mathbf{WH}\right\|_F / \left\|\mathbf{V}^{(p)}\right\|_F$ to the previous iteration falls below 5% of the initial error, which the authors amount to less than ten iterations (Atif et al. 2019). Here, since we used sparse semi-NMF instead, we found that the termination criterion proposed amounts to only one AF-HALS update in the case of STNMF. Omitting the termination criterion removes the necessity of computing the reconstruction errors. Without the need for an initial error, it suffices to create $\mathbf{W}$ from the SVD components. As STNMF is initialized on $\mathbf{W}$ only, computing $\mathbf{H}$ was subsequently inferred from the initialization of $\mathbf{W}$. This further reduced computational cost and provided a very fast initialization procedure. Inspired by A-HALS, we implemented an additional step to avoid numerical instability (Gillis and Glineur 2012). Any columns of $\mathbf{W}$ that contained only zeros were set to a small positive value ($1e^{-16}$). Any zeros residing in nonzero columns remained to accelerate matrix multiplications and to allow sparsity to emerge.

For comparison and for consensus methods, we defined a random initialization as filling $\mathbf{W}$ with values independently sampled from a uniform distribution between zero and unity. The values were drawn with a specified starting seed for reproducibility. In particular, we used the Mersenne Twister (MT19937) pseudorandom number generator (Matsumoto and Nishimura 1998) for potential compatibility across Python, R, and MATLAB with a seed between 0 and 50 for the 50 repetitions of the consensus analysis.

Consensus analysis

We used consensus analysis (Brunet et al. 2004; Monti et al. 2003) to determine the best tradeoff between solution stability and sparsity. Consensus methods allow judging the robustness of the parameter when there is a lack of measurable error. A typical proxy for goodness of decomposition in the field of NMF is the reconstruction error, or residual, as it is non-increasing over the NMF iterations (Cichocki et al. 2007). However, with additional constraints like sparsity regularization, the residual does not correspond to the used objective function. Furthermore, in the context of STNMF, the noisy excess modules may influence the residual, but neglecting them instead, favors solutions that exhibit more localized modules over those with fewer. Consequently, decompositions

based on different parameters cannot be compared easily by their residual. This is one of the reasons, why cross-validation is not suitable for STNMF. For applying consensus methods, we performed 30 randomly initialized repetitions of sparse semi-NMF (50 repetitions for the detailed comparative analysis) for a range of parameters. The parameters correspond to different sparsity regularization strength. Semi-NMF was run for 1000 iterations as we found that to be typically enough to reach convergence. Similarity between solutions was assessed by treating the encodings in **H** as cluster labels and comparing them across solutions. We define cluster labels as the index of the module with the highest absolute weight for a given spike. Although a spike may be elicited by a combination of subunit activations, considering the maximum associated module suffices for sake of comparison. Because the ordering of the emerging modules is arbitrary, the cluster labels cannot be compared across solutions directly. To remove this ambiguity, pairwise comparison of cluster labels among the spikes within each solution are computed in a connectivity matrix **G**. Each entry g_{ij} in the square matrix is one, if spikes i and j share the cluster label and zero otherwise. As the number of unique pairs grows according to $m(m-1)/2$ for increasing number of spikes m, we limit ourselves to 25000 spikes selected at random with a persistent selection across the analysis. That amounts to more than 1 gigabyte of memory at 32-bit floating-point precision for one decomposition. The number may be increased with more memory available or with memory efficient implementations. Nevertheless, we found that this subsample provides a sufficient representation of the resulting cluster memberships. The elementwise average of the connectivity matrices of all 30 solutions expresses the consensus across solutions. The consensus matrix contains values between zero and one, describing the fraction of agreement between the encodings across the 30 runs. We calculated the cophenetic correlation to describe the dispersion of the consensus matrix. It is the Pearson correlation between the consensus matrix and its linkage matrix provided by hierarchical clustering with an average distance metric. The coefficient is a real value between zero and one, with one denoting perfect consensus and identical solutions (Brunet et al. 2004).

STNMF reserves excess modules to capture the noise in the data that are not considered localized subunits and are not representative of the solution. As uncorrelated noise is distributed arbitrarily across the excess modules, they are likely to vary across solutions. Thereby they skew the described metric. Consequently, we only considered modules containing localized subunits as valid cluster labels. Connectivity pairs involving non-localized modules were set to zero in the connectivity matrices. These entries do not reach consensus across solutions. This causes a lower dispersion in the consensus matrix, quenching the CPCCs. The metric becomes more sensitive to differences in detected subunits. In the case of no localized subunits in the decomposition, we did not compute the cophenetic correlation ("not-a-number"). This results in holes in the stability curve.

The CPCC corresponds to the similarity of the recovered subunits across the repeated decompositions. It represents the stability of the solutions and we used it to compare the different sparsity parameters. The optimal trade-off between high stability and little regularization was found at the bend of the curve of sparsity versus stability (**Figure 4 C**). Here, we determined the inflection point by visual inspection. To identify it systematically instead, we propose that one could describe the curve as a piecewise linear function with two or three components and select the intersection of the functions. In case of fluctuations in the curve beyond the bend, the linear fit may fail and a fit to a smoothed copy or an exponential approximation of the curve can be done instead. We defined the selected point on the curve the most suitable sparsity parameter.

Subunit comparison

To compare recovered subunits on a geometrical basis, we computed their relative spatial overlap. Given the contour outlines (described above) of two subunits i and j at two standard deviations, we calculated their relative overlap o_{ij} with the Jaccard index (Jaccard 1912). The intersection is divided by the union of their areas a_i and a_j

$$o_{ij} = \frac{a_i \cap a_j}{a_i \cup a_j}$$

We considered two outlines with $o_{ij} > 0.5$, that is, more than 50% area overlap, to be fully overlapping and candidates of identical subunits.

Inter-mosaic coordination

The spatial coordination of ON- and OFF-type subunit mosaics was determined using the IMCE as proposed previously (Roy et al. 2021). In brief, contour mosaics were reduced to subunit centroids by their center of mass. Within a region of interest (ROI) one of the mosaics (center of mass points) was shifted in horizontal and vertical position on top of the other. The squared inverse distance between each heterotopic pair of centroids r_{ij} was measured with

$$E_{ij} = \begin{cases} \dfrac{1}{(r_{ij} - r_{min})^2}, & r_{ij} > r_{min} \\ \dfrac{1}{r_{min}^2}, & r_{ij} \leq r_{min} \end{cases}$$

where r_{min} is a minimum distance set to 0.2 times the median heterotopic nearest-neighbor distance. The values of E_{ij} were summed and dived by the total number of pairs. The summed energy was computed for different offsets of one mosaic on top of the other before being z-scored and normalized to the median homotypic nearest neighbor distance. The generated topographic map of IMCE values was radially averaged to produce the increasing or decreasing curve. We selected the ROI in two forms. First, we manually chose a region where the density of subunits was approximately homogeneous on both mosaics to minimize effects of gaps. We then confirmed the determined IMCE by following the proposed implementation of finding a suitable convex hull that maximized the number of enclosed heterotopic pairs while excluding outliers (Roy et al. 2021).

REFERENCES

Anishchenko, A., Greschner, M., Elstrott, J., Sher, A., Litke, A. M., Feller, M. B., and Chichilnisky, E. J. (2010). Receptive field mosaics of retinal ganglion cells are established without visual experience. *Journal of Neurophysiology*, *103*(4), 1856–1864.

Atif, S. M., Qazi, S., and Gillis, N. (2019). Improved SVD-based initialization for nonnegative matrix factorization using low-rank correction. *Pattern Recognition Letters*, *122*, 53–59.

Berry, M. J., Warland, D. K., and Meister, M. (1997). The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 5411–5416.

Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P., and Plemmons, R. J. (2007). Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics and Data Analysis*, *52*(1), 155–173.

Boutsidis, C., and Gallopoulos, E. (2008). SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*, *41*(4), 1350–1362.

Boycott, B. B., and Wässle, H. (1991). Morphological Classification of Bipolar Cells of the Primate Retina. *European Journal of Neuroscience*, *3*(11), 1069–1088.

Bro, R., Kjeldahl, K., Smilde, A. K., and Kiers, H. A. L. (2008). Cross-validation of component models: A critical look at current methods. *Analytical and Bioanalytical Chemistry*, *390*(5), 1241–1251.

Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(12), 4164–4169.

Calkins, D. J., and Sterling, P. (2007). Microcircuitry for Two Types of Achromatic Ganglion Cell in Primate Fovea. *Journal of Neuroscience*, *27*(10), 2646–2653.

Candès, E. J., Romberg, J., and Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, *52*(2), 489–509.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, *12*(2), 199–213.

Choromanska, A., Henaff, M., Mathieu, M., Arous, G. Ben, and LeCun, Y. (2014). The Loss Surfaces of Multilayer Networks. *ArXiv*, 1412.0233v3.

Cichocki, A., and Phan, A. H. (2009). Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, *E92-A*(3), 708–721.

Cichocki, A., Zdunek, R., and Amari, S. (2007). Hierarchical ALS Algorithms for Nonnegative Matrix and 3D Tensor Factorization. In *Independent Component Analysis and Signal Separation* (pp. 169–176). Springer Berlin Heidelberg.

Cichocki, A., Zdunek, R., Phan, A. H., and Amari, S. (2009). *Nonnegative Matrix and Tensor Factorizations* (1st ed.). Wiley.

Cohen, E., and Sterling, P. (1990a). Convergence and divergence of cones onto bipolar cells in the central area of cat retina. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *330*(1258), 323–328.

Cohen, E., and Sterling, P. (1990b). Demonstration of cell types among cone bipolar neurons of cat retina. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *330*(1258), 305–321.

Crook, J. D., Packer, O. S., and Dacey, D. M. (2014). A synaptic signature for ON- and OFF-center parasol ganglion cells of the primate retina. *Visual Neuroscience*, *31*(1), 57–84.

Crook, J. D., Peterson, B. B., Packer, O. S., Robinson, F. R., Troy, J. B., and Dacey, D. M. (2008). Y-Cell Receptive Field and Collicular Projection of Parasol Ganglion Cells in Macaque Monkey Retina. *Journal of Neuroscience*, *28*(44), 11277–11291.

d'Aspremont, A., El Ghaoui, L., Jordan, M. I., and Lanckriet, G. R. G. (2007). A Direct Formulation for Sparse PCA Using Semidefinite Programming. *SIAM Review*, *49*(3), 434–448.

Dacey, D. M. (1993). The mosaic of midget ganglion cells in the human retina. *The Journal of Neuroscience*, *13*(12), 5334–5355.

Dacey, D. M. (2000). Parallel Pathways for Spectral Coding in Primate Retina. *Annual Review of Neuroscience*, *23*(1), 743–775.

Dacey, D. M., Packer, O. S., Diller, L., Brainard, D., Peterson, B., and Lee, B. (2000). Center surround receptive field structure of cone bipolar cells in primate retina. *Vision Research*, *40*(14), 1801–1811.

DeVries, S. H., and Baylor, D. A. (1997). Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *Journal of Neurophysiology*, *78*(4), 2048–2060.

Ding, C., Tao Li, and Jordan, M. I. (2010). Convex and Semi-Nonnegative Matrix Factorizations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(1), 45–55.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, *41*(3), 613–627.

Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, *52*(4), 1289–1306.

Drinnenberg, A., Franke, F., Morikawa, R. K., Jüttner, J., Hillier, D., Hantz, P., Hierlemann, A., Azeredo da Silveira, R., and Roska, B. (2018). How Diverse Retinal Functions Arise from Feedback at the First Visual Synapse. *Neuron*, *99*(1), 117-134.e11.

Elad, M. (2006). Why Simple Shrinkage Is Still Relevant for Redundant Representations? *IEEE Transactions on Information Theory*, *52*(12), 5559–5569.

Enroth-Cugell, C., and Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology*, *187*(3), 517–552.

Eriköz, B., Jusuf, P. R., Percival, K. A., and Grünert, U. (2008). Distribution of bipolar input to midget and parasol ganglion cells in marmoset retina. *Visual Neuroscience*, *25*(1), 67–76.

Fu, W., and Perry, P. O. (2017). Estimating the number of clusters using cross-validation. *Journal of Computational and Graphical Statistics*, *29*(1), 162–173.

Gillis, N., and Glineur, F. (2008). Nonnegative Factorization and The Maximum Edge Biclique Problem. *ArXiv*, 0810.4225v1.

Gillis, N., and Glineur, F. (2012). Accelerated Multiplicative Updates and Hierarchical ALS Algorithms for Nonnegative Matrix Factorization. *Neural Computation*, *24*(4), 1085–1105.

Gollisch, T., and Liu, J. K. (2018). *Data: Salamander retinal ganglion cells under finely structured spatio-temporal white noise* [Dataset]. G-Node. https://doi.org/10.12751/g-node.62b65b

Hansen, P. C. (1994). Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, *6*, 1–35.

Hansen, P. C. (1999a). Regularization Tools Version 3.0 for Matlab 5.2. *Numerical Algorithms*, *20*, 195–196.

Hansen, P. C. (1999b). *The L-curve and its use in thenumerical treatment of inverse problems: Vol. 1999:15*. IMM, Department of Mathematical Modelling, Technical University of Denmark.

Ho, N.-D. (2008). *Nonnegative matrix factorization algorithms and applications* [Diss. PhD thesis]. Université catholique de Louvain.

Hoyer, P. O. (2002). Non-negative sparse coding. *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, 557–565.

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, *5*, 1457–1469.

Hyvärinen, A. (1999). Sparse Code Shrinkage: Denoising of Nongaussian Data by Maximum Likelihood Estimation. *Neural Computation*, *11*(7), 1739–1768.

Jaccard, P. (1912). The Distribution of the Flora in the Alpine Zone. *New Phytologist*, *11*(2), 37–50.

Jacoby, R. A., Wiechmann, A. F., Amara, S. G., Leighton, B. H., and Marshak, D. W. (2000). Diffuse bipolar cells provide input to OFF parasol ganglion cells in the macaque retina. *The Journal of Comparative Neurology*, *416*(1), 6–18.

Jacoby, R., Stafford, D., Kouyama, N., and Marshak, D. (1996). Synaptic inputs to ON parasol ganglion cells in the primate retina. *Journal of Neuroscience*, *16*(24), 8041–8056.

Jarsky, T., Cembrowski, M., Logan, S. M., Kath, W. L., Riecke, H., Demb, J. B., and Singer, J. H. (2011). A synaptic mechanism for retinal adaptation to luminance and contrast. *Journal of Neuroscience*, *31*(30), 11003–11015.

Jia, S., Yu, Z., Onken, A., Tian, Y., Huang, T., and Liu, J. K. (2022). Neural System Identification With Spike-Triggered Non-Negative Matrix Factorization. *IEEE Transactions on Cybernetics*, *52*(6), 4772–4783.

Jusuf, P. R., Martin, P. R., and Grünert, U. (2006a). Synaptic connectivity in the midget-parvocellular pathway of primate central retina. *The Journal of Comparative Neurology*, *494*(2), 260–274.

Jusuf, P. R., Martin, P. R., and Grünert, U. (2006b). Random wiring in the midget pathway of primate retina. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(15), 3908–3917.

Kim, H., and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics*, *23*(12), 1495–1502.

Kim, J., and Park, H. (2008). Sparse Nonnegative Matrix Factorization for Clustering. *Georgia Institute of Technology*, *Tech. Rep. GT-CSE-08-01*.

Kim, J., and Park, H. (2011). Fast Nonnegative Matrix Factorization: An Active-Set-Like Method and Comparisons. *SIAM Journal on Scientific Computing*, *33*(6), 3261–3281.

Kolb, H. (1979). The inner plexiform layer in the retina of the cat: electron microscopic observations. *Journal of Neurocytology*, *8*(3), 295–329.

Kolb, H., and Dekorver, L. (1991). Midget ganglion cells of the parafovea of the human retina: A Study by electron microscopy and serial section reconstructions. *The Journal of Comparative Neurology*, *303*(4), 617–636.

Lawson, C. L., and Hanson, R. J. (1974). *Solving Least Squares Problems*. Prentice-Hall.

Lee, B. B., Martin, P. R., and Grünert, U. (2010). Retinal connectivity and primate vision. *Progress in Retinal and Eye Research*, *29*(6), 622–639.

Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791.

Li, H., Calder, C. A., and Cressie, N. (2007). Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. *Geographical Analysis*, *39*(4), 357–375.

Li, L., and Zhang, Y.-J. (2009). FastNMF: highly efficient monotonic fixed-point nonnegative matrix factorization algorithm with good applicability. *Journal of Electronic Imaging*, *18*(3), 033004.

Li, S. Z., Hou, X. W., Zhang, H. J., and Cheng, Q. S. (2001). Learning spatially localized, parts-based representation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *1*, I-207-I–212.

Liu, J. K., Schreyer, H. M., Onken, A., Rozenblit, F., Khani, M. H., Krishnamoorthy, V., Panzeri, S., and Gollisch, T. (2017). Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nature Communications*, *8*(1), 149.

Maheswaranathan, N., Kastner, D. B., Baccus, S. A., and Ganguli, S. (2018). Inferring hidden structure in multilayered neural circuits. *PLOS Computational Biology*, *14*(8), e1006291.

Maheswaranathan, N., McIntosh, L. T., Tanaka, H., Grant, S., Kastner, D. B., Melander, J. B., Nayebi, A., Brezovec, L., Wang, J., Ganguli, S., and Baccus, S. A. (2019). The dynamic neural code of the retina for natural scenes. *BioRxiv*, 340943v5.

Manookin, M. B., and Demb, J. B. (2006). Presynaptic Mechanism for Slow Contrast Adaptation in Mammalian Retinal Ganglion Cells. *Neuron*, *50*(3), 453–464.

Marc, R. E., and Sperling, H. G. (1977). Chromatic organization of primate cones. *Science*, *4288*(196), 454–456.

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

Matsumoto, M., and Nishimura, T. (1998). Mersenne twister. *ACM Transactions on Modeling and Computer Simulation*, *8*(1), 3–30.

McFarland, J. M., Cui, Y., and Butts, D. A. (2013). Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs. *PLoS Computational Biology*, *9*(7), e1003143.

McIntosh, L. T., Maheswaranathan, N., Nayebi, A., Ganguli, S., and Baccus, S. A. (2016). Deep Learning Models of the Retinal Response to Natural Scenes. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 29). Curran Associates, Inc.

Monti, S., Tamayo, P., Mesirov, J. P., and Golub, T. R. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, *52*(1), 91–118.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, *37*(1–2), 17–23.

Owen, A. B., and Perry, P. O. (2009). Bi-cross-validation of the SVD and the nonnegative matrix factorization. *Annals of Applied Statistics*, *3*(2), 564–594.

Pachitariu, M., Steinmetz, N. A., Kadir, S., Carandini, M., and Harris, K. D. (2016). Fast and accurate spike sorting of high-channel count probes with KiloSort. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 29). Curran Associates, Inc.

Qiao, H. (2015). New SVD based initialization strategy for non-negative matrix factorization. *Pattern Recognition Letters*, *63*, 71–77.

Real, E., Asari, H., Gollisch, T., and Meister, M. (2017). Neural Circuit Inference from Function to Structure. *Current Biology*, *27*(2), 189–198.

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

Roy, S., Jun, N. Y., Davis, E. L., Pearson, J., and Field, G. D. (2021). Inter-mosaic coordination of retinal receptive fields. *Nature*, *592*(7854), 409–413.

Saxena, S., Kinsella, I., Musall, S., Kim, S. H., Meszaros, J., Thibodeaux, D. N., Kim, C., Cunningham, J., Hillman, E., Churchland, A. K., and Paninski, L. (2019). Localized semi-nonnegative matrix factorization (LocaNMF) of widefield calcium imaging data. *BioRxiv*, 650093.

Schwartz, G. W., Okawa, H., Dunn, F. A., Morgan, J. L., Kerschensteiner, D., Wong, R. O., and Rieke, F. (2012). The spatial structure of a nonlinear receptive field. *Nature Neuroscience*, *15*(11), 1572–1580.

Shah, N. P., Brackbill, N., Rhoades, C. E., Kling, A., Goetz, G., Litke, A. M., Sher, A., Simoncelli, E. P., and Chichilnisky, E. J. (2020). Inference of nonlinear receptive field subunits with spike-triggered clustering. *ELife*, *9*, e45743.

Sterling, P., Freed, M. A., and Smith, R. G. (1988). Architecture of rod and cone circuits to the on-beta ganglion cell. *Journal of Neuroscience*, *8*(2), 623–642.

Szmajda, B. A., Grünert, U., and Martin, P. R. (2005). Mosaic properties of midget and parasol ganglion cells in the marmoset retina. *Visual Neuroscience*, *22*(4), 395–404.

Tanaka, H., Nayebi, A., Maheswaranathan, N., McIntosh, L. T., Baccus, S. A., and Ganguli, S. (2019). From deep learning to mechanistic understanding in neuroscience: the structure of retinal prediction. *ArXiv*, 1912.06207v1.

Vavasis, S. A. (2010). On the Complexity of Nonnegative Matrix Factorization. *SIAM Journal on Optimization*, *20*(3), 1364–1377.

Victor, J. D., and Shapley, R. M. (1979). The nonlinear pathway of Y ganglion cells in the cat retina. *Journal of General Physiology*, *74*(6), 671–689.

Wässle, H. (1999). Parallel pathways from the outer to the inner retina in primates. In K. R. Gegenfurtner, & L. T. Sharpe (Eds.), *Color Vision: From Genes to Perception* (pp. 145–162). Cambridge University Press.

Wässle, H. (2004). Parallel processing in the mammalian retina. *Nature Reviews Neuroscience*, *5*(10), 747–757.

Watanabe, M., and Rodieck, R. W. (1989). Parasol and midget ganglion cells of the primate retina. *The Journal of Comparative Neurology*, *289*(3), 434–454.

Wild, S. (2003). *Seeding non-negative matrix factorizations with the spherical K-Means clustering* [Master's thesis]. University of Colorado.

Zhou, T., Kang, J., Cong, F., and Li, D. X. (2020). Early childhood developmental functional connectivity of autistic brains with non-negative matrix factorization. *NeuroImage: Clinical*, *26*, 102251.

SPARSE MATRIX DECOMPOSITION WITH ORTHOGONAL REGULARIZATION YIELDS PARTS-BASED REPRESENTATIONS WITH CO-EXISTING POSITIVE AND NEGATIVE VALUES

4

Attached manuscript. Research article, currently in preparation. Reprinted below is the manuscript at its current version.

Authors:

Sören J. Zapp[1,2], Tim Gollisch[1,2,3]

1. University Medical Center Göttingen, Department of Ophthalmology, Göttingen, Germany
2. Bernstein Center for Computational Neuroscience, Göttingen, Germany
3. Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, Göttingen, Germany

Author contributions:

I developed and implemented the factorization algorithm, analyzed the data, and prepared figures and the initial draft of the manuscript. TG supervised the project and contributed to writing the manuscript. All authors helped revise the manuscript. The facial image dataset stems from a publically available repository.

CBCL Face Database No.1. (2000). *MIT Center For Biological and Computation Learning*. http://cbcl.mit.edu/software-datasets/FaceData2.html

Sparse matrix decomposition with orthogonal regularization yields parts-based representations with co-existing positive and negative values

Sören J. Zapp[1,2], Tim Gollisch[1,2,3]

1. University Medical Center Göttingen, Department of Ophthalmology, Göttingen, Germany
2. Bernstein Center for Computational Neuroscience, Göttingen, Germany
3. Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, Göttingen, Germany

ABSTRACT

Parts-based decompositions of high dimensional data are of interest for many industrial and scientific applications, like computer vision, neuroscience, and text data mining. Non-negative matrix factorization is of particular interest because of its inherent additive nature leading to interpretable elements. However, the application in dimensionality reduction is limited, because factorized components are bound to be non-negative and cannot describe co-existing positive and negative values. Here, we explore different forms of regularization, constraints, and subsequent projections to address that challenge in an efficient way. We develop a method of sparse matrix decomposition with orthogonal regularization without parameter and focus on a simple implementation with the attempt to provide an intuitive and mathematically trivial algorithm. The method is compared to similar concepts from the fields of non-negative matrix factorization, convolutional neural networks, and autoencoders, and is put into perspective with potential applications. We demonstrate the technique with the successful parts-based decomposition of a database of facial images.

MAIN TEXT

Non-negative matrix factorization (NMF) is known for its ability to provide parts-based decompositions (Lee and Seung 1999). This is facilitated through its non-negativity constraint that enforces additive building blocks without negative cancelations. A self-evident disadvantage is the exclusivity to non-negative data. Various efforts have been made to accomplish a factorization into parts by replacing non-negativity with sparseness (Jolliffe et al. 2003; Mairal et al. 2009; Witten et al. 2009). These involve intricate computations bound by constraints. Here, we derive an iterative matrix decomposition algorithm based on unconstraint least squares that produces parts-based representations using orthogonal regularization and sparsity endorsement.

We use the well-known face decomposition example (Lee and Seung 1999) to demonstrate parts-based decompositions under positive and negative sign (**Figure 1**). Reducing this dataset of facial images into a lower dimensional representation, principal component analysis (PCA) constructs "eigenfaces" (Turk and Pentland 1991). These holistic components densely span the entire image (**Figure 1B**) and are not informative about the underlying visual structure in the faces (see Lee and Seung 1999). With sparseness and a less confining orthogonal regularization instead, the faces are factorized into localized parts, across negative and positive sign (**Figure 1A**). The overall much sparser components include a few denser factor images. These are distinct form holistic representations, because, instead of a blurring effect, they appear to separate a few fundamental face structures by characteristics like big glasses and neutral or smiling expressions. More prominent are the individual parts of the face, like a nose, regions around the eyes or a mouth. These resemble the parts-based deconstruction of the non-negative faces with NMF (Lee and Seung 1999). A third group are facial characteristics of contrast (**Figure 1B** green outlines). This demonstrates the ability of the algorithm to agglomerate positive and negative pixel values and recognize reoccurring opposing facial features. To explain the implementation in detail and to emphasize the relationship to recently reported efforts, we derive it from NMF.

The most prominent example of parts-based decomposition is set by NMF (Lee and Seung 1999; Paatero and Tapper 1994). The nonnegative matrix V is factorized into features, a tall and skinny matrix W, and encodings, a short and fat matrix H, both nonnegative.

$$\min_{W\geq 0, H\geq 0} \|V - WH\|_F^2 \tag{1}$$

As this is a non-convex problem to solve (Vavasis 2010), NMF is usually formulated as an iterative procedure of alternating optimization of W and H. The computational hurdle of solving the non-negative least squares at the core of these sub-problems can be cleared by implementing an intuitive algorithm called alternating least squares (ALS) (Paatero and Tapper 1994). W and H are updated in alternating fashion while the respective other is held fix by solving a least squares problem. To avoid complex calculations involving a non-negativity constraint, these least squares are unconstrained in the methods of ALS and are followed by setting all negative elements to zero with

$$H \leftarrow \min_{H\geq 0} \|V - WH\|_F^2 \approx \left[W^{\dagger}V\right]_+ = \left[W^{\top}(W^{\top}W)^{-1}V\right]_+ \tag{2}$$

$$W \leftarrow \min_{W\geq 0} \|V - WH\|_F^2 \approx \left[VH^{\dagger}\right]_+ = \left[VH^{\top}(HH^{\top})^{-1}\right]_+ \tag{3}$$

where $A^{\dagger}$ is the Moore-Penrose pseudoinverse of A and $[A]_+$ is the elementwise maximum with zero, $\max\{A, 0\}$. Note, that the rows of H usually are $\ell 2$-normalized after Equation 2, to combat exploding coefficients (Ho 2008). In the context of semi-NMF, that loosens the non-negativity constraint on one of the factor matrices, say W, (Ding et al. 2010), the elementwise maximum in Equation 3 is often replaced with a sparsity constraint (Hoyer 2002, 2004; H. Kim and Park 2007; J. Kim and Park 2008). An alternative common constraint in NMF is orthogonality to retain parts-based solutions. Alongside the non-negativity, it is often used to enforce unique labeling of features, resulting in hard k-means clustering (Ding et al. 2006). In conjunction with semi-NMF, however, orthogonal encodings and non-negative features produce a soft clustering. Semi-orthogonal NMF removes redundancies but preserves uncorrelated features, while allowing samples to belong to multiple features (Li et al. 2018). Instead of the orthogonality constraint, that involves expensive calculations, we implement orthogonal regularization by the following derivation.

Applied to the ALS update of W, the classical regularization method in its most general form (Phillips 1962; Tikhonov 1963) is formulated with

$$W \leftarrow \min_{W}\{\|V - WH\|_F^2 + \lambda^2\|L(W - W_0)\|_F^2\}, \tag{4}$$

where λ is the regularization weight to adjust the impact, W_0 is an initial estimate (typically zero), and L is the regularization matrix that defines the prior of the regularization. For instance, in the context of sparsity regularization, L is a matrix of all ones. The problem in numerically solvable form (Hansen 1999) rephrases to

$$W \leftarrow \min_{W}\left\|\begin{pmatrix} V \\ \lambda L W_0 \end{pmatrix} - \begin{pmatrix} W \\ \lambda L \end{pmatrix} H\right\|_F. \tag{5}$$

Assuming $W_0 = 0$, it reformulates into a "regularized normal equation" (Tikhonov 1963), and can be related back to the Moore-Penrose inverse from Equation 3

$$W \leftarrow VH^{\top}(HH^{\top} + \lambda L)^{-1}, \tag{6}$$

now with the additional term λL. With a regularization prior $L = I - HH^{\top}$, we impose regularization in which orthogonality is encouraged by penalizing the derivation of $HH^{\top}$ from the identitiy matrix I. This replaces the strict orthogonality constraint (Li et al. 2018) with a soft orthogonality. In part, this is related to a similar concept in convolutional neural networks, although applied differently (Brock et al. 2016; Wang et al. 2020). If we now set $\lambda = 1$, we can omit the expensive matrix calculations and the matrix inverse to drastically reduce computational time with

$$W \leftarrow VH^{\top}(HH^{\top} + I - HH^{\top})^{-1} = VH^{\top}I^{-1} = VH^{\top}. \tag{7}$$

This reduction in the update has also been observed when the update of H is explicitly confined to produce an orthogonal $HH^{\top}$ (Li et al. 2018) or when it is subsequently orthogonalized, for example with $H \leftarrow \sqrt{HH^{\top}}H$. Here, however, to state it in other words, there is neither a computational constraint nor explicit regularization. $HH^{\top}$ is merely assumed to be orthogonal (compare Equations 3 and 7). The fast succession of unconstrained updates create a self-fulfilling prophecy as W and H are based on each other in ALS. In the context of semi-NMF, interestingly, this can provide a very fast algorithm, that, judging by the reduced operations, even surpasses ALS in speed.

$$H \leftarrow \left[W^{\dagger}V\right]_{+} \qquad (8)$$
$$W \leftarrow VH^{\top}$$

With suitable dimensions of the input matrix to fall in an appropriate regime of regularization, the implicit regularized orthogonality achieves a parts-based decomposition without the need to tune a parameter (results not shown).

It stands to reason to conclude that an ALS implementation may require two constraints to produce a decomposition into parts: non-negativity on both factors (Lee and Seung 1999), one non-negativity and one sparsity constraint (Ding et al. 2010), or one non-negativity and one orthogonality constraint (Li et al. 2018). Consequently, we select two constraints to eliminate the non-negativity. To satisfy this requirement, we replace the non-negativity in Equation 8 by endorsing sparse factorizations. To avoid increasing the complexity of the updates again, we implement soft thresholding (Donoho 1995) as a regularization-like projection applied after the update of W in Equation 7

$$W \leftarrow \mathrm{sgn}\{W\}\,[|W| - \lambda]_{+}, \qquad (9)$$

where sgn A is the elementwise sign function of A and λ is a parameter analogous to a regularization weight. This projection mimics the induced sparsity evoked by a non-negativity rectification. Instead of cutting off all negative values, a portion of the smallest magnitude values is discarded instead. A suitable choice of λ will produce factorizations into appropriate components. Together, this forms the complete algorithm

Algorithm 1 Combination of Equations 2, 7, and 9.

Inputs: Data $V \in \mathbb{R}^{n\times m}$, parameter λ
Initialize $W \in \mathbb{R}^{n\times r} \sim U(-2\lambda, 2\lambda)$; the range may be freely chosen
Repeat
  Update H with $H \leftarrow W^{\dagger}V$
  $\ell 2$-normalize the rows of H
  Update W with $W \leftarrow VH^{\top}$
  Sparsify W with $W \leftarrow \mathrm{sgn}\{W\}\,[|W| - \lambda]_{+}$
Until convergence is reached

where W is suggested to be initialized from a uniform distribution. Initialization using techniques based on singular value decomposition may provide more stable solutions and earlier convergence (Boutsidis and Gallopoulos 2008). The algorithm is repeated until a convergence criterion is reached.

An alternative implementation (not shown here) is to use quantile-based thresholding for inducing sparsity. There, λ is the fraction of the maximum magnitude of W that determines the amount of values dropped. One advantage is that it is adjustable to affect each component individually. A disadvantage, however, is that it scales with the magnitude of the values over the iterations and always drops a fixed range of values.

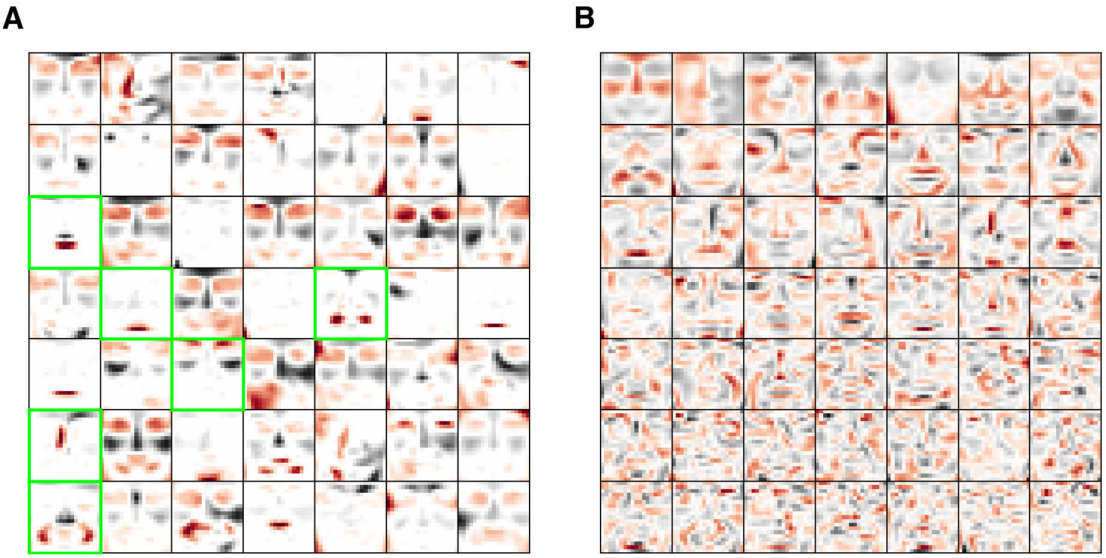


Figure 1 Face decomposition with the proposed matrix factorization **(A)** and PCA **(B)**. While PCA recovers holistic and dense images (eigenfaces), the matrix decomposition offers parts-based representations of different pixel sparsity. Among a few "basis" faces that capture distinct categories (like faces with glasses or smiling expressions), images recover individual parts, like a nose, eyes or a mouth. But also co-occurring features of opposing sign are inferred. Selected examples are outlined in green.

DISCUSSION

The formulation of semi-NMF shows close resemblances of an autoencoder (Kramer 1991). The input layer and prediction in the last layer is V, the latent variables are W and the connection weights are $H^\top$ and H, respectively. This encompasses a rectified linear unit at the elementwise maximum with zero.

$$V \times H^\top \rightarrow \quad ReLu \quad \rightarrow W$$
$$W \times H \rightarrow Identity \rightarrow \hat{V}$$

This relationship could potentially be harnessed to explore updating rules and other forms of constraints and regularization to improve both matrix factorization and autoencoders.

The proposed algorithm of iterative sparse matrix decomposition is a promising approach for neuroscience, because it allows negative components, but is still in line with sparse coding in the visual cortex (Olshausen and Field 1997). A concrete example is the substructure in receptive fields in the visual system. For instance, certain types of retinal ganglion cells perform nonlinear processing of their inputs (Gollisch and Meister 2010). These inputs are integrated into the ganglion cells' spatial receptive fields, the neuron's sensitive region of the visual field. These nonlinear contributions were successfully inferred as subunits in the receptive field with a variation of NMF (Liu et al. 2017). However, the inputs of the presynaptic bipolar cells that represent the subunits exhibit a center-surround receptive field. Non-negative subunits are not able to fully describe them. Along with suitable visual stimulation, our algorithm may address this challenge.

ACKNOWLEDGMENTS

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 724822) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 432680300 (SFB 1456, project B05).

AUTHOR CONTRIBUTIONS

SJZ developed and implemented the factorization algorithm, analyzed the data, and prepared figures and the initial draft of the manuscript. TG supervised the project and contributed to writing the manuscript. All authors helped revise the manuscript.

DISCLOSURE STATEMENT

The authors declare no competing interests.

METHODS

The data consist of 2429 grayscale images of human faces in frontal view at a size of 19 × 19 pixels from a previously available dataset (CBCL Face Database No.1 2000). The pixel values of each image were scaled to a mean and standard deviation of 0.25 before being clipped to a range of [0, 1] as done in previous analyses (Lee and Seung 1999). To obtain data of positive and negative values, the data was subsequently z-scored. The images were compiled into a matrix of the dimensions $n \times m$, with n pixels and m images. The matrix was factorized starting from $W \in \mathbb{R}^{n \times r}$, with $r = 49$, initialized randomly from a uniform distribution in the range of $[-2\lambda, 2\lambda]$ to yield around 50% sparse components after applying the sparsity constraint. Sparsity was applied with $\lambda = 4$. Alternating updates of H and W were repeated for 2000 iterations, which took a few seconds on a 4 GHz 8 core processor. PCA was obtained from the eigenvectors with the highest eigenvalues of the matrix $VV^{\top}$.

REFERENCES

Boutsidis, C., and Gallopoulos, E. (2008). SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*, *41*(4), 1350–1362.

Brock, A., Lim, T., Ritchie, J. M., and Weston, N. (2016). Neural Photo Editing with Introspective Adversarial Networks. *ArXiv*, 1609.07093v3.

CBCL Face Database No.1. (2000). *MIT Center For Biological and Computation Learning* [Dataset]. http://cbcl.mit.edu/software-datasets/FaceData2.html

Ding, C., Li, T., Peng, W., and Park, H. (2006). Orthogonal nonnegative matrix tri-factorizations for clustering. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *2006*, 126–135.

Ding, C., Tao Li, and Jordan, M. I. (2010). Convex and Semi-Nonnegative Matrix Factorizations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(1), 45–55.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, *41*(3), 613–627.

Gollisch, T., and Meister, M. (2010). Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina. *Neuron*, *65*(2), 150–164.

Hansen, P. C. (1999). *The L-curve and its use in thenumerical treatment of inverse problems: Vol. 1999:15*. IMM, Department of Mathematical Modelling, Technical University of Denmark.

Ho, N.-D. (2008). *Nonnegative matrix factorization algorithms and applications* [Diss. PhD thesis]. Université catholique de Louvain.

Hoyer, P. O. (2002). Non-negative sparse coding. *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, 557–565.

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, *5*, 1457–1469.

Jolliffe, I. T., Trendafilov, N. T., and Uddin, M. (2003). A Modified Principal Component Technique Based on the LASSO. *Journal of Computational and Graphical Statistics*, *12*(3), 531–547.

Kim, H., and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics*, *23*(12), 1495–1502.

Kim, J., and Park, H. (2008). Sparse Nonnegative Matrix Factorization for Clustering. *Georgia Institute of Technology*, *Tech. Rep. GT-CSE-08-01*.

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, *37*(2), 233–243.

Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791.

Li, J. Y., Zhu, R., Qu, A., Ye, H., and Sun, Z. (2018). Semi-orthogonal Non-negative Matrix Factorization with an Application in Text Mining. *ArXiv*, 1805.02306v3.

Liu, J. K., Schreyer, H. M., Onken, A., Rozenblit, F., Khani, M. H., Krishnamoorthy, V., Panzeri, S., and Gollisch, T. (2017). Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nature Communications*, *8*(1), 149.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2009). Online dictionary learning for sparse coding. *Proceedings of the 26th Annual International Conference on Machine Learning*, 689--696.

Olshausen, B. A., and Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, *37*(23), 3311–3325.

Paatero, P., and Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, *5*(2), 111–126.

Phillips, D. L. (1962). A Technique for the Numerical Solution of Certain Integral Equations of the First Kind. *Journal of the ACM*, *9*(1), 84–97.

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method (1963 Dokl. Akad. Nauk. SSSR 151, 501–504, Trans.). *Soviet Math. Dokl.*, *4*, 1035–1038.

Turk, M., and Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86.

Vavasis, S. A. (2010). On the Complexity of Nonnegative Matrix Factorization. *SIAM Journal on Optimization*, *20*(3), 1364–1377.

Wang, J., Chen, Y., Chakraborty, R., and Yu, S. X. (2020). Orthogonal Convolutional Neural Networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11502–11512.

Witten, D. M., Tibshirani, R., and Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics (Oxford, England)*, *10*(3), 515–534.

5 OVERALL DISCUSSION

In this thesis, I have developed computational methods and recorded ganglion cell responses in the primate retina using multi-electrode arrays (MEAs), with the goal of investigating nonlinear encoding in the retina. In Chapter 2, I have explored nonlinear spatial integration and how it might be involved in the extraction of various visual features. A comprehensive outline of past scientific findings about the receptive-field substructure of retinal ganglion cells (RGCs) is provided. It is complemented with an outline of investigations of the neural counterparts among the presynaptic RGC circuitry. I have reviewed the strengths of existing computational methods and collected current challenges. To facilitate resolving these, I developed methods of subunit inference with particular focus on their analysis speed and applicability (Chapter 3). With these, I was able to demonstrate that 45 minutes of recorded data allow the recovery of subunits within a few seconds as opposed to several hours. I achieved this by implementing new algorithms and developing tools for hyperparameter tuning. The resulting method is compatible with routine recordings and analyses. In Chapter 4, I further explored the potential of matrix factorization, its algorithms, and its functionality under different constraints in the context of dimensionality reduction. I developed algorithms with the potential to recover the center-surround structure of subunits. In application of my methods, I found subunits in all of the four major ganglion cell types in the primate retina. Differences in subunit properties between cell types and the alignment of their subunit layouts reveal new insights about retinal circuitry. Overall, these relevant findings prove the potential of my methods to aid in understanding the nonlinear computations of the retina.

5.1 RECEPTIVE-FIELD SUBSTRUCTURE

In Chapter 2, I illustrated the challenges of linear models to capture retinal encoding. Mounting evidence suggests that the responses of RGCs to natural stimuli involve nonlinear processing in the retina (Heitman et al. 2016; Karamanlis and Gollisch 2021; J. K. Liu et al. 2022; Turner and Rieke 2016). Linear models cannot explain the detection of various features performed by the retina. For instance, subunits enable nonlinear integration of stimuli at a smaller spatial scale than the receptive field (Enroth-Cugell and Robson 1966; Hochstein and Shapley 1976b). Moreover, inhibitory interactions between subunits and the receptive field shape local operations (Franke et al. 2017; Manookin et al. 2008; Werblin 2010). Beyond these, subunits allow to implement local, as opposed to global, adaptation within the receptive field (Garvert and Gollisch 2013; Khani and Gollisch 2017; Ölveczky et al. 2007), equipping the RGCs with broad means of encoding. These insights establish subunits as a ubiquitous concept of retinal processing, attributed to bipolar cells in various species (Demb et al. 2001a; J. Freeman et al. 2015; J. K. Liu et al. 2017; G. W. Schwartz et al. 2012).

5.2 PARASOL AND MIDGET SUBUNITS

Parasol ganglion cells in the primate retina have been linked to nonlinear Y-cell physiology (Crook et al. 2008a,b; Petrusca et al. 2007) and their function was successfully described with subunit models (Shah et al. 2020; Turner and Rieke 2016). Nonlinear responses in midget ganglion cells are less pronounced (Cafaro and Rieke 2013; Petrusca et al. 2007), but were also modeled successfully (Freedland and Rieke 2022; J. Freeman et al. 2015). Nevertheless, subunits of ON and OFF midget and ON and OFF parasol cells have been mostly investigated separately, not allowing for full comparison under the same stimulus and recording conditions. In this thesis, I identified subunits of each of these major RGC types simultaneously with extracellular recordings using MEAs (Chapter 3). I focused on the subunit layouts and compared them with known morphological properties and across cell types.

Subunit sizes of parasol and midget RGC receptive fields matched the physiologically reported receptive fields of diffuse bipolar and midget bipolar cells, respectively (Dacey et al. 2000). In the case of parasol cells, these findings suggest that subunits likely correspond to bipolar cell inputs. Curiously, I found that the average number of recovered subunits per midget RGC underestimated the calculated and measured convergence of midget bipolar cells to midget ganglion cells in peripheral marmoset retina (Chan et al. 2001; Jusuf et al. 2006b). However, my findings compare well to modelled cone convergence in single-cone stimulation in the macaque retina periphery (G. D. Field et al. 2010). The discrepancy could stem from differences in the cone and bipolar densities between marmoset and macaque retinas and especially their convergence over different eccentricities (Chan et al. 2001). A possible explanation is that gap junctions between midget bipolar cells (Kántor et al. 2017) or modulation from amacrine cells (Kuo et al. 2016; Zaghloul et al. 2007) couple several bipolar cells together to a larger, 'effective' bipolar cell subunit (Vaney et al. 2001). An unlikely alternative reason might be optical blurring of the stimulus pixels, evoking responses in several bipolar cells that consequently form one subunit in the context of that stimulation. Nevertheless, the observed systematic subunit mosaic arrangement suggests that my decomposition reflects meaningful biological subunits.

In fact, I found the subunits to regularly space in distinct mosaics arranged by all ganglion cell types, as it is typical for dendritic trees and receptive fields in the retina (Dacey 1993; DeVries and Baylor 1997; Gauthier et al. 2009; Y. S. Liu et al. 2009). In particular, I identified striking resemblance of parasol and midget subunit mosaics with morphological reports. Subunits are closely tiled corresponding to descriptions of bipolar cell dendritic trees (Dacey 1993). In this configuration, any substantial subunit overlap across adjacent RGCs may suggest shared bipolar cell input (J. K. Liu et al. 2017; see Zapp et al. 2022). I observed such shared subunits more often in the parasol mosaics than for midget cells. This resonates with the dendritic trees of parasol cells overlapping with increasing eccentricity, while midget cell dendritic trees do not (Dacey 1993; B. B. Lee et al. 2010). Apart from that, I did not find concrete differences between the respective ON and OFF mosaics. Consequently, observations of higher reliability in responses of OFF cells compared to ON cells during single-cone stimulation (Li et al. 2014) are not

likely to be explained at the level of bipolar cell arrangement. The limited differences in subunit layouts mirror the distribution and density of bipolar cell input (Eriköz et al. 2008) that do not explain the divergence found in ON and OFF response properties (Chichilnisky and Kalmar 2002; Margolis and Detwiler 2007; Zaghloul et al. 2003). This may be another indication for these asymmetries to either originate from nonlinearities at the inputs of bipolar cells (Schreyer and Gollisch 2021) or even at the level of the photoreceptors (see Endeman and Kamermans 2010; see Kling et al. 2019). This signifies the value of subunit mosaics and how they enable to investigate (shared) bipolar cell input on a physiological level, potentially spawning new ideas, and complementing morphological studies.

5.3 SUBUNIT MOSAIC ALIGNMENT

The parallel channels emerging from the retina are encoding distinct aspects of the visual environment (Grünert and Martin 2020). Each feature is typically encoded in pairs of ON- and OFF-types, indicating increment or decrement of the visual feature (Chichilnisky and Kalmar 2002). Among the spatial organization of RGCs over the retina, the distinct ON-type and OFF-type mosaics are of particular interest. Their alignment with respect to each other has revealed how the primate retina samples the visual space and may offer insight into developmental strategies. The grids of RGCs were found to be anti-aligned as would be predicted by efficient coding (Roy et al. 2021). The anti-alignment has been shown to emerge in noisy signal conditions when aiming to minimize redundancy (Jun et al. 2021). Here, I found evidence for the subunit mosaics to be aligned (Chapter 3). In the following paragraphs, I want to shed light on the implications from three different perspectives.

From an encoding perspective, Jun et al. (2021) demonstrated that alignment is the preferred strategy for center-surround encoding when the system produces low noise at its output. When regarding subunits as bipolar cell inputs, both criteria are likely met, as most bipolar cells exhibit antagonistic surrounds (Protti et al. 2014) and most bipolar cells refrain from noise-inducing spike generation at their output (Asari and Meister 2012; Vries et al. 2011). Under these circumstances it seems, that an aligned bipolar cell receptive field arrangement is optimal for non-redundant encoding.

From a circuitry perspective, by this reasoning, the neural layers in the retina become increasingly more organized. There is no distinction between ON and OFF photoreceptors, such that there is no alignment. Sampling from the finely distributed mosaic of photoreceptors, with little demand for selectivity (Jusuf et al. 2006b), bipolar cells then arrange in opposing pairs of aligned mosaics to decrease redundancy in encoding. Ganglion cells of ON and OFF types then pool from aligned arrangements of their presynaptic counterparts to form an anti-aligned grid to facilitate efficient coding under noisy spike generation.

Following this premise, aligned subunit mosaics fall well into place with a hierarchical strategy of retinal encoding. Such a strategy starting in low-level sensory processing is

not unlikely, as the mosaics of ON and OFF RGCs may be carefully organized to give rise to functional tuning maps in the downstream visual cortex (Song et al. 2021).

5.4 STNMF AS SUBUNIT ANALYSIS

To date, subunit inference is not very accessible. My main goal in this thesis was to develop a method to address these ubiquitous challenges present in current approaches (Chapter 3). I will here summarize briefly on what accounts I have achieved this goal and what make my implementation of spike-triggered non-negative matrix factorization (STNMF) suitable for the problem of subunit recovery.

Current methods rely on complex calculations or time-intensive parameter fitting that require powerful computing hardware. I remedied this dependency by refraining from complex layer-based model training or highly constrained calculations. With my implementation, I have replaced expensive calculations of STNMF with an array of faster algorithms. On the one hand, this removes the demand for high-end hardware and on the other hand, this results in a more than 100-fold speed increase compared to a previous implementation of STNMF (J. K. Liu et al. 2017). Subunits of primate RGCs were recovered within a few seconds, and a population within a few minutes. I am not aware of a method reaching an analysis time in that vicinity.

What makes the recovery of subunits particularly challenging is the high dimensionality of data paired with the lack of prior information about the arrangement and number of subunits. Most methods must be run repeatedly on a range of possible numbers of subunits to identify the most likely decomposition. Based on STNMF, the number of subunits are found implicitly (J. K. Liu et al. 2017) by my method. This eliminates the need for repeated runs which drastically reduces total analysis time per cell.

Based on the same procedure, most methods are bound to implement regularization, in form of adjustable constraints to produce components that are localized in space to resemble subunits. To probe all possible combinations of this hyperparameter together with the number of subunits, subunit inference on large populations of individual cells may take tens or even hundreds of hours. Likewise, my method is subjected to regularization as found in almost all methods of subunit inference. However, I showed that for my method the effect of regularization is shared across cells of the same functional type. This reduces the time spent on hyperparameter tuning to running it on small subsets of cells only.

The long analysis time of current approaches is accompanied by long stimulus recordings that deduct valuable recording time from the experiment. Regularization techniques are proving useful for several of the current subunit recovery methods to shorten the required stimulation time (Maheswaranathan et al. 2018; Shah et al. 2020). Likewise, I have demonstrated here that 45 minutes of recorded data may suffice to get a clear estimate of subunits. Recordings of that duration likely have to be performed already for other spike-triggered analyses.

The center-surround structure of bipolar cell receptive fields is an aspect of nonlinear stimulus integration that has not yet been illuminated well. I did not find center-

surround subunits in the primate retina with my implementation of STNMF. However, I have investigated potential algorithms to discover co-existing positive and negative values within one subunit (Chapter 4). Since bipolar-cell surrounds may span a large extent in space (Dacey et al. 2000), they may not be easy to detect using finely structured white noise. It bears the potential for future study to investigate the capabilities of my developed methods in that regard and whether they would in practice aid in the recovery of subunit surrounds.

All these challenges in the current state of subunit inference paint a highly impractical picture when it comes to routine application. I have here compiled a compact analysis and additional algorithms that pave the way to more in-depth investigations of nonlinear spatial encoding in the retina.

5.5 MATHEMATICAL RELEVANCE

While this work was highly geared towards application in sensory neuroscience, the exploration of the subunit inference carries relevance for other fields of computational analysis. By publishing my developed algorithms, the field of mathematics may benefit from the gained insights.

5.5.1 *Exploration of non-negative matrix factorization*

Despite its old age and the seemingly daunting growth of deep learning, the field of non-negative matrix factorization (NMF) (D. D. Lee and Seung 1999; Paatero and Tapper 1994) is still under active investigation with numerous publications in recent years (for instance Ang and Gillis 2019; Hautecoeur et al. 2022; Y. Peng et al. 2019) reporting faster or more accurate implementations. Because certain aspects, like its non-convex nature (Vavasis 2010) or the convergence to a stable solution (Gillis and Glineur 2008), have not been fully elucidated, NMF remains interesting to mathematicians and computer scientists from a theoretical perspective.

In my implementation of Chapter 3, I am relying on sparse semi-non-negative matrix factorization (semi-NMF), a combination of semi-NMF (Ding et al. 2010) and sparsity NMF (Hoyer 2002, 2004; H. Kim and H. Park 2007; J. Kim and H. Park 2008). As there are not many algorithms of this particular combination, I implemented it by fusing state-of-the-art techniques from both areas. While this created in a fast and reliable solution, it introduced an additional benefit. In the typical iterative implementation of NMF, the use of accelerated hierarchical alternating least squares performs multiple smaller updates within one iteration (Gillis and Glineur 2012). My additional sparsity regularization was consequently also applied multiple times per iteration. This resulted in continuous refinement of the subunit decomposition at each subsequent update. The factorization became much more dominated by the subunits, while less prominent structures were progressively suppressed. With fast diminishing uncorrelated noise, the signal-to-noise ratio was amplified. This procedure may provide valuable solutions for efficiently denoising data in matrix factorization applications.

5.5.2 *Hyperparameter tuning for non-negative matrix factorization*

Sparsity regularization is typically implemented by a penalty term that mediates the solution. The impact of the penalty term is specified by a hyperparameter that is carefully selected to achieve the desired result. An approach to determine a suitable sparsity regularization weight did not exist in the context of NMF universal to various implementations (Bro et al. 2008; Owen and Perry 2009). I have undertaken the process of determining an appropriate weight by utilizing consensus methods as a robust method of hyperparameter tuning for NMF (Chapter 3). Consensus methods have been used previously to alleviate different challenges (Brunet et al. 2004; Monti et al. 2003; Zhou et al. 2020). My use of consensus analysis does not depend on the exact formulation of the problem. As it only relies on the measured consensus among repeated applications, it is practical for any NMF implementation irrespective of the use of intricate algorithms that employ vectorized computations with factored out terms. Furthermore, although measuring the consensus extends to clustering techniques, it is particularly well suited for NMF, because of its naturally non-unique solutions space. Lastly, it offers a robust strategy for sparse NMF that is usually described without any means to justify the choice of the regularization parameter (Hoyer 2002, 2004; H. Kim and H. Park 2007; J. Kim and H. Park 2008). The above-mentioned aspects go to show that consensus analysis for hyperparameter selection is relevant beyond the scope of subunit inference.

5.5.3 *Orthogonal regularization*

In Chapter 4, I have explored various constraints of NMF and the alternative of the less aggressive and computationally beneficial regularization. My conclusions might be of interest for the communities of NMF, convolutional neural networks (CNNs), and other fields of machine learning like autoencoders. In particular, I have investigated the option of orthogonal regularization, which has mostly only found application in CNN models (to limited degree) so far and had not yet enjoyed thorough theoretical investigation. Orthogonality as a variable regularization instead of a rigid constraint offers great potential in other practical applications that are currently almost exclusively dominated by $\ell 1$-norm and $\ell 2$-norm regularization.

5.6 APPLICATION OF SUBUNIT INFERENCE

With my developed method, I have attempted to shed light on the differences of parasol and midget ganglion cells in the retina with the investigation of their subunit mosaic (Chapter 3). Following that example, my efforts to develop a practical implementation of STNMF has made subunits accessible for analyses as a compact part of larger experiments of different focus.

5.6.1 *Characterization of RGC cell types*

With my versatile version of STNMF (Chapter 3), the exploration of the subunit layouts may reveal how bipolar cell inputs are integrated into different RGC types. This may range from characterizing the size and number of subunits per RGC, to investigations of systematic sharing of their input.

On the other end of the spectrum is the classification of RGC cell types (Goetz et al. 2022; Rhoades et al. 2019). Cell type clustering based on physiological properties like the spatial receptive field size and their temporal dynamics could be further extended by properties of the subunits. This could potentially lead to a more fine-grained segregation of functional cell types to match the numerous morphologically identified types yet to be explained in their physiology (Bae et al. 2018; Yan et al. 2020a).

Yet another approach is to investigate the stability curve of the consensus analysis (Chapter 3). I have demonstrated that RGCs of the same functional type closely match in solution stability with respect to varying sparsity regularization. The stability curve may offer an alternative to reveal deviations among the cells' nonlinear integration.

5.6.2 *Pairs of subunit mosaics*

Unlike midget ganglion cells, parasol ganglion cells pool inputs from two types of bipolar cells (Bordt et al. 2006; Calkins and Sterling 2007). Because the dendritic trees of bipolar cells tile the retina tightly (Wässle et al. 1994), the subunit layout of a parasol cell should show substantial overlap as evidence of two distinct mosaics (see J. Freeman et al. 2015). However, I observed a uniform tiling of the subunit grids that appear even to be aligned across ON- and OFF-type parasol cells (Chapter 3). As I have established, the subunit sizes match diffuse bipolar cell dendritic trees and even underestimate the measured receptive field sizes (Dacey et al. 2000). Given that it is unlikely that the recovered subunits represent merged bipolar cell receptive fields, this observation begs the question whether and why only one bipolar cell mosaic was recovered.

As speculated in Chapter 2, STNMF shows potential of recovering spatially overlapping subunits. STNMF was proposed to enforce sparse reconstructions (J. K. Liu et al. 2017) penalized by each pixel across all spatial components. My implementation, on the contrary, regularizes within each spatial component. On the one hand that disassociates the individual subunits from each other and on the other hand, it allows a pixel to exhibit high values in multiple subunits, effectively enabling subunit overlap. Since I found no evidence for subunits with relative displacement, the two populations of bipolar cells may operate in separate stimulus regimes. The diffuse bipolar cells DB2 respond to lower temporal frequencies (Freed 2000), while DB3 cells react faster (Calkins and Sterling 2007). The latter represent around 60% of the ribbon contacts to OFF parasol ganglion cells (Calkins and Sterling 2007; Jacoby et al. 2000). Given the high temporal frequency of the stimulation used in the study of Chapter 3, one would hypothesize that the ganglion cell activation may be dominated by the richer DB3 input. With my fast version of STNMF at hand, adjusting the temporal dynamics of the stimulus would enable to confirm a secondary subunit mosaic or to rectify that hypothesis.

5.6.3 *Beyond the retina*

In Chapter 2, I iterated the successes in explaining the sensory processing with a linear-nonlinear-linear-nonlinear (LNLN) model in the primary and secondary visual cortex (Almasi et al. 2020; Bartsch et al. 2022; Lochmann et al. 2013; Rust et al. 2005; Vintch et al. 2015; Wu et al. 2015) and in areas of visual motion processing (Beyeler et al. 2016; Mineault et al. 2012). Chapter 2 also touched upon the potential opportunities for LNLN models in other sensory systems like in auditory processing (Ahrens et al. 2008; Keshishian et al. 2020; McFarland et al. 2013). The converging signal propagation of connected neurons as found between the layers of the retina, may underlie a reoccurring circuit motif in the sensory systems of the brain (R. J. Douglas et al. 1989; Fukushima 1980; Heeger et al. 1996; Riesenhuber and Poggio 1999). Advances in experimental techniques, like optogenetic stimulation (Papagiakoumou 2013), may allow for simulation and spike recordings of these intricate networks in the cortex. That would allow the inference of intermediate signalling using methods of subunit recovery. As the sensory systems essentially perform dimensionality reduction through sparse representation (Olshausen and D. J. Field 1996, 1997), my implementation of STNMF might be particularly suitable, because it only presupposes sparsity and is otherwise agnostic to the modality. To extend upon this, my efforts from Chapter 4 may offer insights concerning joint representations of positive and negative valued components.

5.6.4 *Closed-loop analyses*

Unlike classical recordings, closed-loop experiments allow iterative adjustments and allow more targeted stimulation (Bölinger and Gollisch 2012; Edin et al. 2004). These valuable studies have been performed to limited extent in the context of subunit analysis (Rhoades et al. 2019; Shah et al. 2020; Takeshita and Gollisch 2014). These analyses are difficult, because they rely on fast cell identification and analysis methods. Due to its fast convergence, STNMF is able to make subunits available during close-loop experiments in a practical time frame. This offers the possibility to confirm nonlinear stimulus integration online and to record the activity of bipolar cells - indirectly through RGC spiking. Furthermore, it poses the opportunity to perform more in-depth experiments on the live subunits. This could involve probing the identified subunits individually or in different pairs to obtain a better understanding of their integration and interplay. Direct investigations of retinal computation phenomena would be more accessible. Of great interest are studies on the temporal dynamics of nonlinear signal integration, like motion-detection (Demb et al. 2001b; B. Liu et al. 2021; Ölveczky et al. 2003) and local adaptation (Garvert and Gollisch 2013; Khani and Gollisch 2017). Also, subunit mosaics in different light intensities or stimulus regimes could be explored. Moreover, pharmacological experiments would allow to decompose nonlinear processing on a cellular basis. Inhibitory inputs like amacrine cells could be blocked to understand their impact on 'effective' bipolar-cell subunits.

5.6.5 *Vision restoration*

In Chapter 1, I have discussed the implications of photoreceptor degeneration in the presence of retinitis pigmentosa and the importance of understanding retinal encoding for developing optogenetic tools for vision restoration. As the optogenetic targeting of bipolar cells is showing promise (Cehajic-Kapetanovic et al. 2015; Kralik et al. 2022; Lagali et al. 2008), the computations in the inner retina can be preserved, as well as the emergence of parallel encoding channels (M. H. Berry et al. 2019; Wyk et al. 2015).

Developments on selective bipolar-cell targeting progresses constantly. An example are ON bipolar promoters (Hulliger et al. 2020) that have enabled natural-like ganglion cell encoding, including transient and sustained responses and adaptation to ambient light intensities in mice (Kralik et al. 2022).

As systematic optogenetic experiments in marmosets come into reach (J. E. Park and Silva 2019; Sasaki et al. 2009), deciphering the properties and differences in the subunit mosaics of parasol and midget ganglion cells may offer hypotheses for optogenetic studies. A first step may be to run subunit analyses in experiments of optogenetic manipulated retinas and compare if stimulation drives the inner retina in a way that subunits are preserved. I have demonstrated the ease at which my method reconstructs bipolar-cell subunit mosaics of multiple RGC types of a single recording. To verify type-selectivity of bipolar cells in optogenetic studies, my efforts will be a valuable contribution.

5.7 CONCLUSION

In this thesis, I have established the importance and implications of nonlinear spatial computations of the retina. Many features that the retina extracts are based on these principles but are not well understood. This impacts the general understanding of early visual encoding, insights into reoccurring motifs of sensory processing, and medical applications for vision restoration. Flexible computational methods are needed to decipher the presynaptic code of ganglion cells and its nonlinear integration. I have outlined the challenges of the current state of computational approaches and addressed them with the development of new implementations. With multi-electrode recordings, I have demonstrated the ease of use and unmatched analysis speed, allowing the reconstruction of bipolar-cell subunit mosaics of several ganglion cell types in a population. Investigations of regularization revealed similarities across cells of the same type and offer a comprehensive hyperparameter selection. In conclusion, I have paved the way for more accessible subunit inference, demonstrated new insights in the arrangement of the primate retina, and raised new questions about the visual processing in and beyond the retina.

BIBLIOGRAPHY

Aertsen, A. M. H. J. and Johannesma, P. I. M. (1981). "The Spectro-Temporal Receptive Field". *Biol. Cybern.* 42.2, 133–143 (cit. on p. 4).

Ahrens, M. B., Linden, J. F., and Sahani, M. (2008). "Nonlinearities and Contextual Influences in Auditory Cortical Responses Modeled with Multilinear Spectrotemporal Methods". *J. Neurosci.* 28.8, 1929–1942 (cit. on p. 88).

Almasi, A., Meffin, H., Cloherty, S. L., Wong, Y., Yunzab, M., and Ibbotson, M. R. (2020). "Mechanisms of Feature Selectivity and Invariance in Primary Visual Cortex". *Cereb. Cortex* 30.9, 5067–5087 (cit. on p. 88).

Ang, A. M. S. and Gillis, N. (2019). "Accelerating Nonnegative Matrix Factorization Algorithms Using Extrapolation". *Neural Comput.* 31.2, 417–439 (cit. on p. 85).

Asari, H. and Meister, M. (2012). "Divergence of visual channels in the inner retina". *Nat. Neurosci.* 15.11, 1581–1589 (cit. on pp. 5, 9, 10, 83).

Atick, J. J. and Redlich, A. N. (1992). "What Does the Retina Know about Natural Scenes?" *Neural Comput.* 4.2, 196–210 (cit. on p. 4).

Baden, T., Berens, P., Franke, K., Román Rosón, M., Bethge, M., and Euler, T. (2016). "The functional diversity of retinal ganglion cells in the mouse". *Nature* 529.7586, 345–350 (cit. on p. 9).

Baden, T. and Osorio, D. (2019). "The Retinal Basis of Vertebrate Color Vision". *Annu. Rev. Vis. Sci.* 5.1, 177–200 (cit. on p. 11).

Bae, J. A., Mu, S., Kim, J. S., Turner, N. L., Tartavull, I., Kemnitz, N., Jordan, C. S., Norton, A. D., Silversmith, W. M., Prentki, R., Sorek, M., David, C., Jones, D. L., Bland, D., Sterling, A. L. R., Park, J., Briggman, K. L., and Seung, H. S. (2018). "Digital Museum of Retinal Ganglion Cells with Dense Anatomy and Physiology". *Cell* 173.5, 1293–1306.e19 (cit. on pp. 4, 87).

Barlow, H. B. (1953). "Summation and inhibition in the frog's retina". *J. Physiol.* 119.1, 69–88 (cit. on p. 4).

Barlow, H. B. (1961). "Possible principles underlying the transformation of sensory messages". *Sens. Commun.*, 217–234 (cit. on p. 4).

Barlow, H. B., Hill, R. M., and Levick, W. R. (1964). "Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit". *J. Physiol.* 173.3, 377–407 (cit. on p. 11).

Bartsch, F., Cumming, B. G., and Butts, D. A. (2022). "Model-based characterization of the selectivity of neurons in primary visual cortex". *J. Neurophysiol.* 128.2, 350–363 (cit. on p. 88).

Benardete, E. A., Kaplan, E., and Knight, B. W. (1992). "Contrast gain control in the primate retina: P cells are not X-like, some M cells are". *Vis. Neurosci.* 8.5, 483–486 (cit. on p. 8).

Berry, M. H., Holt, A., Salari, A., Veit, J., Visel, M., Levitz, J., Aghi, K., Gaub, B. M., Sivyer, B., Flannery, J. G., and Isacoff, E. Y. (2019). "Restoration of high-sensitivity and adapting vision with a cone opsin". *Nat. Commun.* 10.1, 1221 (cit. on pp. 1, 89).

Beyeler, M., Dutt, N., and Krichmar, J. L. (2016). "3D Visual Response Properties of MSTd Emerge from an Efficient, Sparse Population Code". *J. Neurosci.* 36.32, 8399–8415 (cit. on p. 88).

Bohl, J. M., Shehu, A., Hellmer, C. B., and Ichinose, T. (2022). "Patch clamp recording from bipolar cells in the wholemount mouse retina". *STAR Protoc.* 3.3, 101482 (cit. on p. 10).

Bölinger, D. and Gollisch, T. (2012). "Closed-Loop Measurements of Iso-Response Stimuli Reveal Dynamic Nonlinear Stimulus Integration in the Retina". *Neuron* 73.2, 333–346 (cit. on pp. 5, 88).

Bordt, A. S., Hoshi, H., Yamada, E. S., Perryman-Stout, W. C., and Marshak, D. W. (2006). "Synaptic input to OFF parasol ganglion cells in macaque retina". *J. Comp. Neurol.* 498.1, 46–57 (cit. on p. 87).

Borghuis, B. G., Marvin, J. S., Looger, L. L., and Demb, J. B. (2013). "Two-Photon Imaging of Nonlinear Glutamate Release Dynamics at Bipolar Cell Synapses in the Mouse Retina". *J. Neurosci.* 33.27, 10972–10985 (cit. on p. 5).

Bro, R., Kjeldahl, K., Smilde, A. K., and Kiers, H. A. L. (2008). "Cross-validation of component models: A critical look at current methods". *Anal. Bioanal. Chem.* 390.5, 1241–1251 (cit. on p. 86).

Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2004). "Metagenes and molecular pattern discovery using matrix factorization." *Proc. Natl. Acad. Sci. U. S. A.* 101.12, 4164–9 (cit. on p. 86).

Cafaro, J. and Rieke, F. (2013). "Regulation of Spatial Selectivity by Crossover Inhibition". *J. Neurosci.* 33.15, 6310–6320 (cit. on p. 82).

Cajal, S. R. y (1893). "La rétine des vertébrés". *Cellule* 9, 119–257 (cit. on pp. 4, 8).

Calkins, D. J. and Sterling, P. (2007). "Microcircuitry for Two Types of Achromatic Ganglion Cell in Primate Fovea". *J. Neurosci.* 27.10, 2646–2653 (cit. on p. 87).

Carlson, D. and Carin, L. (2019). "Continuing progress of spike sorting in the era of big data". *Curr. Opin. Neurobiol.* 55, 90–96 (cit. on p. 9).

Cehajic-Kapetanovic, J., Eleftheriou, C., Allen, A. E., Milosavljevic, N., Pienaar, A., Bedford, R., Davis, K. E., Bishop, P. N., and Lucas, R. J. (2015). "Restoration of Vision with Ectopic Expression of Human Rod Opsin". *Curr. Biol.* 25.16, 2111–2122 (cit. on pp. 1, 89).

Chan, T. L., Goodchild, A. K., and Martin, P. R. (1997). "The morphology and distribution of horizontal cells in the retina of a New World monkey, the marmoset Callithrix jacchus : A comparison with macaque monkey". *Vis. Neurosci.* 14.1, 125–140 (cit. on p. 8).

Chan, T. L. and Grünert, U. (1998). "Horizontal cell connections with short wavelength-sensitive cones in the retina: A comparison between New World and Old World primates". *J. Comp. Neurol.* 393.2, 196–209 (cit. on p. 8).

Chan, T. L., Martin, P. R., Clunas, N., and Grünert, U. (2001). "Bipolar cell diversity in the primate retina: Morphologic and immunocytochemical analysis of a new world monkey, the marmoset Callithrix jacchus". *J. Comp. Neurol.* 437.2, 219–239 (cit. on pp. 8, 82).

Chichilnisky, E. J. (2001). "A simple white noise analysis of neuronal light responses". *Netw. Comput. Neural Syst.* 12.2, 199–213 (cit. on pp. 4, 9).

Chichilnisky, E. J. and Kalmar, R. S. (2002). "Functional Asymmetries in ON and OFF Ganglion Cells of Primate Retina". *J. Neurosci.* 22.7, 2737–2747 (cit. on p. 83).

Croner, L. J. and Kaplan, E. (1995). "Receptive fields of P and M ganglion cells across the primate retina". *Vision Res.* 35.1, 7–24 (cit. on p. 8).

Crook, J. D., Peterson, B. B., Packer, O. S., Robinson, F. R., Troy, J. B., and Dacey, D. M. (2008a). "Y-Cell Receptive Field and Collicular Projection of Parasol Ganglion Cells in Macaque Monkey Retina". *J. Neurosci.* 28.44, 11277–11291 (cit. on pp. 8, 82).

Crook, J. D., Peterson, B. B., Packer, O. S., Robinson, F. R., Gamlin, P. D., Troy, J. B., and Dacey, D. M. (2008b). "The Smooth Monostratified Ganglion Cell: Evidence for Spatial Diversity in the Y-Cell Pathway to the Lateral Geniculate Nucleus and Superior Colliculus in the Macaque Monkey". *J. Neurosci.* 28.48, 12654–12671 (cit. on pp. 8, 82).

Dacey, D. M. (1993). "The mosaic of midget ganglion cells in the human retina". *J. Neurosci.* 13.12, 5334–5355 (cit. on pp. 2, 8, 82).

Dacey, D. M. (2004). "Origins of perception: retinal ganglion cell diversity and the creation of parallel visual pathways". In: *Cogn. Neurosci.* Ed. by M. S. Gazzaniga. 3rd ed. Cambridge: MIT Press, 281–301 (cit. on pp. 1, 4, 8).

Dacey, D. M. and Petersen, M. R. (1992). "Dendritic field size and morphology of midget and parasol ganglion cells of the human retina". *Proc. Natl. Acad. Sci. U. S. A.* 89.20, 9666–9670 (cit. on p. 8).

Dacey, D. M., Packer, O. S., Diller, L., Brainard, D., Peterson, B., and Lee, B. (2000). "Center surround receptive field structure of cone bipolar cells in primate retina". *Vision Res.* 40.14, 1801–1811 (cit. on pp. 82, 85, 87).

De Boer, E. and Kuyper, P. (1968). "Triggered Correlation". *IEEE Trans. Biomed. Eng.* 15.3, 169–179 (cit. on p. 4).

DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1995). "Receptive-field dynamics in the central visual pathways". *Trends Neurosci.* 18.10, 451–458 (cit. on p. 4).

Deisseroth, K., Feng, G., Majewska, A. K., Miesenbock, G., Ting, A., and Schnitzer, M. J. (2006). "Next-Generation Optical Technologies for Illuminating Genetically Targeted Brain Circuits". *J. Neurosci.* 26.41, 10380–10386 (cit. on p. 1).

Demb, J. B., Haarsma, L., Freed, M. A., and Sterling, P. (1999). "Functional circuitry of the retinal ganglion cell's nonlinear receptive field". *J. Neurosci.* 19.22, 9756–9767 (cit. on p. 5).

Demb, J. B., Zaghloul, K. A., Haarsma, L., and Sterling, P. (2001a). "Bipolar Cells Contribute to Nonlinear Spatial Summation in the Brisk-Transient (Y) Ganglion Cell in Mammalian Retina". *J. Neurosci.* 21.19, 7447–7454 (cit. on pp. 2, 5, 81).

Demb, J. B., Zaghloul, K. A., and Sterling, P. (2001b). "Cellular Basis for the Response to Second-Order Motion Cues in Y Retinal Ganglion Cells". *Neuron* 32.4, 711–721 (cit. on pp. 7, 88).

Derrington, A. M., Lennie, P., and Wright, M. J. (1979). "The mechanism of peripherally evoked responses in retinal ganglion cells." *J. Physiol.* 289.1, 299–310 (cit. on p. 5).

Desrosiers, J., Wanet-Defalque, M.-C., Témisjian, K., Gresset, J., Dubois, M.-F., Renaud, J., Vincent, C., Rousseau, J., Carignan, M., and Overbury, O. (2009). "Participation in daily

activities and social roles of older adults with visual impairment". *Disabil. Rehabil.* 31.15, 1227–1234 (cit. on p. 1).

DeVries, S. H. and Baylor, D. A. (1997). "Mosaic arrangement of ganglion cell receptive fields in rabbit retina". *J. Neurophysiol.* 78.4, 2048–2060 (cit. on p. 82).

DiCarlo, J. J. and Johnson, K. O. (2000). "Spatial and Temporal Structure of Receptive Fields in Primate Somatosensory Area 3b: Effects of Stimulus Scanning Direction and Orientation". *J. Neurosci.* 20.1, 495–510 (cit. on p. 4).

Ding, C., Tao Li, and Jordan, M. I. (2010). "Convex and Semi-Nonnegative Matrix Factorizations". *IEEE Trans. Pattern Anal. Mach. Intell.* 32.1, 45–55 (cit. on p. 85).

Douglas, R. J., Martin, K. A. C., and Whitteridge, D. (1989). "A Canonical Microcircuit for Neocortex". *Neural Comput.* 1.4, 480–488 (cit. on p. 88).

Drasdo, N. and Fowler, C. W. (1974). "Non-linear projection of the retinal image in a wide-angle schematic eye". *Br. J. Ophthalmol.* 58.8, 709–714 (cit. on p. 2).

Drinnenberg, A., Franke, F., Morikawa, R. K., Jüttner, J., Hillier, D., Hantz, P., Hierlemann, A., Azeredo da Silveira, R., and Roska, B. (2018). "How Diverse Retinal Functions Arise from Feedback at the First Visual Synapse". *Neuron* 99.1, 117–134.e11 (cit. on p. 9).

Edin, F., Machens, C. K., Schütze, H., and Edin, F. (2004). "Searching for Optimal Sensory Signals: Iterative Stimulus Reconstruction in Closed-Loop Experiments". *J. Comput. Neurosci.* 17.1, 47–56 (cit. on p. 88).

Endeman, D. and Kamermans, M. (2010). "Cones perform a non-linear transformation on natural stimuli". *J. Physiol.* 588.3, 435–446 (cit. on p. 83).

Enroth-Cugell, C. and Robson, J. G. (1966). "The contrast sensitivity of retinal ganglion cells of the cat". *J. Physiol.* 187.3, 517–552 (cit. on pp. 4, 6, 7, 81).

Enroth-Cugell, C. and Pinto, L. (1970). "Algebraic summation of centre and surround inputs to retinal ganglion cells of the cat". *Nature* 226.5244, 458–459 (cit. on p. 4).

Enroth-Cugell, C. and Robson, J. G. (1984). "Functional characteristics and diversity of cat retinal ganglion cells. Basic characteristics and quantitative description". *Invest. Ophthalmol. Vis. Sci.* 25.3, 250–67 (cit. on p. 6).

Enroth-Cugell, C. and Freeman, A. W. (1987). "The receptive-field spatial structure of cat retinal Y cells". *J. Physiol.* 384.1, 49–79 (cit. on p. 5).

Eriköz, B., Jusuf, P. R., Percival, K. A., and Grünert, U. (2008). "Distribution of bipolar input to midget and parasol ganglion cells in marmoset retina". *Vis. Neurosci.* 25.1, 67–76 (cit. on pp. 8, 83).

Field, G. D. and Chichilnisky, E. J. (2007). "Information Processing in the Primate Retina: Circuitry and Coding". *Annu. Rev. Neurosci.* 30.1, 1–30 (cit. on p. 8).

Field, G. D., Gauthier, J. L., Sher, A., Greschner, M., Machado, T. A., Jepson, L. H., Shlens, J., Gunning, D. E., Mathieson, K., Dabrowski, W., Paninski, L., Litke, A. M., and Chichilnisky, E. J. (2010). "Functional connectivity in the retina at the resolution of photoreceptors". *Nature* 467.7316, 673–677 (cit. on p. 82).

Fischer, B., Krüger, J., and Droll, W. (1975). "Quantitative aspects of the shift-effect in cat retinal ganglion cells". *Brain Res.* 83.3, 391–403 (cit. on p. 5).

Foster, R. G., Provencio, I., Hudson, D., Fiske, S., De Grip, W., and Menaker, M. (1991). "Circadian photoreception in the retinally degenerate mouse (rd/rd)". *J. Comp. Physiol. A* 169.1, 39–50 (cit. on p. 2).

Franke, K., Berens, P., Schubert, T., Bethge, M., Euler, T., and Baden, T. (2017). "Inhibition decorrelates visual feature representations in the inner retina". *Nature* 542.7642, 439–444 (cit. on p. 81).

Freed, M. A. (2000). "Parallel Cone Bipolar Pathways to a Ganglion Cell Use Different Rates and Amplitudes of Quantal Excitation". *J. Neurosci.* 20.11, 3956–3963 (cit. on p. 87).

Freedland, J. and Rieke, F. (2022). "Systematic reduction of the dimensionality of natural scenes allows accurate predictions of retinal ganglion cell spike outputs". *Proc. Natl. Acad. Sci. U. S. A.* 119.46 (cit. on pp. 9, 82).

Freeman, J., Field, G. D., Li, P. H., Greschner, M., Gunning, D. E., Mathieson, K., Sher, A., Litke, A. M., Paninski, L., Simoncelli, E. P., and Chichilnisky, E. J. (2015). "Mapping nonlinear receptive field structure in primate retina at single cone resolution". *Elife* 4.39, 55–68 (cit. on pp. 5, 9, 10, 81, 82, 87).

Fukushima, K. (1980). "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". *Biol. Cybern.* 36.4, 193–202 (cit. on p. 88).

Garvert, M. M. and Gollisch, T. (2013). "Local and Global Contrast Adaptation in Retinal Ganglion Cells". *Neuron* 77.5, 915–928 (cit. on pp. 81, 88).

Gauthier, J. L., Field, G. D., Sher, A., Greschner, M., Shlens, J., Litke, A. M., and Chichilnisky, E. J. (2009). "Receptive Fields in Primate Retina Are Coordinated to Sample Visual Space More Uniformly". *PLoS Biol.* 7.4, e1000063 (cit. on pp. 9, 82).

Ghosh, K. K., Goodchild, A. K., Sefton, A. E., and Martin, P. R. (1996). "Morphology of retinal ganglion cells in a New World monkey, the marmoset Callithrix jacchus". *J. Comp. Neurol.* 366.1, 76–92 (cit. on p. 8).

Gillis, N. and Glineur, F. (2008). "Nonnegative Factorization and The Maximum Edge Biclique Problem". *arXiv*, 0810.4225v1 (cit. on p. 85).

Gillis, N. and Glineur, F. (2012). "Accelerated Multiplicative Updates and Hierarchical ALS Algorithms for Nonnegative Matrix Factorization". *Neural Comput.* 24.4, 1085–1105 (cit. on p. 85).

Goetz, J., Jessen, Z. F., Jacobi, A., Mani, A., Cooler, S., Greer, D., Kadri, S., Segal, J., Shekhar, K., Sanes, J. R., and Schwartz, G. W. (2022). "Unified classification of mouse retinal ganglion cells using function, morphology, and gene expression". *Cell Rep.* 40.2, 111040 (cit. on pp. 9, 87).

Gollisch, T. (2013). "Features and functions of nonlinear spatial integration by retinal ganglion cells". *J. Physiol.* 107.5, 338–348 (cit. on p. 11).

Gollisch, T. and Meister, M. (2010). "Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina". *Neuron* 65.2, 150–164 (cit. on pp. 1, 5, 11).

Gomes, F. L., Silveira, L. C. L., Saito, C. A., and Yamada, E. S. (2005). "Density, proportion, and dendritic coverage of retinal ganglion cells of the common marmoset (Callithrix jacchus jacchus)". *Brazilian J. Med. Biol. Res.* 38.6, 915–924 (cit. on p. 8).

Goodchild, A. K., Ghosh, K. K., and Martin, P. R. (1996). "Comparison of photoreceptor spatial density and ganglion cell morphology in the retina of human, macaque monkey, cat, and the marmosetCallithrix jacchus". *J. Comp. Neurol.* 366.1, 55–75 (cit. on p. 8).

Grimes, W. N., Songco-Aguas, A., and Rieke, F. (2018). "Parallel Processing of Rod and Cone Signals: Retinal Function and Human Perception". *Annu. Rev. Vis. Sci.* 4.1, 123–141 (cit. on p. 8).

Grünert, U. and Martin, P. R. (2020). "Cell types and cell circuits in human and non-human primate retina". *Prog. Retin. Eye Res.* 78, 100844 (cit. on pp. 1, 83).

Hartline, H. K. (1938). "The response of single optic nerve fibers of the vertebrate eye to illumination of the retina". *Am. J. Physiol.* 121, 400–415 (cit. on p. 2).

Hartong, D., Berson, E., and Dryja, T. (2006). "Retinitis pigmentosa Prevalence and inheritance patterns". *Lancet* 368, 1795–1809 (cit. on p. 1).

Hautecoeur, C., De Lathauwer, L., Gillis, N., and Glineur, F. (2022). "Least-squares methods for nonnegative matrix factorization over rational functions". *arXiv*, 2209.12579v1 (cit. on p. 85).

Haverkamp, S., Reinhard, K., Peichl, L., and Mietsch, M. (2022). "No evidence for age-related alterations in the marmoset retina". *Front. Neuroanat.* 16, 62 (cit. on p. 8).

Heeger, D. J., Simoncelli, E. P., and Movshon, J. A. (1996). "Computational models of cortical visual processing". *Proc. Natl. Acad. Sci. U. S. A.* 93.2, 623–627 (cit. on p. 88).

Heine, C. and Browning, C. J. (2002). "Communication and psychosocial consequences of sensory loss in older adults: overview and rehabilitation directions". *Disabil. Rehabil.* 24.15, 763–773 (cit. on p. 1).

Heitman, A., Brackbill, N., Greschner, M., Sher, A., Litke, A. M., and Chichilnisky, E. J. (2016). "Testing pseudo-linear models of responses to natural scenes in primate retina". *bioRxiv*, 045336 (cit. on p. 81).

Hochstein, S. and Shapley, R. M. (1976a). "Quantitative analysis of retinal ganglion cell classifications". *J. Physiol.* 262.2, 237–264 (cit. on pp. 5, 7).

Hochstein, S. and Shapley, R. M. (1976b). "Linear and nonlinear spatial subunits in Y cat retinal ganglion cells". *J. Physiol.* 262.2, 265–284 (cit. on pp. 5, 81).

Hoyer, P. O. (2002). "Non-negative sparse coding". In: *Proc. 12th IEEE Work. Neural Networks Signal Process.* IEEE, 557–565 (cit. on pp. 85, 86).

Hoyer, P. O. (2004). "Non-negative matrix factorization with sparseness constraints". *J. Mach. Learn. Res.* 5, 1457–1469 (cit. on pp. 85, 86).

Hubel, D. H. and Wiesel, T. N. (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". *J. Physiol.* 160.1, 106–154 (cit. on p. 4).

Hulliger, E. C., Hostettler, S. M., and Kleinlogel, S. (2020). "Empowering Retinal Gene Therapy with a Specific Promoter for Human Rod and Cone ON-Bipolar Cells". *Mol. Ther. - Methods Clin. Dev.* 17.June, 505–519 (cit. on p. 89).

Jacobs, G. H. (2008). "Primate color vision: A comparative perspective". *Vis. Neurosci.* 25.5-6, 619–633 (cit. on p. 7).

Jacoby, R. A., Wiechmann, A. F., Amara, S. G., Leighton, B. H., and Marshak, D. W. (2000). "Diffuse bipolar cells provide input to OFF parasol ganglion cells in the macaque retina". *J. Comp. Neurol.* 416.1, 6–18 (cit. on p. 87).

Jadzinsky, P. D. and Baccus, S. A. (2015). "Synchronized amplification of local information transmission by peripheral retinal input". *Elife* 4, 1–24 (cit. on p. 7).

Jun, N. Y., Field, G. D., and Pearson, J. (2021). "Scene statistics and noise determine the relative arrangement of receptive field mosaics". *Proc. Natl. Acad. Sci. U. S. A.* 118.39 (cit. on p. 83).

Jusuf, P. R., Lee, S. C. S., and Grünert, U. (2004). "Synaptic connectivity of the diffuse bipolar cell type DB6 in the inner plexiform layer of primate retina". *J. Comp. Neurol.* 469.4, 494–506 (cit. on p. 8).

Jusuf, P. R., Martin, P. R., and Grünert, U. (2006a). "Synaptic connectivity in the midget-parvocellular pathway of primate central retina". *J. Comp. Neurol.* 494.2, 260–274 (cit. on p. 8).

Jusuf, P. R., Martin, P. R., and Grünert, U. (2006b). "Random wiring in the midget pathway of primate retina." *J. Neurosci.* 26.15, 3908–17 (cit. on pp. 82, 83).

Kaardal, J., Fitzgerald, J. D., Berry, M. J., and Sharpee, T. O. (2013). "Identifying Functional Bases for Multidimensional Neural Computations". *Neural Comput.* 25.7, 1870–1890 (cit. on p. 11).

Kántor, O., Varga, A., Nitschke, R., Naumann, A., Énzsöly, A., Lukáts, Á., Szabó, A., Németh, J., and Völgyi, B. (2017). "Bipolar cell gap junctions serve major signaling pathways in the human retina". *Brain Struct. Funct.* 222.6, 2603–2624 (cit. on p. 82).

Karamanlis, D. and Gollisch, T. (2021). "Nonlinear Spatial Integration Underlies the Diversity of Retinal Ganglion Cell Responses to Natural Images". *J. Neurosci.* 41.15, 3479–3498 (cit. on p. 81).

Keat, J., Reinagel, P., Reid, R. C., and Meister, M. (2001). "Predicting Every Spike". *Neuron* 30.3, 803–817 (cit. on p. 4).

Keshishian, M., Akbari, H., Khalighinejad, B., Herrero, J. L., Mehta, A. D., and Mesgarani, N. (2020). "Estimating and interpreting nonlinear receptive field of sensory neural responses with deep neural network models". *Elife* 9, 1–24 (cit. on p. 88).

Khani, M. H. and Gollisch, T. (2017). "Diversity in spatial scope of contrast adaptation among mouse retinal ganglion cells". *J. Neurophysiol.* 118.6, 3024–3043 (cit. on pp. 81, 88).

Kim, H. and Park, H. (2007). "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis". *Bioinformatics* 23.12, 1495–1502 (cit. on pp. 85, 86).

Kim, J. and Park, H. (2008). "Sparse Nonnegative Matrix Factorization for Clustering". *Georg. Inst. Technol.* Tech. Rep. GT-CSE-08-01 (cit. on pp. 85, 86).

Kistler, W. M., Gerstner, W., and Hemmen, J. L. van (1997). "Reduction of the Hodgkin-Huxley Equations to a Single-Variable Threshold Model". *Neural Comput.* 9.5, 1015–1045 (cit. on p. 4).

Kling, A., Field, G. D., Brainard, D. H., and Chichilnisky, E. J. (2019). "Probing Computation in the Primate Visual System at Single-Cone Resolution". *Annu. Rev. Neurosci.* 42.1, annurev–neuro–070918–050233 (cit. on p. 83).

Kling, A., Gogliettino, A. R., Shah, N. P., Wu, E. G., Brackbill, N., Sher, A., Litke, A. M., Silva, R. A., and Chichilnisky, E. J. (2020). "Functional Organization of Midget and Parasol Ganglion Cells in the Human Retina". *bioRxiv*, 240762v3 (cit. on p. 7).

Kolb, H., Nelson, R., and Mariani, A. (1981). "Amacrine cells, bipolar cells and ganglion cells of the cat retina: A Golgi study". *Vision Res.* 21.7, 1081–1114 (cit. on p. 4).

Kolb, H., Linberg, K. A., and Fisher, S. K. (1992). "Neurons of the human retina: A Golgi study". *J. Comp. Neurol.* 318.2, 147–187 (cit. on pp. 2, 8).

Kolb, H. and Marshak, D. (2003). "The midget pathways of the primate retina". *Doc. Ophthalmol.* 106.1, 67–81 (cit. on p. 8).

Kralik, J., Wyk, M. van, Stocker, N., and Kleinlogel, S. (2022). "Bipolar cell targeted optogenetic gene therapy restores parallel retinal signaling and high-level vision in the degenerated retina". *Commun. Biol.* 5.1, 1116 (cit. on pp. 1, 89).

Krubitzer, L. (2007). "The Magnificent Compromise: Cortical Field Evolution in Mammals". *Neuron* 56.2, 201–208 (cit. on p. 7).

Kuffler, S. W. (1953). "Discharge Patterns and Functional Organization of Mammalian Retina". *J. Neurophysiol.* 16.1, 37–68 (cit. on pp. 4, 11).

Kuo, S. P., Schwartz, G. W., and Rieke, F. (2016). "Nonlinear Spatiotemporal Integration by Electrical and Chemical Synapses in the Retina". *Neuron* 90.2, 320–332 (cit. on pp. 5, 82).

Lagali, P. S., Balya, D., Awatramani, G. B., Münch, T. A., Kim, D. S., Busskamp, V., Cepko, C. L., and Roska, B. (2008). "Light-activated channels targeted to ON bipolar cells restore visual function in retinal degeneration". *Nat. Neurosci.* 11.6, 667–675 (cit. on pp. 1, 89).

Lee, B. B. (1996). "Receptive field structure in the primate retina". *Vision Res.* 36.5, 631–644 (cit. on p. 8).

Lee, B. B., Martin, P. R., and Grünert, U. (2010). "Retinal connectivity and primate vision". *Prog. Retin. Eye Res.* 29.6, 622–639 (cit. on p. 82).

Lee, D. D. and Seung, H. S. (1999). "Learning the parts of objects by non-negative matrix factorization". *Nature* 401.6755, 788–791 (cit. on pp. 11, 85).

Lee, J., Mitelut, C., Shokri, H., Kinsella, I., Dethe, N., Wu, S., Li, K., Reyes, E. B., Turcu, D., Batty, E., Kim, Y. J., Brackbill, N., Kling, A., Goetz, G., Chichilnisky, E. J., Carlson, D., and Paninski, L. (2020). "YASS: Yet Another Spike Sorter applied to large-scale multi-electrode array recordings in primate retina". *bioRxiv*, 997924v1 (cit. on p. 9).

Li, P. H., Field, G. D., Greschner, M., Ahn, D., Gunning, D. E., Mathieson, K., Sher, A., Litke, A. M., and Chichilnisky, E. J. (2014). "Retinal Representation of the Elementary Visual Signal". *Neuron* 81.1, 130–139 (cit. on p. 82).

Lindner, M., Gilhooley, M. J., Hughes, S., and Hankins, M. W. (2022). "Optogenetics for visual restoration: From proof of principle to translational challenges". *Prog. Retin. Eye Res.* 91.February, 101089 (cit. on p. 1).

Liu, B., Hong, A., Rieke, F., and Manookin, M. B. (2021). "Predictive encoding of motion begins in the primate retina". *Nat. Neurosci.* 24.9, 1280–1291 (cit. on p. 88).

Liu, J. K., Schreyer, H. M., Onken, A., Rozenblit, F., Khani, M. H., Krishnamoorthy, V., Panzeri, S., and Gollisch, T. (2017). "Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization". *Nat. Commun.* 8.1, 149 (cit. on pp. 2, 10, 11, 81, 82, 84, 87).

Liu, J. K., Karamanlis, D., and Gollisch, T. (2022). "Simple model for encoding natural images by retinal ganglion cells with nonlinear spatial integration". *PLOS Comput. Biol.* 18.3, e1009925 (cit. on p. 81).

Liu, Y. S., Stevens, C. F., and Sharpee, T. O. (2009). "Predictable irregularities in retinal receptive fields". *Proc. Natl. Acad. Sci. U. S. A.* 106.38, 16499–16504 (cit. on p. 82).

Livingstone, M. S. and Hubel, D. H. (1988). "Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception". *Science* 240.4853, 740–749 (cit. on p. 8).

Lochmann, T., Blanche, T. J., and Butts, D. A. (2013). "Construction of Direction Selectivity through Local Energy Computations in Primary Visual Cortex". *PLoS One* 8.3, e58666 (cit. on p. 88).

Maheswaranathan, N., Kastner, D. B., Baccus, S. A., and Ganguli, S. (2018). "Inferring hidden structure in multilayered neural circuits". *PLOS Comput. Biol.* 14.8, e1006291 (cit. on pp. 2, 5, 10, 84).

Maheswaranathan, N., McIntosh, L. T., Tanaka, H., Grant, S., Kastner, D. B., Melander, J. B., Nayebi, A., Brezovec, L., Wang, J., Ganguli, S., and Baccus, S. A. (2019). "The dynamic neural code of the retina for natural scenes". *bioRxiv*, 340943v5 (cit. on p. 10).

Manookin, M. B., Beaudoin, D. L., Ernst, Z. R., Flagel, L. J., and Demb, J. B. (2008). "Disinhibition Combines with Excitation to Extend the Operating Range of the OFF Visual Pathway in Daylight". *J. Neurosci.* 28.16, 4136–4150 (cit. on p. 81).

Margolis, D. J. and Detwiler, P. B. (2007). "Different Mechanisms Generate Maintained Activity in ON and OFF Retinal Ganglion Cells". *J. Neurosci.* 27.22, 5994–6005 (cit. on p. 83).

Masland, R. H. (2012). "The Neuronal Organization of the Retina". *Neuron* 76.2, 266–280 (cit. on p. 2).

Masri, R. A., Percival, K. A., Koizumi, A., Martin, P. R., and Grünert, U. (2019a). "Survey of retinal ganglion cell morphology in marmoset". *J. Comp. Neurol.* 527.1, 236–258 (cit. on pp. 1, 4, 8).

Masri, R. A., Lee, S. C. S., Madigan, M. C., and Grünert, U. (2019b). "Particle-Mediated Gene Transfection and Organotypic Culture of Postmortem Human Retina". *Transl. Vis. Sci. Technol.* 8.2, 7 (cit. on p. 8).

McFarland, J. M., Cui, Y., and Butts, D. A. (2013). "Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs". *PLoS Comput. Biol.* 9.7, e1003143 (cit. on pp. 5, 88).

McIntosh, L. T., Maheswaranathan, N., Nayebi, A., Ganguli, S., and Baccus, S. A. (2016). "Deep Learning Models of the Retinal Response to Natural Scenes". In: *Adv. Neural Inf. Process. Syst.* Vol. 29. Curran Associates, Inc. (cit. on p. 10).

Meister, M., Pine, J., and Baylor, D. A. (1994). "Multi-neuronal signals from the retina: acquisition and analysis". *J. Neurosci. Methods* 51.1, 95–106 (cit. on pp. 9, 10).

Meister, M. and Berry, M. J. (1999). "The Neural Code of the Retina". *Neuron* 22.3, 435–450 (cit. on p. 4).

Merigan, W. H. and Maunsell, J. H. R. (1993). "How Parallel are the Primate Visual Pathways?" *Annu. Rev. Neurosci.* 16.1, 369–402 (cit. on p. 8).

Merwine, D. K., Amthor, F. R., and Grzywacz, N. M. (1995). "Interaction between center and surround in rabbit retinal ganglion cells". *J. Neurophysiol.* 73.4, 1547–1567 (cit. on p. 11).

Mineault, P. J., Khawaja, F. A., Butts, D. A., and Pack, C. C. (2012). "Hierarchical processing of complex motion along the primate dorsal visual pathway". *Proc. Natl. Acad. Sci. U. S. A.* 109.16 (cit. on p. 88).

Mitchell, J. F. and Leopold, D. A. (2015). "The marmoset monkey as a model for visual neuroscience". *Neurosci. Res.* 93, 20–46 (cit. on p. 8).

Monasterio, F. M. de (1978). "Properties of concentrically organized X and Y ganglion cells of macaque retina". *J. Neurophysiol.* 41.6, 1394–1417 (cit. on p. 8).

Monti, S., Tamayo, P., Mesirov, J. P., and Golub, T. R. (2003). "Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data". *Mach. Learn.* 52.1, 91–118 (cit. on p. 86).

Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. (1978). "Spatial summation in the receptive fields of simple cells in the cat's striate cortex". *J. Physiol.* 283.1, 53–77 (cit. on p. 4).

Münch, T. A., Silveira, R. A. da, Siegert, S., Viney, T. J., Awatramani, G. B., and Roska, B. (2009). "Approach sensitivity in the retina processed by a multifunctional neural circuit". *Nat. Neurosci.* 12.10, 1308–1316 (cit. on pp. 7, 11).

Nassi, J. J. and Callaway, E. M. (2009). "Parallel processing strategies of the primate visual system". *Nat. Rev. Neurosci.* 10.5, 360–372 (cit. on p. 7).

Nelson, R. (1977). "Cat cones have rod input: A comparison of the response properties of cones and horizontal cell bodies in the retina of the cat". *J. Comp. Neurol.* 172.1, 109–135 (cit. on p. 2).

Olshausen, B. A. and Field, D. J. (1996). "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". *Nature* 381.6583, 607–609 (cit. on p. 88).

Olshausen, B. A. and Field, D. J. (1997). "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vision Res.* 37.23, 3311–3325 (cit. on p. 88).

Ölveczky, B. P., Baccus, S. A., and Meister, M. (2003). "Segregation of object and background motion in the retina". *Nature* 423.6938, 401–408 (cit. on pp. 7, 11, 88).

Ölveczky, B. P., Baccus, S. A., and Meister, M. (2007). "Retinal Adaptation to Object Motion". *Neuron* 56.4, 689–700 (cit. on pp. 7, 81).

Owen, A. B. and Perry, P. O. (2009). "Bi-cross-validation of the SVD and the nonnegative matrix factorization". *Ann. Appl. Stat.* 3.2, 564–594 (cit. on p. 86).

Paatero, P. and Tapper, U. (1994). "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values". *Environmetrics* 5.2, 111–126 (cit. on p. 85).

Pachitariu, M., Steinmetz, N. A., Kadir, S., Carandini, M., and Harris, K. D. (2016). "Fast and accurate spike sorting of high-channel count probes with KiloSort". In: *Adv. Neural Inf. Process. Syst.* Vol. 29. Curran Associates, Inc. (cit. on p. 9).

Papagiakoumou, E. (2013). "Optical developments for optogenetics". *Biol. Cell* 105.10, n/a–n/a (cit. on p. 88).

Park, J. E. and Silva, A. C. (2019). "Generation of genetically engineered non-human primate models of brain function and neurological disorders". *Am. J. Primatol.* 81.2, e22931 (cit. on pp. 7, 8, 89).

Paxinos, G., Watson, C., Petrides, M., Rosa, M. G. P., and Tokuno, H. (2012). *The marmoset brain in stereotaxic coordinates*. London: Academic Press, Elsevier (cit. on p. 8).

Peng, Y.-R., Shekhar, K., Yan, W., Herrmann, D., Sappington, A., Bryman, G. S., Zyl, T. van, Do, M. T. H., Regev, A., and Sanes, J. R. (2019). "Molecular Classification and Comparative Taxonomics of Foveal and Peripheral Cells in Primate Retina". *Cell* 176.5, 1222–1237.e22 (cit. on pp. 1, 4).

Peng, Y., Long, Y., Qin, F., Kong, W., Nie, F., and Cichocki, A. (2019). "Flexible Non-negative Matrix Factorization with Adaptively Learned Graph Regularization". In: *ICASSP 2019 - 2019 IEEE Int. Conf. Acoust. Speech Signal Process.* IEEE, 3107–3111 (cit. on p. 85).

Petrusca, D., Grivich, M. I., Sher, A., Field, G. D., Gauthier, J. L., Greschner, M., Shlens, J., Chichilnisky, E. J., and Litke, A. M. (2007). "Identification and Characterization of a Y-Like Primate Retinal Ganglion Cell Type". *J. Neurosci.* 27.41, 11019–11027 (cit. on pp. 8, 82).

Poljak, S. (1935). "Structure of the retina in primates". *Acta Ophthalmol.* 13, 52–60 (cit. on p. 8).

Polyak, S. L. (1941). *The Retina*. Chicago, IL: University of Chicago Press (cit. on p. 8).

Preuss, T. M. (2007). "Evolutionary Specializations of Primate Brain Systems". In: *PRIMATE Orig. Adapt. Evol.* Boston, MA: Springer US, 625–675 (cit. on p. 7).

Protti, D. A., Di Marco, S., Huang, J. Y., Vonhoff, C. R., Nguyen, V., and Solomon, S. G. (2014). "Inner retinal inhibition shapes the receptive field of retinal ganglion cells in primate". *J. Physiol.* 592.1, 49–65 (cit. on pp. 8, 83).

Prusky, G. T. and Douglas, R. M. (2004). "Characterization of mouse cortical spatial vision". *Vision Res.* 44.28, 3411–3418 (cit. on p. 7).

Qiu, Y., Zhao, Z., Klindt, D., Kautzky, M., Szatko, K. P., Schaeffel, F., Rifai, K., Franke, K., Busse, L., and Euler, T. (2021). "Natural environment statistics in the upper and lower visual field are reflected in mouse retinal specializations". *Curr. Biol.* 31.15, 3233–3247.e6 (cit. on p. 7).

Ravi, S., Ahn, D., Greschner, M., Chichilnisky, E. J., and Field, G. D. (2018). "Pathway-Specific Asymmetries between ON and OFF Visual Signals". *J. Neurosci.* 38.45, 9728–9740 (cit. on p. 9).

Raviola, E. and Gilula, N. B. (1973). "Gap Junctions between Photoreceptor Cells in the Vertebrate Retina". *Proc. Natl. Acad. Sci. U. S. A.* 70.6, 1677–1681 (cit. on p. 2).

Real, E., Asari, H., Gollisch, T., and Meister, M. (2017). "Neural Circuit Inference from Function to Structure". *Curr. Biol.* 27.2, 189–198 (cit. on pp. 5, 10).

Reinhard, K. and Münch, T. A. (2021). "Visual properties of human retinal ganglion cells". *PLoS One* 16.2, e0246952 (cit. on p. 7).

Rhoades, C. E., Shah, N. P., Manookin, M. B., Brackbill, N., Kling, A., Goetz, G., Sher, A., Litke, A. M., and Chichilnisky, E. J. (2019). "Unusual Physiological Properties of Smooth Monostratified Ganglion Cell Types in Primate Retina". *Neuron* 103.4, 658–672.e6 (cit. on pp. 9, 87, 88).

Riesenhuber, M. and Poggio, T. (1999). "Hierarchical models of object recognition in cortex". *Nat. Neurosci.* 2.11, 1019–1025 (cit. on p. 88).

Rockhill, R. L., Daly, F. J., MacNeil, M. A., Brown, S. P., and Masland, R. H. (2002). "The Diversity of Ganglion Cells in a Mammalian Retina". *J. Neurosci.* 22.9, 3831–3843 (cit. on p. 4).

Rodieck, R. W. (1965). "Quantitative analysis of cat retinal ganglion cell response to visual stimuli". *Vision Res.* 5.12, 583–601 (cit. on p. 4).

Rodieck, R. W. (1973). *The vertebrate retina*. San Francisco: Freeman, W. H. (cit. on p. 4).

Rodieck, R. W. and Stone, J. (1965). "Analysis of receptive fields of cat retinal ganglion cells". *J. Neurophysiol.* 28.5, 833–849 (cit. on p. 4).

Rodieck, R. W., Binmoeller, K. F., and Dineen, J. (1985). "Parasol and midget ganglion cells of the human retina". *J. Comp. Neurol.* 233.1, 115–132 (cit. on p. 8).

Roy, S., Jun, N. Y., Davis, E. L., Pearson, J., and Field, G. D. (2021). "Inter-mosaic coordination of retinal receptive fields". *Nature* 592.7854, 409–413 (cit. on pp. 4, 83).

Rust, N. C., Schwartz, O., Movshon, J. A., and Simoncelli, E. P. (2005). "Spatiotemporal Elements of Macaque V1 Receptive Fields". *Neuron* 46.6, 945–956 (cit. on p. 88).

Samengo, I. and Gollisch, T. (2013). "Spike-triggered covariance: geometric proof, symmetry properties, and extension beyond Gaussian stimuli". *J. Comput. Neurosci.* 34.1, 137–161 (cit. on p. 10).

Sanes, J. R. and Masland, R. H. (2015). "The Types of Retinal Ganglion Cells: Current Status and Implications for Neuronal Classification". *Annu. Rev. Neurosci.* 38.1, 221–246 (cit. on p. 4).

Santos, A., Humayun, M. S., Juan, E. de, Greenburg, R. J., Marsh, M. J., Klock, I. B., and Milam, A. H. (1997). "Preservation of the inner retina in retinitis pigmentosa: a morphometric analysis". *Arch. Ophthalmol.* 115.4, 511–515 (cit. on p. 1).

Sasaki, E., Suemizu, H., Shimada, A., Hanazawa, K., Oiwa, R., Kamioka, M., Tomioka, I., Sotomaru, Y., Hirakawa, R., Eto, T., Shiozawa, S., Maeda, T., Ito, M., Ito, R., Kito, C., Yagihashi, C., Kawai, K., Miyoshi, H., Tanioka, Y., Tamaoki, N., Habu, S., Okano, H., and Nomura, T. (2009). "Generation of transgenic non-human primates with germline transmission". *Nature* 459.7246, 523–527 (cit. on pp. 7, 8, 89).

Schreyer, H. M. and Gollisch, T. (2021). "Nonlinear spatial integration in retinal bipolar cells shapes the encoding of artificial and natural stimuli". *Neuron* 109.10, 1692–1706.e8 (cit. on pp. 5, 83).

Schwartz, G. W. and Rieke, F. (2011). "Nonlinear spatial encoding by retinal ganglion cells: when $1 + 1 \neq 2$". *J. Gen. Physiol.* 138.3, 283–290 (cit. on p. 1).

Schwartz, G. W., Okawa, H., Dunn, F. A., Morgan, J. L., Kerschensteiner, D., Wong, R. O., and Rieke, F. (2012). "The spatial structure of a nonlinear receptive field". *Nat. Neurosci.* 15.11, 1572–1580 (cit. on pp. 5, 7, 10, 81).

Schwartz, O., Pillow, J. W., Rust, N. C., and Simoncelli, E. P. (2006). "Spike-triggered neural characterization". *J. Vis.* 6.4, 13 (cit. on p. 10).

Shah, N. P., Brackbill, N., Rhoades, C. E., Kling, A., Goetz, G., Litke, A. M., Sher, A., Simoncelli, E. P., and Chichilnisky, E. J. (2020). "Inference of nonlinear receptive field subunits with spike-triggered clustering". *Elife* 9, e45743 (cit. on pp. 2, 11, 82, 84, 88).

Sherrington, C. S. (1906). *The integrative action of the nervous system*. Vol. 35. New Haven: Yale University Press (cit. on p. 2).

Shkolnik-Yarros, E. G. (1971). "Neurons of the cat's retina". *Vision Res.* 11.1, 7–26 (cit. on p. 4).
Silveira, R. A. da and Roska, B. (2011). "Cell Types, Circuits, Computation". *Curr. Opin. Neurobiol.* 21.5, 664–671 (cit. on p. 1).
Simoncelli, E. P., Paninski, L., Pillow, J. W., and Schwartz, O. (2004). "Characterization of Neural Responses with Stochastic Stimuli". *Cogn. Neurosci.* 3, 327–338 (cit. on p. 4).
Solomon, S. G. and Rosa, M. G. P. (2014). "A simpler primate brain: the visual system of the marmoset monkey". *Front. Neural Circuits* 8, 96 (cit. on p. 8).
Song, M., Jang, J., Kim, G., and Paik, S.-B. (2021). "Projection of Orthogonal Tiling from the Retina to the Visual Cortex". *Cell Rep.* 34.1, 108581 (cit. on p. 84).
Soto, F., Hsiang, J.-C., Rajagopal, R., Piggott, K., Harocopos, G. J., Couch, S. M., Custer, P., Morgan, J. L., and Kerschensteiner, D. (2020). "Efficient Coding by Midget and Parasol Ganglion Cells in the Human Retina". *Neuron* 107.4, 656–666.e5 (cit. on p. 8).
Stevens, J. K. and Gerstein, G. L. (1976). "Spatiotemporal organization of cat lateral geniculate receptive fields". *J. Neurophysiol.* 39.2, 213–238 (cit. on p. 4).
Stone, J. L., Barlow, W. E., Humayun, M. S., Juan, E. de, and Milam, A. H. (1992). "Morphometric analysis of macular photoreceptors and ganglion cells in retinas with retinitis pigmentosa." *Arch. Ophthalmol.* 110.11, 1634–1639 (cit. on p. 1).
Szmajda, B. A., Grünert, U., and Martin, P. R. (2008). "Retinal ganglion cell inputs to the koniocellular pathway". *J. Comp. Neurol.* 510.3, 251–268 (cit. on p. 8).
Takeshita, D. and Gollisch, T. (2014). "Nonlinear Spatial Integration in the Receptive Field Surround of Retinal Ganglion Cells". *J. Neurosci.* 34.22, 7548–7561 (cit. on pp. 11, 88).
Tanaka, H., Nayebi, A., Maheswaranathan, N., McIntosh, L. T., Baccus, S. A., and Ganguli, S. (2019). "From deep learning to mechanistic understanding in neuroscience: the structure of retinal prediction". In: *Adv. Neural Inf. Process. Syst.* Vol. 32. Curran Associates, Inc. (cit. on p. 10).
Toris, C. B., Eiesland, J. L., and Miller, R. F. (1995). "Morphology of ganglion cells in the neotenous tiger salamander retina". *J. Comp. Neurol.* 352.4, 535–559 (cit. on p. 4).
Troilo, D., Rowland, H. C., and Judge, S. J. (1993). "Visual optics and retinal cone topography in the common marmoset (Callithrix jacchus)". *Vision Res.* 33.10, 1301–1310 (cit. on p. 8).
Turner, M. H. and Rieke, F. (2016). "Synaptic Rectification Controls Nonlinear Spatial Integration of Natural Visual Inputs". *Neuron* 90.6, 1257–1271 (cit. on pp. 9, 81, 82).
Vaney, D. I., He, S., Taylor, W. R., and Levick, W. R. (2001). "Direction-Selective Ganglion Cells in the Retina". In: *Motion Vis.* Ed. by J. M. Zanker and J. Zeil. Berlin, Heidelberg: Springer Berlin Heidelberg. Chap. 1, 14–57 (cit. on pp. 5, 82).
Vavasis, S. A. (2010). "On the Complexity of Nonnegative Matrix Factorization". *SIAM J. Optim.* 20.3, 1364–1377 (cit. on p. 85).
Victor, J. D., Shapley, R. M., and Knight, B. W. (1977). "Nonlinear analysis of cat retinal ganglion cells in the frequency domain." *Proc. Natl. Acad. Sci. U. S. A.* 74.7, 3068–3072 (cit. on p. 5).
Victor, J. D. and Shapley, R. M. (1979a). "Receptive field mechanisms of cat X and Y retinal ganglion cells". *J. Gen. Physiol.* 74.2, 275–298 (cit. on p. 7).

Victor, J. D. and Shapley, R. M. (1979b). "The nonlinear pathway of Y ganglion cells in the cat retina". *J. Gen. Physiol.* 74.6, 671–689 (cit. on pp. 5, 7).

Victor, J. D. and Shapley, R. M. (1980). "A method of nonlinear analysis in the frequency domain". *Biophys. J.* 29.3, 459–483 (cit. on p. 4).

Vintch, B., Zaharia, A. D., Movshon, J. A., and Simoncelli, E. P. (2012). "Efficient and direct estimation of a neural subunit model for sensory coding". In: *Adv. Neural Inf. Process. Syst.* Vol. 25. Curran Associates, Inc. (cit. on p. 5).

Vintch, B., Movshon, J. A., and Simoncelli, E. P. (2015). "A Convolutional Subunit Model for Neuronal Responses in Macaque V1". *J. Neurosci.* 35.44, 14829–14841 (cit. on p. 88).

Vries, S. E. J. de, Baccus, S. A., and Meister, M. (2011). "The Projective Field of a Retinal Amacrine Cell". *J. Neurosci.* 31.23, 8595–8604 (cit. on pp. 9, 10, 83).

Wang, F., Li, E., De, L., Wu, Q., and Zhang, Y. (2021). "OFF-transient alpha RGCs mediate looming triggered innate defensive response". *Curr. Biol.* 31.11, 2263–2273.e3 (cit. on p. 7).

Wang, J., Jacoby, R., and Wu, S. M. (2016). "Physiological and morphological characterization of ganglion cells in the salamander retina". *Vision Res.* 119, 60–72 (cit. on p. 4).

Wässle, H. (2004). "Parallel processing in the mammalian retina". *Nat. Rev. Neurosci.* 5.10, 747–757 (cit. on p. 2).

Wässle, H., Grünert, U., Martin, P. R., and Boycott, B. B. (1994). "Immunocytochemical characterization and spatial distribution of midget bipolar cells in the macaque monkey retina". *Vision Res.* 34.5, 561–579 (cit. on p. 87).

Watanabe, M. and Rodieck, R. W. (1989). "Parasol and midget ganglion cells of the primate retina". *J. Comp. Neurol.* 289.3, 434–454 (cit. on p. 8).

Wei, W. (2018). "Neural Mechanisms of Motion Processing in the Mammalian Retina". *Annu. Rev. Vis. Sci.* 4.1, 165–192 (cit. on p. 11).

Werblin, F. S. (2010). "Six different roles for crossover inhibition in the retina: Correcting the nonlinearities of synaptic transmission". *Vis. Neurosci.* 27.1-2, 1–8 (cit. on p. 81).

Werblin, F. S. and Dowling, J. E. (1969). "Organization of the retina of the mudpuppy, Necturus maculosus. II. Intracellular recording." *J. Neurophysiol.* 32.3, 339–355 (cit. on p. 5).

White, A. J. R., Sun, H., Swanson, W. H., and Lee, B. B. (2002). "An Examination of Physiological Mechanisms Underlying the Frequency-Doubling Illusion". *J. Vis.* 43.11, 3590–3599 (cit. on p. 8).

Wu, A., Park, I. M., and Pillow, J. W. (2015). "Convolutional spike-triggered covariance analysis for neural subunit models". In: *Adv. Neural Inf. Process. Syst.* Vol. 28. Curran Associates, Inc. (cit. on p. 88).

Wyk, M. van, Pielecka-Fortuna, J., Löwel, S., and Kleinlogel, S. (2015). "Restoring the ON Switch in Blind Retinas: Opto-mGluR6, a Next-Generation, Cell-Tailored Optogenetic Tool". *PLOS Biol.* 13.5, e1002143 (cit. on pp. 1, 89).

Yan, W., Laboulaye, M. A., Tran, N. M., Whitney, I. E., Benhar, I., and Sanes, J. R. (2020a). "Mouse Retinal Cell Atlas: Molecular Identification of over Sixty Amacrine Cell Types". *J. Neurosci.* 40.27, 5177–5195 (cit. on pp. 4, 87).

Yan, W., Peng, Y.-R., Zyl, T. van, Regev, A., Shekhar, K., Juric, D., and Sanes, J. R. (2020b). "Cell Atlas of The Human Fovea and Peripheral Retina". *Sci. Rep.* 10.1, 9802 (cit. on pp. 1, 4, 8, 11).

Yger, P., Spampinato, G. L. B., Esposito, E., Lefebvre, B., Deny, S., Gardella, C., Stimberg, M., Jetter, F., Zeck, G., Picaud, S., Duebel, J., and Marre, O. (2018). "A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo". *Elife* 7, e34518 (cit. on p. 9).

Yu, Z., Turner, M. H., Baudin, J., and Rieke, F. (2022). "Adaptation in cone photoreceptors contributes to an unexpected insensitivity of primate On parasol retinal ganglion cells to spatial structure in natural images". *Elife* 11, 1–28 (cit. on p. 11).

Zaghloul, K. A., Boahen, K., and Demb, J. B. (2003). "Different Circuits for ON and OFF Retinal Ganglion Cells Cause Different Contrast Sensitivities". *J. Neurosci.* 23.7, 2645–2654 (cit. on p. 83).

Zaghloul, K. A., Manookin, M. B., Borghuis, B. G., Boahen, K., and Demb, J. B. (2007). "Functional Circuitry for Peripheral Suppression in Mammalian Y-Type Retinal Ganglion Cells". *J. Neurophysiol.* 97.6, 4327–4340 (cit. on pp. 5, 82).

Zapp, S. J., Nitsche, S., and Gollisch, T. (2022). "Retinal receptive-field substructure: scaffolding for coding and computation". *Trends Neurosci.* 45.6, 430–445 (cit. on p. 82).

Zhang, Y., Kim, I.-j., Sanes, J. R., and Meister, M. (2012). "The most numerous ganglion cell type of the mouse retina is a selective feature detector". *Proc. Natl. Acad. Sci. U. S. A.* 109.36 (cit. on pp. 7, 11).

Zhou, T., Kang, J., Cong, F., and Li, D. X. (2020). "Early childhood developmental functional connectivity of autistic brains with non-negative matrix factorization". *NeuroImage Clin.* 26, 102251 (cit. on p. 86).

ACRONYMS

A-HALS	Accelerated hierarchical alternating least squares
AF-HALS	Accelerated fast hierarchical alternating least squares
ALS	Alternating least squares
CNN	Convolutional neural network
CPCC	Cophenetic correlation coefficient
Fast HALS	Fast hierarchical alternating least squares
HALS	Hierarchical alternating least squares
IMCE	Inter-mosaic coordination energy
LGN	Lateral geniculate nucleus
LN	Linear-nonlinear
LNLN	Linear-nonlinear-linear-nonlinear
MEA	Multi-electrode array
NMF	Non-negative matrix factorization
NNLS	Non-negative least squares
NNSVD-LRC	Non-negative singular value decomposition with low-rank correction
PCA	Principal component analysis
PSTH	Peristimulus time histogram
RGC	Retinal ganglion cell
ROI	Region of interest
semi-NMF	Semi-non-negative matrix factorization
STA	Spike-triggered average
STC	Spike-triggered covariance
STE	Spike-triggered stimulus ensemble
STNMF	Spike-triggered non-negative matrix factorization
SVD	Singular value decomposition

ACKNOWLEDGMENTS

Throughout this project I had a lot of support from many people around me who I am truly grateful for. First and foremost, I extend my sincere gratitude to Tim Gollisch for his support from start to finish through this PhD. I am grateful for the enthusiasm, the encouragement, and the understanding. I have learned a lot about scientific work and how to manage a project. I valued the creative freedom and I enjoyed our discussions. I cherish the past years in the lab, and I thank him for arranging the *twice-in-a-life-time* marmoset experiments. I would like to thank the members of my thesis committee, Jan Clemens and Hansjörg Scherberger, for their questions and constructive feedback during our meetings, their quick correspondence, and their flexibility for online meetings. I would also like to thank Alexander Ecker, Viola Priesemann, and Caspar Schwiedrzik, for being part of the examination board.

I am very grateful to the members of the sensory processing in the retina lab for the adventures and great memories. First of all, I am thankful for everyone that participated in the marmoset experiments! I want to thank Helene Schreyer and Mohammad Khani that have shaped the lab with their knowledge, energy, selfless work, and warm personalities. They have been very helpful in and outside of the lab and taught me a lot. I would like to thank Dimokratis Karamanlis for the great discussions and advice and especially for the countless hours poured into developing the spike-sorting pipeline. I appreciate how much you three have sacrificed for the lab! In no particular order I thank Fernando Rozenblit and Yunus Can Erol for the technical discussions and their work on the data conversion and phy-plugins, Varsha Ramakrishna and Shashwat Sridhar for the late night discussions in the lab, Steffen Nitsche for the combined effort on the review manuscript, Juan Diego Prieto Ramirez, Margaret Young, Michael Weick and Neda Shahidi for the fun and scientific discussions. I also want to thank Jian Liu for his initial work on STNMF and Dario Protti for teaching the lab the way of the marmoset!

I could not have finished this project without Christiane Westermann and Daniela Proto who kept the lab and office running. I would also like to thank the GGNB office for the ongoing support with administrative matters and for being kind and understanding. The workplace was always clean and lively, thanks to the UMG cleaning staff who contributed immensely to the fruitful working environment.

I want to thank my fiancée Hyojin Kim who went on this journey with me together and for being patient, supportive and encouraging, and for cooking so well during the last weeks of writing this thesis. I would also like to thank her parents, Sangho Kim and Heekyung Kang, and her sister, Hyorim Kim, for the kind encouragement and exciting times together.

I am very grateful to my parents Winfried and Christiane Zapp for encouraging and supporting me throughout this project, especially during the corona quarantine and in the last few weeks. I would also like to thank my siblings Julia Weaver, Mareike Glier, and Hendrik Zapp and their significant others for their encouragement and for rooting for me during the last few weeks especially.

DECLARATION

Herewith I declare, that I prepared the Doctoral thesis "Functional Decomposition of Retinal Ganglion Cell Receptive Fields" on my own and with no other sources and aids than quoted.

Sören Johannes Zapp

Göttingen, December 14, 2022