INVESTIGATION OF THE *DROSOPHILA* MOTION VISION CIRCUIT USING REMOTE-FOCUSING TWO-PHOTON MICROSCOPY
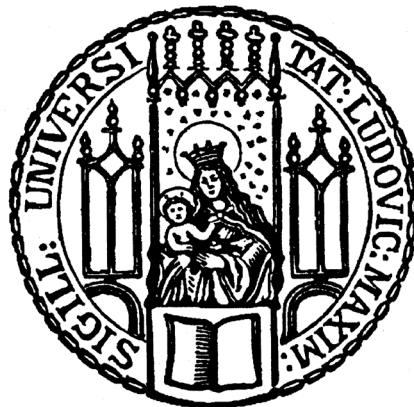
Anna Schützenberger

Dissertation der
Graduate School of Systemic Neurosciences der
Ludwig-Maximilians-Universität München

24 June 2019

Investigation of the *Drosophila* motion vision circuit using remote-focusing two-photon microscopy

Anna Schützenberger

Dissertation der
Graduate School of Systemic Neurosciences der
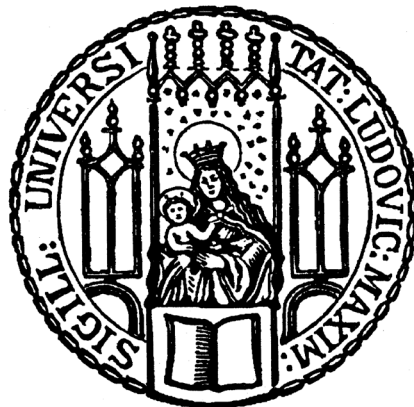Ludwig-Maximilians-Universität München

24 June 2019

Summary

Neural circuits in the brain perform countless computations at every moment. Understanding the principles that govern neural computations is a central goal in neuroscience. In the visual system, neural circuits process the information we receive through our eyes, extracting features that are relevant for us in terms of our behavior. One such processing step is the computation of visual motion. The sensory neurons in the eye, the photoreceptors, represent the visual world as a two-dimensional map of luminance levels. In this representation, motion is only contained implicitly as a spatiotemporal correlation; simply speaking, when something moves, you will see it at one position at a certain time, and then at a neighboring position shortly after. To gain an explicit representation of motion, subsequent layers of neural circuits have to compute motion. Seeing motion is such a fundamental aspect of vision that most animals' brains are equipped with circuits to perform this computation. The fruit fly *Drosophila melanogaster* is one example, and it is the species that I chose to study in this thesis. For almost a century, the fruit fly has served as a model organism to biology, making it one of the best-understood animals. Throughout this time, a key limiting factor of scientific progress has been – and still is – the availability of tools to access and probe neural circuits. Before the invention of electrophysiology in the 1950s, it was not possible to record the electrical signals of neurons, and only since the advent of two-photon microscopy in the 1990s, the signals of the small neurons in the fruit fly brain can be recorded.

Thus, in my thesis, I first expanded our toolset for microscopy, and then applied it to the study of visual motion computation in the *Drosophila* brain.

For the first part of my PhD, I developed a technique to image arbitrarily rotated, oblique planes that we can align with the retinotopic plane of the fly visual system (**Manuscript 1**). In standard two-photon microscopes, the laser focus is scanned across the focal plane by two mirrors that tilt the beam. Using the remote focusing principle, I developed a module based on a light mirror and a second objective lens, which enables us to rapidly move the laser focus along the third dimension, orthogonal to the focal plane. This allowed us independent control over all three scanning axes, making it possible to freely rotate the imaging plane in three dimensions. With this technique, we can for the first time visualize the two-dimensional maps in the fly visual system, or in other words, reconstitute the fly's view of the surroundings through the filter of a specific processing step.

In subsequent work, I used this method to study the two-dimensional representation of natural images by the interneuron L2 (**Manuscript 1**). L2 receives direct input from the photoreceptors, depolarizes to light decrements (OFF stimuli), and sends its signals to other neurons in the motion vision circuit. Consistently, I could show that L2 codes for the

inverse luminance of natural scenes, and that it represents natural images as predicted from a linear response model.

The extensive body of literature on *Drosophila* motion vision has led to a detailed understanding of the circuit. T4 and T5 are the first cell types to explicitly represent motion; that is, they preferentially respond to motion in a particular direction, and weakly to motion in any other direction. Algorithmic models of motion detection predict T4 and T5 responses to a variety of visual stimuli; however, they fail when presented with natural images. Natural scenes can vary substantially in their statistics, and a large body of evidence suggests that visual neurons dynamically adapt to the statistics of the surround. Thus, in a collaborative effort, we set out to study the effect of contrast variability on the motion detection circuit (**Manuscript 2**). We found that several cell types presynaptic to T4 and T5 adapt their responses depending on surround contrast, which renders T4 and T5 cells, as well as motion-based fly behavior, robust against contrast fluctuations.

T4 as well as T5 cells are arranged in a retinotopic fashion, each cell coding for local motion at a specific location in visual space. The computation of local motion takes place on individual dendrites of T4/T5, where input from a number of presynaptic cells is integrated to produce a direction selective signal. However, recent electron-microscopy studies revealed that these local motion detectors are also synaptically connected. Therefore, in a third project, I investigated the effect of surround motion on the local motion computation of T4 and T5 (**Manuscript 3**). I showed that T4 and T5 responses are more direction selective when presented with global motion than local motion. Moreover, I showed that synaptic connections between T4/T5 indeed shape local motion computation, revealing a previously disregarded complexity of the circuit.

Overall, with this thesis, I have advanced the technologies for studying the fly visual system, as well as contributed to our understanding of neuronal computation in the model circuit of fly motion vision.

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

1.1 STUDYING NEURAL CIRCUITS

1.1.1 NEURAL COMPUTATIONS AND INFORMATION PROCESSING IN THE BRAIN

Neurons are the fundamental building blocks of brains in all animals. They receive and pass on electrical signals. A neural circuit is made up by a group of neurons that are connected in a specific way. It acts as a signal processing unit, computing quantities that are relevant for the animal. All the neural circuits taken together form a vast network that is the nervous system.

The brain is the control center of an animal's behavior. As such, it is involved in all tasks the animal faces, whether it is evading predators, locating food or finding a mating partner. To be able to regulate these behaviors, the brain needs to have access to different types of information. First of all, information about the environment, which it receives via sensory systems. With our eyes, we take in visual information through the light reaching our retina. Similarly, we can hear and smell, and thus gather sound and odor information about our environment. Through the autonomous nervous system, the brain also obtains information about our body's internal state, e.g. whether we are hungry, cold, or need to go to the bathroom. The information about the position of our body is gathered through our proprioceptive sense (this means that you can close your eyes and still know whether your arm is bent or stretched out). Furthermore, information is stored in the brain itself in the form of memories and experiences we have made throughout our life.

This abundance of information that the brain receives at every moment is processed by neural circuits and converted it into a representation that is relevant for the animal. The processing steps that are performed by neural circuits are illustrated by the following example: Many sound waves reach our ears at once, but only a complex combination of frequencies and temporal patterns of the sound will be relevant to us, such as those that correspond to human speech. Neural circuits in the brain need to process the frequencies that reach the ear to identify the relevant aspects. By comparing the information from both ears, neural circuits can evaluate where the sounds are coming from. That the letter "a" was spoken can be determined based on a specific temporal combination of frequencies. Similar processes are in place in the visual system. The retina represents the surrounding visual space as a two-dimensional map of brightness. However, before we become aware of the information at a level that is relevant to us (for instance that we are seeing an apple), our visual system has already performed many processing steps. These encompass adapting to the overall brightness levels, detecting the edges and contours in the image, understanding

that the central red area is a separate object in the foreground, and eventually distinguishing it from a tomato or a pear in red light. Yet other neural circuits in turn will receive this information, combine it with the information from your stomach, and pass it on to your speech centers, which will make you say, "Yes, I would like the apple."

The fundamental principles of neural communication, together with the tasks we know the brain can accomplish, raise many questions. How does a neural circuit perform these computations? How does an individual neuron contribute? How can a neuron process information? How do sensory systems decide what information is important, and what can be discarded? These questions concerning the computations that take place in our brains form the framework for this thesis.

1.1.2 THE VISUAL SYSTEM AND THE VISUAL ENVIRONMENT

Sensory organs are the entry point of environmental information into the brain. Sensory cells translate physical phenomena like sound, light, or odors, into electrical signals. The signals of sensory cells then represent features such as the sound amplitude at a certain frequency, or the concentration of a certain volatile molecule, and are passed on to other neurons in the brain. In vision, the sensory cells are called photoreceptors and they represent the brightness at a specific location in space.

Throughout evolution, a large part of the animal kingdom has developed a sense of sight. Eyes have evolved independently several times. In the mammalian eye, light is focused through a single lens onto the retina, which houses many light-sensitive photoreceptors. In the compound eye of insects, many small lenses sample light from the visual space, which is then transduced by a single or few photoreceptors per lens. Despite the differences found in the structure of mammalian compared to insect eyes, the early processing stages in their visual systems show striking similarities. The first representation of the visual environment in both mammals and flies is in the form of a two-dimensional map (formed by the array of photoreceptors) of local luminances in the surround. The subsequent layers in the mammalian retina and in the visual system of insects also form visual maps. Each map, composed of many cells of a specific cell type, is extracting a simple visual feature such as local contrast, local motion or color, by combining and processing the information from the photoreceptors and other maps. Crucially, the visual environment is similar for both mammals and insects. Thus, similar needs for visual processing may have led to this parallel evolution of neuronal circuits. The following paragraphs will elaborate on the rules governing the visual scenes around us and the ways that visual systems adapt to these rules.

The visual image that we see at any moment in time is defined by the light sources and objects in our surroundings. The statistics of this image are not random – instead, we are more likely to encounter certain patterns of light than others. The statistics of natural scenes themselves have been subject to detailed investigation (Simoncelli and Olshausen, 2001), and they can be described on several timescales and levels. For example, as a general rule

natural scenes contain spatial correlations, or, in other words, it is likely for a neighboring area in an image to have a similar brightness. Conversely, other image statistics can change abruptly; stepping out of the door into a sunny day will shift the mean luminance levels drastically.

Accordingly, visual systems have adapted in the course of evolution to the general principles governing natural scenes. In addition, they are dynamically adapting to a change in the statistics of the current visual environment. For instance, moths that are active during the night only encounter low luminances, and consequently, their visual systems have adapted throughout evolution to be able to see in low-light conditions (Stöckl et al., 2016). Despite these optimizations in the course of evolution, short-term adaptation of the visual system is still necessary. The dynamic range of a neuron (the range across which a neuron can transmit information with changes in its membrane potential) usually only encompasses one or two orders of magnitude, while the changes in luminance, from a bright day to a moonless night, can span more than ten orders of magnitude (Rieke and Rudd, 2009). Mechanical systems as the changing pupil size in humans can in part compensate for this change. Moreover, neuronal mechanisms shift the dynamic range of neurons depending on the surround statistics (Carandini and Heeger, 2012). As stated by an influential theory on efficient coding, for a visual system with a limited dynamic range, this adaptation to the statistics of the surround is critical for efficiently representing visual information.

A statistic of natural scenes that can vary considerably over space and time is contrast (Ruderman and Bialek, 1994). The contrast of a natural image is often defined for a local patch of pixels as the standard deviation of luminances in this area, also called root-mean-square contrast, and it is essentially a measure for the width of the luminance distribution (Peli, 1990). In a forest, contrasts may be low. Contrasts in the sky areas of an image are generally low, but during mid-day sun, contrasts are very stark. Two examples of contrast and luminance distributions in natural images are shown in Figure 1. However, to what extend, on what timescale and in which neurons contrast adaptations take place in the fly is still an open question. Therefore, how neurons dynamically adapt to changes in surround contrast statistics forms a part of this thesis.
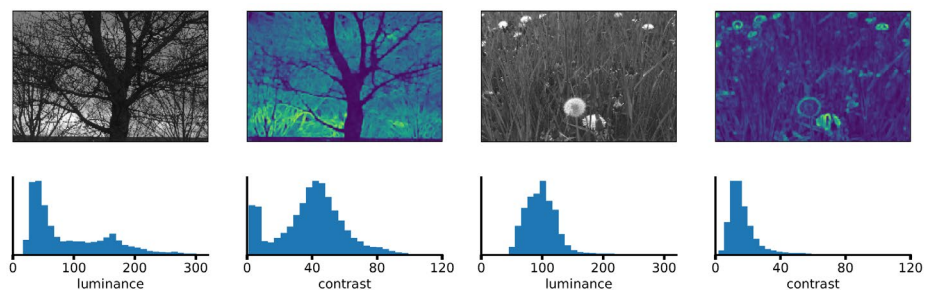

Figure 1 | Luminance and contrast distribution of natural images. The image on the left contains bright as well as dark areas, and stark contrasts in the entire image except for the tree stem. The image on the left is contains only low contrasts and luminances, except for the flower areas, where contrast and luminance are high.

1.1.3 TOOLS AND APPROACHES TO DISSECT A NEURAL CIRCUIT

1.1.3.1 TOOLS

Scientific progress is frequently determined by the availability of tools and instruments. If not for the invention of microscopy, we would still be oblivious to everything too small to see with our bare eyes. If not for our understanding of genes and inheritance, we could not have used it to build genetic tools for dissecting neural circuits. Today, a countless variety of tools is available to study the brain, ranging from advanced microscopy techniques to intricate molecular approaches and computational methods. In section 1.3 and 1.4, the tools used in this thesis will be introduced in detail. However, still, the types of questions about the brain that we can ask - and answer - are evidently limited by the tools we have.

For example, it is fundamentally important to know which neurons connect to each other. It allows us to formulate hypotheses about functional units in the brain, about the type of information processing taking place in neural circuits, and about general principles governing neural circuit architecture. However, until very recently it was impossible to comprehensively map the synaptic connections between neurons. Apart from the small nematode worm *C. elegans*, where all 5600 connections between its 302 neurons were mapped in 1986 (White et al., 1986), we do not have a full connectome, i.e. the complete map of neuronal connections, for any other species. Nevertheless, rapid development of electron microscopy technologies made it possible to reconstruct the connectomes of chemical synapses for several smaller brain regions in a number of organisms (Ohno et al., 2016; Swanson and Lichtman, 2016).

Similarly, the advent of electrophysiology in the 1950s, and subsequently two-photon calcium imaging in the 1990s, made it possible for the first time to probe the electrical signals of neurons. In electrophysiology, by inserting an electrode into the brain or into individual cells, the electrical potential of the cells can be measured. This method enabled the detailed characterization of neuronal signaling, which is based on different ionic currents flowing across the membrane. While extremely powerful in its own right, only very few cells can be recorded from at once, and only cells with large cell bodies. The calcium concentration in neurons is tightly linked to their electrical potential, and thus is a proxy for neural activity. With two-photon imaging, calcium concentration can be measured in axons and dendrites, as well as in neurons that were previously inaccessible to electrophysiology. Today, two-photon imaging is used extensively across neuroscience, enabling the detailed study of functional properties of neurons.

Seeing that the right tools are crucial for scientific progress, as part of this thesis, I developed a novel module for two-photon microscopes that allows us ask new types of research questions.

1.1.3.2 APPROACHES FOR SYSTEMS IDENTIFICATION

As an animal moves through the environment, its brain is in a dynamic state, and neuronal activity is constantly changing. As the input through sensory systems changes, the animal will change its behavior, which consequently causes changes in sensory input once more. To investigate a complex, closed-loop and dynamic system such as this, it is very common in biology to take a reductionist approach. In doing so, individual components of the system are studied in isolation, assuming that their properties in isolation also reflect their role in the bigger system. For example, animals can be immobilized while being shown visual stimuli, as if watching a movie. This way, the visual input can be controlled precisely, and the feedback loop from behavior to the sensory input is interrupted. The properties of the neuronal responses can then be set into direct relationship with the visual stimuli that are being shown. These stimuli, rather than being naturalistic visual movies, can reflect very simple properties of visual features. Stimuli such as flashes, moving edges, moving gratings, or noise stimuli are very common in vision research as they allow for identification of neuronal properties based on simple stimulus features. With this powerful approach, detailed functional properties of many cell types in the visual system have been identified. It will also be used throughout this thesis.

It is equally important, however, to challenge this assumption, and ask whether the properties of the circuit are different when part of the bigger system rather than in isolation. In order to eventually understand the workings of the system in the natural context, it is necessary to investigate its properties also on an increasingly systemic level. To this end, for example, visual system responses have been studied in behaving animals, behavioral responses to visual stimuli have been studied in closed-loop, and naturalistic visual stimuli have been employed. In the course of this thesis, natural scenes will be used in conjunction with simple stimuli to probe visual system adaptation to surround contrast as well as the representation of natural images by a specific cell type.

1.2   THE MODEL SYSTEM: *DROSOPHILA* MOTION VISION

The motion vision circuit in the fruit fly serves as an excellent study object for investigating neural computation for a number of reasons, which will be elaborated upon in the following sections.

1.2.1   THE FRUIT FLY

The fruit fly *Drosophila melanogaster* (in this thesis also referred to as *Drosophila*, or simply, the fly) has been subject to scientific investigation for more than a century. In the early 20th century, Morgan picked the fruit fly as the model organism to study the rules of heredity. He discovered many basic principles without at the time knowing about the underlying structure of the DNA (Bellen et al., 2010). In 1910, he identified the *white* gene, a gene that is responsible for the red eye color of *Drosophila* (Morgan, 1910). Ever since, *Drosophila* has

been used heavily for research in many fields, for instance genetics, developmental biology and neuroscience (Bellen et al., 2010).

The advantages of *Drosophila* as a model organism are numerous: Its generation time is around 10 days, which makes generating new fly lines simple and quick. Fruit flies are cheap and easy to keep and maintain. A female lays several thousand eggs in her lifetime. Importantly, the nervous system of the fly is largely hard-wired, thus excluding factors such as plasticity. It comprises only about 300 000 neurons (Hofbauer and Campos-Ortega, 1990), and yet the fly shows a rich behavioral repertoire, exhibiting, for example, elaborate courtship behavior and fast flight maneuvers. Finally, the decade-long use as a model organism makes it an exceptionally well-understood animal, and has lead to the development of a unique set of genetic tools that can be used to modify its neuronal circuitry with unprecedented precision (Venken et al., 2011).

1.2.2 MOTION VISION

Being able to perceive visual motion is crucial for many of the tasks any animal faces. Objects as for example prey, a predator or conspecifics, may be hard to see when stationary, but can be identified easily when moving, based on the relative motion of the object to the background. The visual motion produced by the movement of the animal itself is used for orientation and stabilization, and is an indication for the distance of objects. In insects, a multitude of reflexes and behaviors are based on visual motion. For example, male fruit flies chase female flies during courtship, always keeping the female in front of them (Land and Collett, 1974). Hunting insects, like some robber flies or dragonflies, can keep their prey, which itself is moving, precisely in front of them during their approach (Olberg et al., 2007; Wardill et al., 2017). In flight, fruit flies and other insects are able keep a stable course, which is accomplished by stabilizing the visual environment around them.

When the entire visual image moves across the retina of an animal, as if the animal was either turning by itself or being moved by environmental conditions like wind, the animal
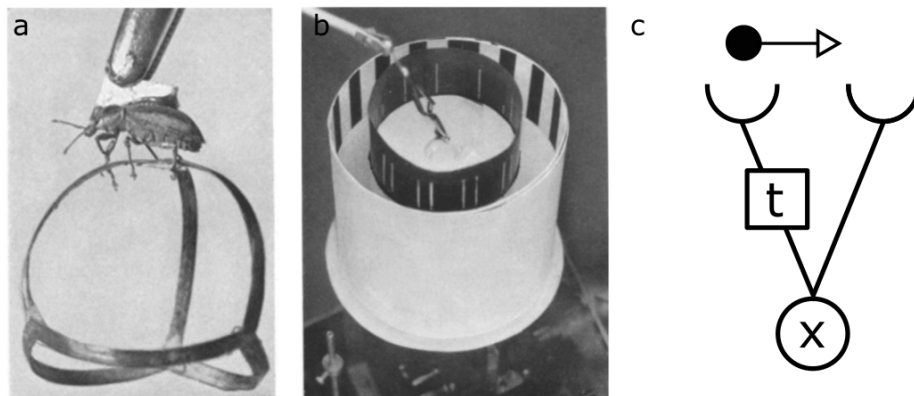

Figure 2 | Optomotor experiments in the beetle. a) The beetle *Chlorophanus* on a spherical maze with decision points. b) Visual stimulus area around the beetle. c) Reichardt detector model characterizing the beetles walking behavior. a) and b) taken with permission from Hassenstein and Reichardt (1956).

can counteract this motion by turning in the opposite direction. This behavior has been described in many species, including humans and flies, and is called optomotor response (e.g.: Abdeljalil et al., 2005; Fermi and Richardt, 1963; Goetz, 1964; Rock and Smith, 1986). Seminal studies on the optomotor reflex in a beetle were carried out by Reichardt and colleagues in the 1950's (Hassenstein, 1951; Hassenstein and Reichardt, 1956). In these experiments, the beetle *Chlorophanus viridis* walked on a spherical maze where it repeatedly had to make a choice whether to continue walking to the right or left (Fig. 2a). The beetle was positioned inside a rotating drum, on which visual stimuli were shown. When a pattern was rotated in a certain direction, the beetle walked in the same direction, attempting to stabilize the visual image on its retina.

The detailed characterization of this optomotor behavior led to the development of an algorithmic model for motion detection that could predict the beetle's behavior (Fig. 2b, Hassenstein and Reichardt, 1956). In this model, two local luminance signals from adjacent points in space are compared. The two signals are multiplied with each other (x), after one signal was delayed in time (t). This results in a large signal when an object moves in one direction, but a small one when an object moves the other direction. Numerous details of the beetle's behavior could be explained with this model (Mauss et al., 2017b; Reichardt, 2013, 1987).

The Hassenstein-Reichardt model can explain motion detection on the algorithmic level; however, it can only hypothesize about the neural implementation of motion detection. This motivate the central question: How do neurons in the brain compute visual motion? At the level of the retina, photoreceptors represent the local luminance of the surround. In this representation, motion information is contained only implicitly, i.e. in the spatio-temporal correlations of luminance. In order to explicitly represent visual motion, neurons therefore have to extract motion information from the retinal signals.

Direction selective cells have been found for the first time in frog retina (Maturana and Lettvin, 1960). These cells explicitly represent motion in that they preferentially respond to visual motion in a particular direction. Subsequent studies in the visual system of the blow fly *Calliphora vicina* also found motion selective cells (Dvorak et al., 1975a, 1975b), which have since been characterized in detail (e.g. Borst and Haag, 2002; Hausen, 1982). The neural substrates encoding visual motion vision have subsequently been studied extensively in the mammalian retina and cortex and in many other animals, including the fruit fly *Drosophila* (Borst and Helmstaedter, 2015). How then do these neurons become direction selective?

In summary, the computation of visual motion is a prime example for a neural computation that can be described mathematically, can be studied in a tractable model organism and is relevant behaviorally for many animals. The section 2.5. on '*Drosophila* motion vision' in this

introduction will give a detailed account on what is already known about the computation of motion in *Drosophila*.

1.3   TOOLS TO STUDY THE *DROSOPHILA* VISUAL SYSTEM

Over the last century of research on *Drosophila*, a multitude of tools has been developed, in general for biological science and specifically for *Drosophila*. In the following, tools that have played a key role in advancing *Drosophila* research, and tools that I made use of throughout this thesis, are introduced.

1.3.1   INVESTIGATING BRAIN STRUCTURE

To gain insight into the workings of the brain, its structures – cell types, brain areas and synaptic connections - need to be identified. Already in 1873, Golgi developed a staining method whereby a random fraction of cells in the brain is stained. With this tool and simple light microscopy, Ramon y Cajal and Sanches (1915) characterized the cell types in the visual system of several fly species. Fischbach and Dittrich (1989) used this method comprehensively to describe the majority of cell types in the visual system of *Drosophila*. They also speculated about possible circuits and connections, based on the arborizations of neurons in different layers, and already proposed a potential circuit for motion detection.

More recently, advances in electron microscopy made it possible to image brain tissue with a resolution that can identify individual chemical synapses. From an extremely detailed image stack, the connectome, the matrix of all chemical connections between neurons, can be reconstructed. Recent studies have published parts of the connectome of the *Drosophila* visual system (Shinomiya et al., 2019, 2014; Takemura, 2014; Takemura et al., 2008, 2017; Zheng et al., 2018). In the motion vision circuit, the cells presynaptic to the first motion selective cells could be identified (Shinomiya et al., 2019, 2014; Takemura et al., 2017, 2015).

Knowing which neurons form synapses with each other is an invaluable resource to form hypotheses about the functional properties of the cells and circuits. In fact, most of the hypotheses tested and formed in this thesis are based on the information from the connectome. While today we are able to reconstruct the connectome of chemical synapses of small brain areas, we are still lacking large-scale methods to identify electrical synapses, or to ascertain the strength and type of a synapse.

1.3.2   BEHAVIORAL READOUTS

When the surrounding world coherently moves in one direction, fruit flies perform an optomotor reflex during walking or flight, by turning their head and body with the direction of motion, in the same way as was already shown in the beetle (Goetz, 1964). This reflex can be used as a readout for the motion that the fly perceives. By tethering a fly on a holder and allowing it to walk on an air-cushioned ball or allowing it to freely move its wings, visual stimuli can be delivered in a controlled manner, either in open loop (the stimulus motion is defined by the experimenter) or in closed loop (the stimulus is moved on-line depending

on the fly's intended motion). These behavioral readouts allow for detailed investigation of how motion is perceived by the fly. For example, flies do not only perform an optomotor response to simple motion (that is, two-point spatio-temporal correlations), but they are also susceptible to motion illusions (Bahl et al., 2015; Clark et al., 2011; Leonhardt et al., 2017; Tuthill et al., 2011) To define the role of a specific cell type for behavior, blocking tools as introduced in the next section in combination with behavioral experiments can be used. Overall, behavioral experiments can provide valuable insights into the possible implementations of motion detection in the fly brain.

1.3.3   TARGETING CELL TYPES: GAL4-UAS SYSTEM

In order to investigate the structure or function of certain cell types or individual cells, tools to access these exact cell types are necessary. Intensive research in the field of *Drosophila* genetics resulted not only in a detailed understanding of the fly genome, but also in an array of genetic tools for cell-type specific targeting .

The most prevalent of these systems is the binary expression system Gal4-UAS (Brand and Perrimon, 1993). In this system, a transcriptional activator Gal4, which is derived from yeast and not endogenous to *Drosophila*, binds to an upstream activating sequence UAS. This activates the transcription of the gene upstream to UAS. As a consequence, this then so-called effector gene is expressed only when and where the Gal4 protein is present. To create fly lines with Gal4 present only in a subset of cells, endogenous enhancer fragments are used to drive Gal4 expression. The enhancer fragments together with the Gal4 gene are inserted into specific genomic location by site-specific integration. Using this method, large libraries of fly lines, expressing Gal4 in different subsets of cells, have been created (Jenett et al., 2012; Pfeiffer et al., 2008). By crossing a Gal4 fly line that has a specific Gal4 expression pattern with a fly line containing UAS and an effector gene, the progeny will express the effector gene in the subset of cells that also express Gal4 (Fig. 3). Since its invention, this system has been improved for stability and expression strength (St Johnston, 2002), and is used extensively in *Drosophila* research.

The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* (Chalfie et al., 1994) is one of the most important effector genes. With the Gal4-UAS system, GFP can be expressed in specific cell types, enabling the study of their morphological and anatomical properties. Other effectors include tools to block cells or to record their calcium concentration, and will be discussed below.

Several intersectional expression strategies to further restrict the expression of the effector gene became available recently. Among them is the split-Gal4 system (Luan et al., 2006). Here, the Gal4 gene is split into two parts. Only in the cells where both parts are expressed they combine, allowing them bind to the UAS sequence, activating the expression of the effector gene. By combining two expression patterns with this method, only the cells that
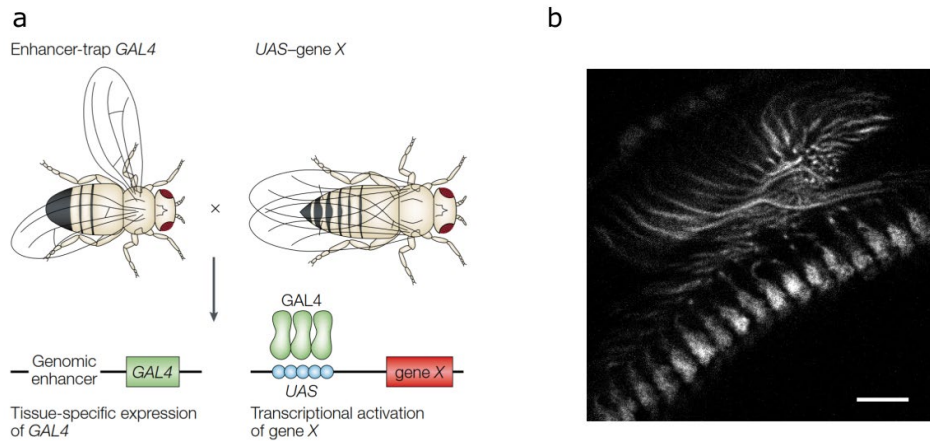
Figure 3 | The Gal4-UAS system. a) Schematic of the Gal4-UAS method. A fly expressing Gal4 in a subpopulation of cells is crossed to a fly with the UAS-effector gene construct. In the progeny, the effector gene will be expressed in the subpopulation of cells expressing Gal4. b) A fly expressing the fluorescent calcium indicator GCaMP6f in lamina monopolar cells L2 via the Gal4-UAS system. Scale bar 20 μm. a) taken with permission from St Johnston (2002).

are a part of both will express the effector gene, thus resembling the logical *and* operation. Another option to refine the expression pattern is with the Gal80 system. The presence of a protein called Gal80 inhibits Gal4 binding to UAS (Lee and Luo, 1999) and certain cells can be excluded from the expression pattern that way. Moreover, there are methods with which Gal4 expression can be controlled stochastically (e.g. Golic, 1991; Golic and Lindquist, 1989).

In addition to the Gal4-UAS system, other binary expression systems have been developed, which, combined with the Gal4-UAS system, allow for the expression of different effectors in different cell types, and increase the possibilities of circuit interventions (LexA system: (Lai and Lee, 2006) and QF system: (Potter et al., 2010)).

1.3.4 MEASURING AND MODIFYING NEURONAL ACTIVITY

1.3.4.1 ELECTROPHYSIOLOGY

Neurons encode information in their electrical potentials. Thus, recording the potential of a neuron is the most direct way to study its function. For decades, the blowfly *Calliphora vicina* was used in visual research. Its large cell bodies allowed for extra- or intracellular electrical recordings, and its visual system responses were characterized in detail (e.g. Hausen, 1984, 1976; (Borst and Haag, 2002). Only after the technique of whole-cell patch clamp recording (Hamill et al., 1981) developed further, large cells in the *Drosophila* visual system became accessible to electrical recordings (Joesch et al., 2008; Wilson et al., 2004).

Using electrophysiology, the voltage of a cell can be recorded with high temporal resolution, allowing to temporally resolve the fast dynamics of neuronal signaling. However, usually only very few cells at a time (mostly only one) can be recorded from, and many of the cells in the *Drosophila* visual system are so small that they are not accessible to electrophysiology. Therefore, other techniques as calcium imaging are often used as an alternative.

1.3.4.2 CALCIUM SENSORS

One of the most important ions in synaptic transmission is calcium, which is present in very low concentrations when a neuron is at rest. The concentration increases many-fold in the pre- and postsynaptic terminals when the cell is depolarized. In that, calcium concentration can be used as a proxy for neuronal activity. Genetically encoded fluorescent calcium sensors are proteins that emit a photon upon binding calcium. Among the most prominent are so-called GCaMP sensors which are based on the fusion of the fluorescent protein GFP and a calcium binding protein calmodulin (J. Nakai et al., 2001). Upon binding of a calcium ion by the calmodulin unit, the protein conformation changes and enables the GFP part to absorb the excitation light and in turn to emit fluorescence. Since it was first introduced, several improved versions of GCaMP have been developed (e.g. Chen et al., 2013). They can be genetically encoded in the genome of *Drosophila* as an effector gene with the Gal4-UAS system, and thus the calcium activity of specific cell types can be recorded. With this method, cell types too small for electrophysiological recordings can be investigated, which applies to large majority of cells in the *Drosophila* motion vision pathway. Due to the slow dynamics of the sensor (in the range of several hundred milliseconds), the measured fluorescence signals are slow, low-pass-filtered forms of the actual cellular calcium concentrations; a disadvantage compared to the time-resolved recordings with electrophysiology.

1.3.4.3 MODIFYING NEURONAL ACTIVITY

Analogous to calcium sensors, also other protein tools can be introduced into specific cell types with the Gal4-UAS system. Among them are proteins that block synaptic activity by interfering with different parts of cellular signaling, or proteins that increase synaptic output of the cells. These tools used in conjunction with readouts of neuronal activity as with calcium imaging, provide a powerful way to understand the roles of specific cell types in neuronal computation.

Tetanustoxin (TNT) is used to block synaptic activity (Sweeney et al., 1995). TNT cleaves to synaptobrevin, a protein that is necessary for vesicle release at the presynapse. This effectively inhibits the neuron from releasing neurotransmitter. As TNT is expressed throughout the fly's lifetime, it might, however, also interfere with the formation of neural circuits. Shibire[ts] is an alternative tool for blocking synaptic activity (Kitamoto, 2001). It is a temperature-sensitive, dominant-negative allele of dynamin, which is necessary for synaptic vesicle recycling. At room temperature, shibire[ts] is active normally, but above 29 ° it becomes ineffective. With this tool, the synaptic activity can therefore be switched off reversibly in an individual fly, and control measurements can be performed in the same animal. More recently, genetically encoded optogenetic tools have been developed (e.g. Boyden et al., 2005; Deisseroth and Hegemann, 2017; Govorunova et al., 2015; Klapoetke et al., 2014; Nagel et al., 2002). These light-sensitive ion channels in the cell membrane are opened with light. Depending on the ion channel's selectivity, light can therefore depolarize

or hyperpolarize a neuron reversibly. This light-based method is much faster than temperature-based methods in switching on and off neurons; thus it allows for on-line interaction with a neural circuit.

1.4   MICROSCOPY: IMAGING THE *DROSOPHILA* VISUAL SYSTEM

Microscopy enables us to see things so small that they would otherwise be hidden from our eyes. Lenses have been used for eyeglasses since the antiques, and telescopes reformed the worldview in the 17th century (Masters, 2008). Eventually, Antoni van Leeuwenhoek (1632–1723) was the first to observe living microorganisms with a simple microscope that he constructed using his refined method of making lenses. From then onwards, the study of the small things revolutionized our understanding of infectious disease and biology in general. Fluorescent dyes were used to increase the contrast and resolution for light microscopy. With fluorescence microscopy, many otherwise invisible structures, such as the nuclei of cells or chromosomes, could be studied *in vivo*.

In 1978, Thomas and Christoph Cremer invented the laser scanning confocal microscope, bringing together the confocal principle (Minsky, 1955) with laser scanning techniques (Cremer and Cremer, 1978). For the first time, by blocking the fluorescence emitted from all planes but the focal plane, samples could be resolved in 3D. A laser, scanning the sample, was used to excite the fluorophores. The shape of the laser beam exited a conical volume in the sample. A confocal pinhole then ensured that only the fluorescence emitted from the focal point of the laser – and not the rest of the cone – would reach the detection system. This way, the detected fluorescence signal could be localized to a small point in the 3D sample. Confocal microscopy is a technique used extensively in today's life science. Some of its disadvantages have since been overcome by a technique that emerged in the 1990s: two-photon microscopy.

1.4.1   2-PHOTON MICROSCOPY

The physical phenomenon, after which two-photon microscopy is named, has been described in theoretical work by Maria Goeppert-Mayer in 1931 in her doctoral thesis (Göppert-Mayer, 1931). According to her work, two photons can be absorbed by one molecule in one quantum event. 30 years later, after the invention of lasers, Kaiser and Garret could verify the two-photon effect experimentally (Kaiser and Garrett, 1961). The probability of two photons being absorbed at the same time is exceedingly low. Therefore, it was not until the invention of powerful, pulsed lasers with very short pulse length that the effect could be used in practice (Gosnell and Taylor, 1991). Eventually, in 1990, Denk and colleagues invented the two-photon microscope, combining laser scanning microscopy with two-photon absorption (Denk et al., 1990).
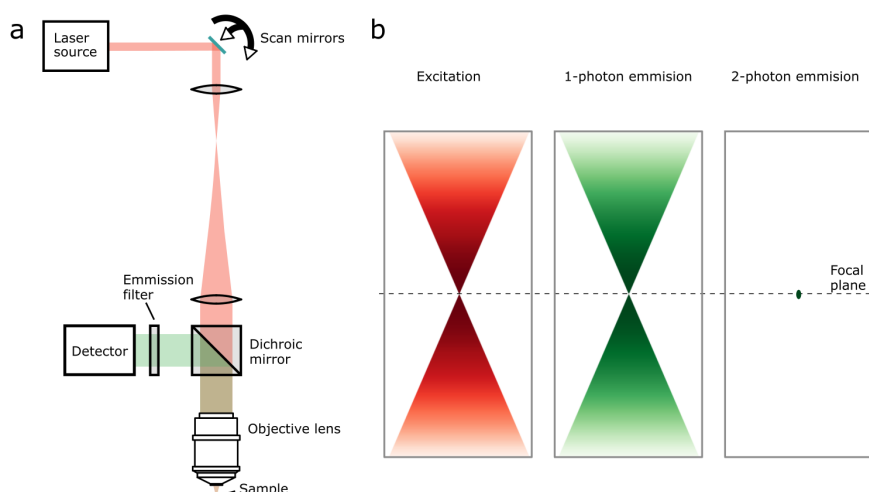
Figure 4 | Two-photon laser scanning microscopy. a) Schematic of a simple two-photon microscope. A pulsed laser is the source of excitation light. With two moving scan mirrors the beam is tilted, which results in a lateral movement of the laser focus in the focal plane of the objective lens. Resulting emitted fluorescence is collected through the same lens by separating excitation and emission wavelengths. b) Left: Schematic of the laser beam in the sample for both 1-photon and 2-photon fluorescence microscopy. Middle: Resulting fluorescence emission for 1-photon excitation. Right: Resulting fluorescence emission for 2-photon excitation. The photon density high enough for the two-photon effect to occur only in the focal spot.

In laser scanning two-photon fluorescence microscopy (Fig. 4a, for reviews, see e.g.: Helmchen and Denk, 2005; So et al., 2000; Svoboda and Yasuda, 2006; Weber and Helmchen, 2014), a pulsed laser is used to excite fluorophores in the sample. A long wavelength, approximately double the wavelength that is needed to excite the fluorophore with one photon, is used. The probability of two-photon absorption is sufficiently high only in the focal spot of the laser, so that all detected fluorescence can be localized to the small focal spot of the laser (Fig. 4b). The size of the focal spot, and thus the resolution of the system, depends on the wavelength used and the numerical aperture (NA) of the objective lens (Svoboda and Yasuda, 2006). The most commonly used lasers for two-photon microscopes are Ti:Sapphire lasers, since they provide short, < 100 ps, pulses at high repetition rates as is optimal for the excitation of fluorophores (Guild et al., 1997; Svoboda and Yasuda, 2006).

Two-photon imaging has several advantages over one photon excitation in confocal microscopy. All emitted fluorescence originates from the small focal spot of the laser. Therefore, firstly, no pinhole has to be used and so, no fluorescence signal has to be discarded. Secondly, scattered fluorescent light also contributes to the signal, since its origin is defined by the excitation (Denk et al., 1994). Thirdly, it reduces the averse effects of fluorescence, photobleaching and phototoxicity outside of the focal volume, to a minimum. Bleaching is the process by which fluorescent molecules lose their ability to cycle between ground and excited state. Phototoxicity occurs when the fluorophore in the excited state alters molecules around it, e.g. by creating reactive oxygen species that can interact with and damage tissue (Lichtman and Conchello, 2005). Finally, it is possible to image structures

deeper in the sample. The higher wavelength can penetrate tissue further as it is less absorbed by tissue and scatters less (for review, see: Helmchen and Denk, 2005).

Today, two-photon microscopes are commercially available and are used extensively in the scientific community. Combined with methods such as calcium imaging, they provide a powerful technique to assess the function of neurons (Grienberger and Konnerth, 2012; Svoboda and Yasuda, 2006). Many additional developments, for example faster scanning motors, more efficient photodetection, or resonant scanners (Fan et al., 1999; Nguyen et al., 2001) have since improved this method further.

1.4.2 IMAGING IN THE 3D VOLUME

One focus for improving two-photon microscopy lies in expanding the possibilities to image the sample in three dimensions, as opposed to imaging a 2D section, as is the case for most microscopes. Horizontal planes evidently do not reflect the complexity and 3D structure of most neuronal circuits, and in that, imaging methods that enable access to the entire sample volume are in high demand.
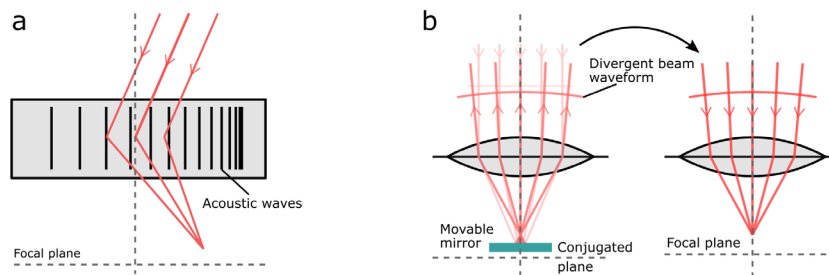
Figure 5 | Approaches for 3D imaging. a) Schematic of an acousto-optical modulator, differentially reflecting the laser beam such that it focuses outside the nominal focal plane. b) Schematic of the remote-focusing principle. A beam with non-zero vergence is created in a conjugated lens system, which then creates a focal spot outside the nominal focal plane of the imaging lens.

The two scan mirrors in a two-photon microscope tilt the laser beam, which moves the focal spot of the laser in the sample laterally in x and y. This way, the focal plane of the objective lens is scanned. In order to image a different plane, that is, to refocus, the distance between objective lens and sample has to be changed, by moving either the lens or the sample itself. For imaging *in vivo*, however, it is common to use water immersion lenses, and fast movement will introduce motion artefacts. In addition, moving heavy objects such as the objective lens or the sample cannot be achieved at high speed.

Several approaches to circumvent this limitation have been devised, that allow refocussing without moving objective lens or sample. Among the different strategies are electrically tunable lenses (Grewe et al., 2011), spatial and temporal multiplexing (Cheng et al., 2011; Durst et al., 2006) or acousto-optical lenses (e.g. Duemani Reddy et al., 2008; Katona et al., 2012; Kirkby et al., 2010). The latter are based on a principle stating that acoustic waves can change the refractive index of a medium, which in turn can modulate light waves (Fig. 5a).

Acousto-optical lenses can be used for scanning sparse, arbitrary regions of interest, as the position of the laser focus is determined by the acoustic waves, which, in contrast to the movement of mirrors, are not subject to inertia. However, microscopes using acousto-optical lenses also require precise alignment and elaborate beam control with beam stabilizers and dispersion control (Iyer et al., 2003).

Recently, an alternative refocusing approach has been described (Botcherby et al., 2008, 2007). Botcherby et al. (2008) showed that refocusing could be accomplished remotely, in the back of the microscope in a plane conjugate to the imaging plane. They applied this principle called remote focusing to two-photon imaging (Botcherby et al., 2012). It is based on a lightweight mirror and a second objective lens, which form a conjugate plane to the imaging plane (Fig. 5b). The collimated laser beam enters the objective lens and is reflected by the mirror. As it travels back through the objective lens, depending on the position of the mirror with respect to the focal plane of the lens, the waveform of the beam will be convergent or divergent. This beam is then relayed to the original objective lens, where it forms a focal spot above or below the nominal focal plane of that lens. When certain conditions are met, a remote focusing system like this can bypass spherical aberrations that are usually introduced when focusing a beam outside the focal plane of a lens that obeys the sine-condition (Botcherby et al., 2008, 2007), and in that enable diffraction-limited imaging in the entire volume.

The microscope design by Botcherby et al. (2012) implies that the remote focusing unit is placed in the beam path following the lateral scan mirrors, just before the objective lens. However, the microscope design of many standard two-photon microscopes does not allow additional optical elements in this part of the microscope. Therefore, as part of this thesis, I developed a more flexible remote focusing module that can be used in conjunction with typical two-photon microscopes. I used this module to image oblique planes, which we can align on-line with the retinotopic plane in the *Drosophila* visual system.

1.5 *DROSOPHILA* MOTION VISION

1.5.1 ANATOMY OF THE EYE

The visual system of the fruit fly *Drosophila* is comprised of around 100 000 neurons (both hemispheres). In that, it takes up about one third of the fly's brain in terms of neuron numbers. Each eye is comprised of about 750 individual units, called ommatidia (Ready et al., 1976). The eyes face almost opposite directions (panoramic vision), such that together they sample light from 85 % of the space around the animal (Heisenberg and Wolf, 1984). Each ommatidium consists of a lens, accessory cells, and eight photoreceptor cells. Following the photoreceptor layer, the subsequent structures in the visual system, the lamina, medulla and lobula complex (consisting of lobula and lobula plate), are also arranged in retinotopic columns, each column processing information from a small point in space (Fig. 6a). These structures are neuropils, made up by the neuronal processes of many
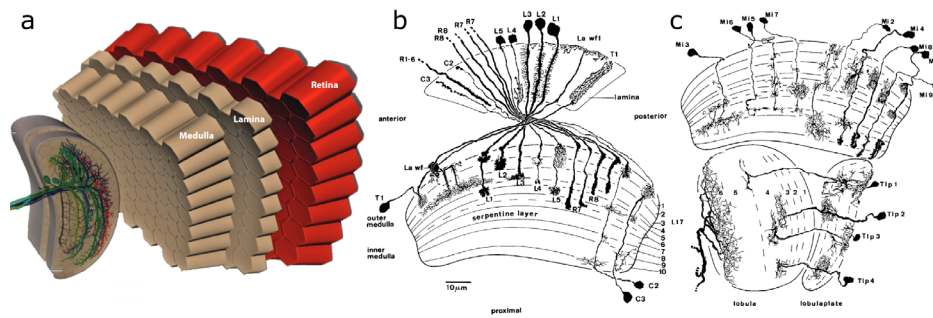
Figure 6 | Anatomy of the Drosophila visual system. a) Schematic of the fly visual system. Retinotopically arranged neuropils form the visual system. b) Cell types in the lamina and medulla as characterized by Fischbach and Dittrich (1989). c) Cell types in the medulla and lobula complex. a) is modified with permission after Borst et al. (2010). b) and c) are taken with permission from Fischbach and Dittrich (1989).

different cell types. Comprehensive anatomical studies by Fischbach and Dittrich (1989, Fig. 6b-c) described the majority of cell types in the *Drosophila* visual system. Cells in the lobula complex then project to different areas in the central brain.

1.5.2 MOTION SENSITIVE CELLS

Lobula plate tangential cells (LPTCs) are large, prominent cells that have their dendrites in the lobula plate. They were characterized in detail in the blow fly *Calliphora* (e.g. Hausen, 1982, 1976; Hengstenberg, 1982) and consequently also in *Drosophila* (Fig. 7a, Joesch et al., 2008; Schnell et al., 2010). LPTCs are motion sensitive to large areas in the visual field. They depolarize in response to one direction of motion (their preferred direction) and hyperpolarize to the opposite direction (Fig. 7b). Their dendrites are localized in the lobula plate, where they innervate specific layers. Cells in the four different layers of the lobula plate show activity to motion in four different directions, as measured with deoxyglucose labelling (Buchner et al., 1984). Cells that also stratify in specific layers of the lobula plate are T4 and T5 cells. Based on their anatomy, these cells have long been hypothesized to be sensitive to local motion and to give input to LPTCs (Fischbach and Dittrich, 1989). Of both T4 and T5, there are four different anatomical subtypes (T4a, T4b, T4c and T4d, and

analogous for T5). The axon terminals of each subtype project to a specific layer in the lobula plate, the a-subtypes to the first layer, the b-subtypes to the second and so on. Using two-photon calcium imaging, it could finally be shown that T4 and T5 cells indeed are direction selective (Fig. 8, Maisak et

Figure 7 | Lobula-plate tangential cells (LPTCs) in the Drosophila visual system. a) Staining of two LPTCs (green) and neuropil staining (purple). Scale bar 25 μm. b) Electrophysiological recordings from HS cells shows their direction selectivity; they depolarize to motion in one direction and hyperpolarize to motion in the opposite direction. a) and b) taken with permission from Schnell et al. (2010).

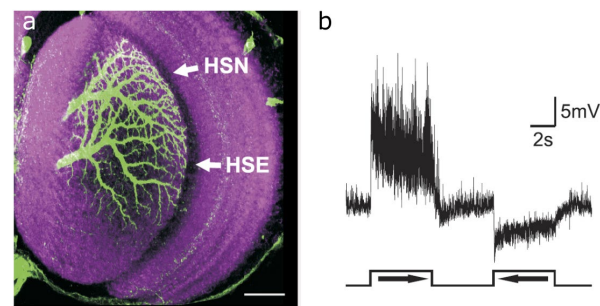al., 2013), and, as predicted from their innervations in the lobula plate, type a responds preferentially to front-to-back motion, type b to back-to-front motion, type c to upward motion and type d to downward motion. Each subtype is found once per column, and in that, T4 and T5 sample local motion from the complete visual field.
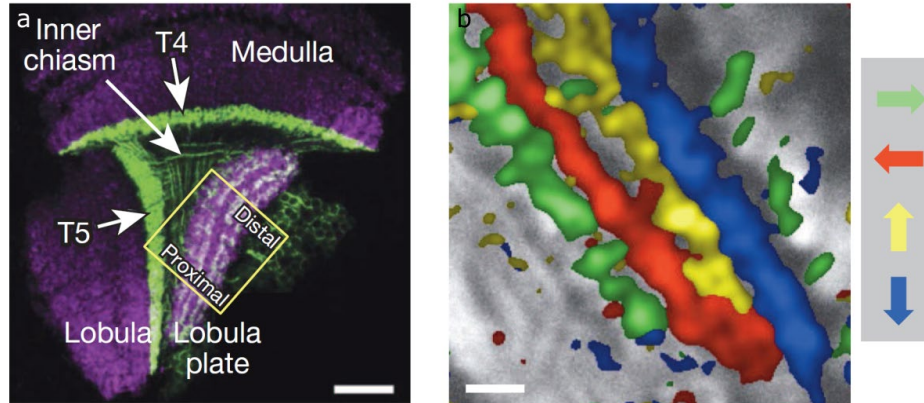
Figure 8 | Local motion sensitive T4 and T5 cells. a) Anatomy of T4 and T5 cells. In the lobula plate, axonal arborizations in four layers represent the four cardinal directions. The dendrites of T4 are localized in the medulla, while the T5 dendrites are localized in the lobula. Scale bar 20 μm. b) Calcium imaging of axonal terminals of T4 and T5 in the lobula plate reveal the layered direction selectivity of T4 and T5. Cells of subtype a, arborizing in the first layer, responds preferentially to front-to-back motion, and accordingly for the other subtypes. Scale bar 5 μm. a) and b) taken with permission from Maisak et al. (2013).

1.5.3 MODELS OF MOTION DETECTION

The behavioral studies on the beetle and other insects in the 1950s, together with early recordings from motion sensitive cells, led to the development of a group of theoretical models for motion detection before it was possible to study its neural implementation. Visual motion is manifested in the retinal signals as correlations in space-time. The most common models for motion detection therefore compare signals from at least two points in space and time. The Hassenstein-Reichardt detector (Fig. 9a), proposed by Reichardt (Hassenstein and Reichardt, 1956) to explain the beetle's turning responses, receives input from two adjacent points in space. The input from one of the points is delayed in time. Both inputs are then compared by a multiplication. In this model, a large response is elicited when motion occurs in a particular direction (the preferred direction, PD), that first elicits a signal at the point in space where the input is delayed. In this case, the signals from the two input lines arrive at the multiplication stage at the same time. A smaller response is elicited when the movement is in the opposite direction (null direction, ND). An alternative model, the Barlow-Levick detector (Fig. 9b), was proposed to explain the direction selective responses of the rabbit retinal ganglion cells (Barlow and Levick, 1965). As in the previous model, two input lines, where one of them is delayed in time, are compared. In contrast to the Hassenstein-Reichardt detector, however, the signals are then compared by a division. Here, when the signals from the two input lines arrive at the same time, the resulting signal will be small; hence, this direction of movement is the null direction. Conversely, the
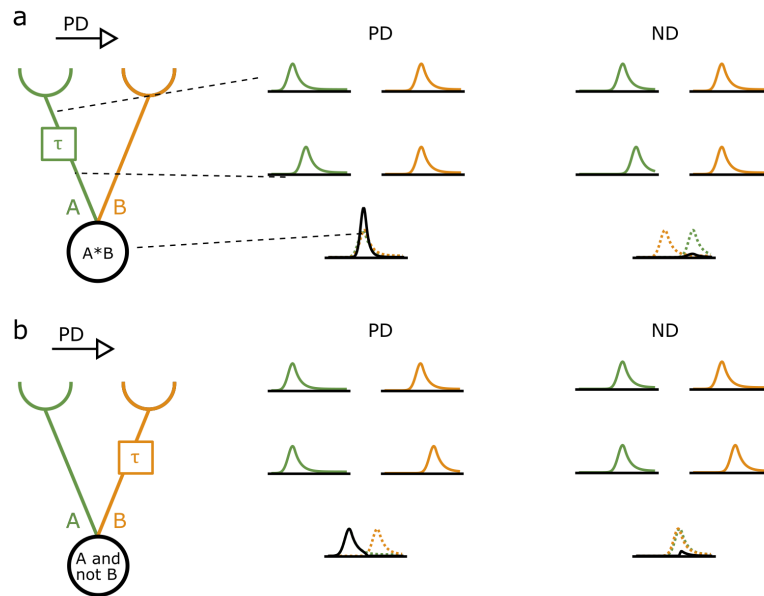
Figure 9 | Models of motion detection. a) A version of the Hassenstein-Reichardt detector, enhancing preferred-direction (PD) signals. b) A version of the Barlow-Levick detector, suppressing null-direction (ND) signals.

movement in the opposite direction, in which the divisive signal is not arriving at the same time, will elicit a larger response.

Due to their nature of facilitating the PD response or suppressing the ND response, the mechanism of the two models is also termed preferred-direction enhancement (PDE, Hassenstein-Reichardt detector) or null-direction suppression (NDS, Barlow-Levick detector). These models were influential in that they are able to explain many characteristics of motion sensitive cells as well as of the optomotor response. Recent results, as introduced below, have led to more elaborate models of motion detection, for example, a combined model with three input lines that implements both PDE and NDS (Haag et al., 2016).

1.5.4   *Drosophila* motion vision circuit

Two-photon calcium imaging together with the genetic tools that *Drosophila* provides made it possible to investigate the neural implementation of the motion detector in great detail. The following sections will introduce the circuitry presynaptic to T4 and T5 and the current efforts to pinpoint the mechanisms for motion detection.

1.5.4.1   *Photoreceptors*

There are eight photoreceptor cells per ommatidium (R1 – R8), which form specialized structures called rhabdomeres, dense stacks of microvilli. In these structures, incident light is converted into electrical potential through a biochemical cascade (Hardie and Raghu, 2001), starting with light sensitive photopigments. Photoreceptors R1 – R6 express rhodopsin 1, a photopigment that senses a broad range of wavelengths. Photoreceptors R7 and R8 express rhodopsins 3 - 6. R1 – R6 are critical for motion detection (Heisenberg and Buchner, 1977; Yamaguchi et al., 2008), although R7 and R8 have a modulatory function

(Wardill et al., 2012). The photoreceptors are arranged in the ommatidium such that they sense incident light from slightly offset positions in space. Accordingly, they project the axons to neighboring cartridges in the lamina, such that one lamina cartridge receives input from six photoreceptors in different ommatidia, that all encode information from one point in visual space (Strausfeld, 1971). This principle called neural superposition increases the sensitivity compared to a simple, direct transmission, with no trade-off in resolution (Braitenberg, 1967).

The functional properties of photoreceptors have been characterized in detail. For instance, the phototransduction cascade is remarkably fast, such that the temporal resolution of fly vision reaches up to 200 Hz (Hardie, 1991). All photoreceptors release the neurotransmitter histamine upon stimulation with light (Hardie, 1989). In addition, adaptation mechanisms within the phototransduction cascade shift the dynamic range of the photoreceptors such that it is matched to the surrounding light levels (Juusola and Hardie, 2001; Laughlin and Hardie, 1978).

1.5.4.2 *LAMINA*

Photoreceptors R1 – R6 project to the lamina neuropil, which is also arranged retinotopically. Out of the 12 cell types in the lamina (Fischbach and Dittrich, 1989; Meinertzhagen and O'Neil, 1991), eight are unicolumnar. Among them are the lamina monopolar cells L1 – 5, which all project to different layers in the medulla. Other cell types in the lamina include C2, C3, T1, and the multicolumnar cells Lai, Lawf1 and Lawf2. Electron-microscopic studies have shown that the cells form a complex network within the lamina (Rivera-Alba et al., 2011). For example, while L1 – L3 receive direct input from the photoreceptors, L4 and L5 receive input from L2, L4 forms reciprocal synapses with L2 (Rivera-Alba et al., 2011). Moreover, at least L1 and L2 are coupled with electrical synapses (Joesch et al., 2010).

L1 and L2 respond with depolarization to light decrements (OFF stimuli) (Clark et al., 2011; Reiff et al., 2010). They are necessary for motion vision, as blocking the output of both L1 and L2 renders the flies motion blind (Rister et al., 2007). Furthermore, the motion sensitive LPTCs become unresponsive when L1 and L2 output is blocked (Joesch et al., 2010). Blocking L1 and L2 individually renders the LPTCs responsive either only to ON motion stimuli (bright, moving edges) in the case of L2 block, or to OFF motion stimuli (dark moving edges) in the case of L1 block (Joesch et al., 2010). Fly behavior shows similar results (Clark et al., 2011). In that, L1 and L2 form the beginning of parallel (ON and OFF) pathways for processing motion. While L1 and L2 play major roles for motion detection, several other lamina cells have also been implicated in motion vision (Meier et al., 2014; Silies et al., 2013; Tuthill et al., 2013), which is not surprising given the intricate connectivity within the lamina. In addition, the functional response properties of L2, L3, L4 and Lawf1 have been characterized (Freifeld et al., 2013; Meier et al., 2014; Silies et al., 2013; Tuthill et al., 2014). In *Calliphora*, in electrophysiological recordings, lamina cells show center-surround receptive fields, and biphasic responses, as well as adaptation to surround

luminance (Laughlin and Hardie, 1978). In *Drosophila*, recordings with voltage indicators also show the biphasic voltage response of L1 and L2 (H. H. H. Yang et al., 2016).

1.5.4.3 MEDULLA

In addition to its columnar, retinotopic structure, the medulla is subdivided into 10 layers. Over 60 cell types are housed in this neuropil, many of them also unicolumnar (Fischbach and Dittrich, 1989). The photoreceptors R7 and R8 project their axons into the medulla, as do many lamina cell types. Electron microscope studies over the last years successively
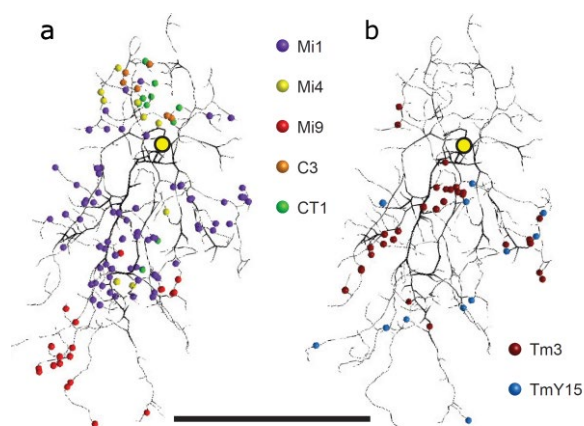
Figure 10 | Synaptic inputs to T4c from reconstructed EM data. a) Synapse distribution of several presynaptic cells on the T4 dendrite. b) Synapse distribution of other presynaptic cells on the T4 dendrite. Scale bar 10 μm. a) and b) taken with permission from Takemura et al. (2017).

revealed the complexity of the network in the medulla leading up to T4 and T5 cells (Shinomiya et al., 2019, 2014; Takemura et al., 2017, 2015). Presynaptic to T4 in the medulla are the cell types Mi1, Tm3, Mi4, Mi9, C3, CT1 and TmY15, as well as T4 cells of the same subtype (Takemura et al., 2017, Fig. 10). Presynaptic to T5 in the lobula are Tm1, Tm2, Tm4, Tm9, CT1, TmY15 as well as LT33 and TM23 (Shinomiya et al., 2019, 2014). The neurotransmitter types of these cells have also been described: Mi1 and Tm3 release acetylcholine, Mi9 uses glutamate and Mi4, TmY15, C3 and CT1 are GABAergic (Takemura et al., 2017). All cell types presynaptic to T5, except for CT1, release acetylcholine (Shinomiya et al., 2019).

Functional studies on the cell types of the medulla characterized their properties in detail. Mi1, Tm3, Mi4 and Mi9 differ with respect to their spatial and temporal response properties, exhibiting different sizes of center-surround receptive field structures, and varying temporal low-pass or band-pass characteristics (Arenz et al., 2017; Strother et al., 2017; H. H. H. Yang et al., 2016). The input cells to T5 also show a similar diversity (Arenz et al., 2017; Fisher et al., 2015a; Meier et al., 2014; Serbe et al., 2016).

1.5.4.4 NEURAL IMPLEMENTATION OF MOTION DETECTION

The motion sensitive T4 and T5 cells are required for optomotor behavior (Bahl et al., 2013; Maisak et al., 2013; Schilling and Borst, 2015) as well as for the responses of LPTCs (Bahl et al., 2015; Maisak et al., 2013; Mauss et al., 2014; Schnell et al., 2012). Since most likely none of the presynaptic cells to T4 and T5 are motion sensitive (Arenz et al., 2017; Serbe et al., 2016), direction selectivity must arise in T4 and T5 cells.

Several studies have investigated the role of the cells presynaptic to T4 and T5 for motion detection. Silencing the output of individual cell types presynaptic to T4 and T5 affects LPTC responses as well as T4 and T5 responses to different degrees (Ammer et al., 2015; Serbe et al., 2016; Strother et al., 2017). Optogenetic experiments show that activation of individual input cell types only elicits small responses in T4 (Strother et al., 2017). In addition, several studies have tried to pinpoint the motion detection algorithm that is implemented in the fly brain. According to some studies, T4 as well T5 cells show both preferred-direction enhancement and null-direction suppression at spatially offset positions (Haag et al., 2017, 2016; Leong et al., 2016). Other studies could confirm only the presence of either PDE or NDS in T4 and T5 (Fisher et al., 2015b; Gruntman et al., 2018; Salazar-Gatzimas et al., 2016). To date, the reasons for these discrepancies are not entirely clear. However, given the complexity of the synaptic connections between the presynaptic cells, further studies will be necessary to pin down the implementation of motion detection.

In the lobula plate, T4 and T5 project to LPTCs as well as to lobula plate intrinsic cells (LPis). These cells receive signals from one layer of the lobula plate and project to the layer with the opposite preferred direction, where they inhibit LPTCs, thus rendering them motion opponent (depolarizing to motion in one direction and hyperpolarizing to motion in the opposite direction; Mauss et al., 2015).

Recent electron microscopy studies show that T4 and T5 also form synapses with cells of the same subtype (Shinomiya et al., 2019; Takemura et al., 2017). Possible interactions between T4 and T5 cells have not been taken into account in previous studies on the implementation of motion detection. Therefore, part of this thesis investigates the role of interactions between T4 and T5 for motion detection.

1.6   CONCLUSION

Motion vision in the fruit fly poses an excellent example of a computation performed by a neural circuit. In studying this circuit, we may be able to gain insight into the principles of neural computation in general. A crucial limiting factor for the investigation of neural circuits is the availability of tools. In this thesis, I therefore first focused on developing a technique for imaging the retinotopic plane of the fruit fly visual system (Manuscript 1). I applied this method to study the representation of natural scenes by the cell type L2 (Manuscript 1). Furthermore, in a collaborative effort, we investigated the neural computations underlying the robustness of motion vision despite varying contrast statistics (Manuscript 2). We could show that cells presynaptic to T4 and T5 dynamically adjust to changing surround contrast, rendering T4 and T5 as well as LPTCs responses and optomotor behavior robust with respect to contrast statistics. Lastly, I investigated the influence of surround motion on the computation of local motion in T4 and T5 (Manuscript 3) and found that indeed, the local motion signals of T4 and T5 are modulated by motion in the surround.

MANUSCRIPT 1

Seeing through the eye of a fly: Reconstructing *Drosophila*'s visual map with remote focusing two-photon microscopy

Anna Schuetzenberger[1,2] and Alexander Borst[1,2]

1 Department Circuits-Computation-Models, Max-Planck-Institute of Neurobiology, Martinsried, Germany

2 Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität, Munich, Germany

ABSTRACT

In the visual system of most animals, neurons of a specific type or population form retinotopic maps. Each map processes a specific visual feature that is relevant for the animal's behavior. In the mammalian retina, the retinotopic array of neurons can be recorded from with microelectrode arrays or two-photon microscopy, while the retina is mounted to a flat surface. In the fruit fly *Drosophila*, genetic access to many cell types and advanced techniques for circuit intervention have resulted in a detailed understanding of its visual system. However, while it is possible to record calcium signals from a small number of cells, the retinotopic maps of its visual system cannot be accessed in an efficient way. Here, we present a technique that allows us to image oblique planes that we can align *on-line* with the retinotopic plane of the visual system. With a remote-focusing module, we gain fast control over the laser focus along the optical axis, and control software allows for arbitrary rotation of the imaging plane, with no trade off in resolution or scan speed. To demonstrate the usefulness of this system, we record the representation of natural images in the retinotopic map formed by the interneuron L2. We show that L2 encodes the luminance of natural images, as is well predicted by its linear response properties. In summary, we developed a method that not only permits us to record from a significant fraction of the fly's visual field but also expands the toolbox available for two-photon microscopy in other applications.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGEMENTS

INTRODUCTION

Vision is a crucial sense for many animals, including humans. Through vision, we gather rich information about our surroundings, which is processed in the visual system to extract relevant features that then guide behavioral choices. In the initial stages of visual processing, visual information is represented as a set of 2D maps of the surrounding space (Chklovskii and Koulakov, 2004; Land and Fernald, 1992). In the fruit fly *Drosophila melanogaster,* the visual system collects information through about 750 ommatidia (Ready et al., 1976), each made up of a lens and housing eight photoreceptors. In the first neuropil, the lamina, photoreceptor inputs from neighboring ommatidia converge such that one lamina column receives information from one point in visual space (neural superposition principle). Each of the 750 columns is sampling light from a different point in space (Land, 1997), and together, they retinotopically map the surround. Similarly, the cell types in the consecutive layers of neuropil, the medulla and lobula complex, constitute retinotopic maps, each of which codes for a visual feature like local contrast or motion (for review, see e.g: Mauss et al., 2017; Song and Lee, 2018).

The vertebrate retina is constructed in a strikingly similar way (for review, see e.g.: Borst and Helmstaedter, 2015; Sanes and Zipursky, 2010). Importantly, due to its anatomical separation from the rest of the brain, it can be mounted on a flat surface and the retinotopic array of neurons can be recorded from by means of microelectrode arrays or two-photon calcium imaging (Baden et al., 2016; Denk and Detwiler, 1999; Meister et al., 1994; Pillow et al., 2008; Segev et al., 2004). This ease of access probably has made the retina one of the best-understood neural circuits (Demb and Singer, 2015; Diamond, 2017; Dowling, 2012). In *Drosophila*, two-photon calcium imaging together with elaborate genetic tools (Akerboom et al., 2013; Brand and Perrimon, 1993; Pfeiffer et al., 2008) has enabled recording activity from defined cell types allowing the visual system of *Drosophila* to be studied in great detail (for reviews, see: Aptekar and Frye, 2013; Mauss et al., 2017; Song and Lee, 2018). However, there is no analogue to the technique for the vertebrate retina with which the two-dimensional spatial representation could be assessed efficiently.

Recent advances in imaging technologies are opening the doors to more flexible investigation of neural structures (for review see e.g.: Ji et al., 2016; Ronzitti et al., 2018). The retinotopic plane in the *Drosophila* visual system, when mounted for *in vivo* calcium imaging, is an arbitrarily rotated, oblique plane with respect to the imaging system (Fig. 1a-c). In this study, we developed a two-photon microscopy module for imaging oblique planes which we can align *on-line* with the retinotopic plane of the visual system of *Drosophila* (Fig. 1d). To this end, we employ the remote focusing principle (Botcherby et al., 2008). With remote focusing, the focal spot of the laser can be moved fast along the optical axis by means of the movement of a third, remote mirror. The independent control over all three imaging axes obtained this way, together with a software module then enables us to freely rotate the imaging plane in 3D. While other implementations of remote focusing have compromised on optical resolution (Rupprecht et al., 2016), are only partly
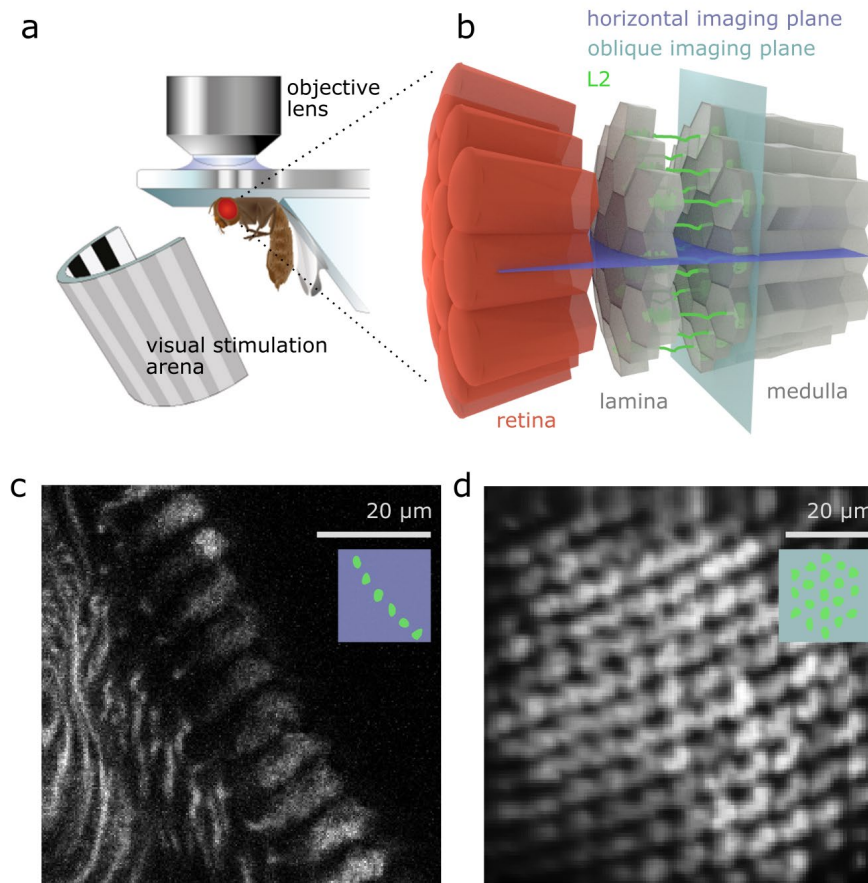
Figure 1 | Imaging the Drosophila visual system. a) Setup for *in vivo* calcium imaging of *Drosophila* during visual stimulation. The fly is fixed on a holder while the brain is viewed from the dorso-posterior side. Visual stimuli are shown on a cylindrical screen surrounding the fly. Schematic not to scale. b) The visual system of *Drosophila* consists of retinotopically-arranged layers of neuropil. The horizontal imaging plane (purple) sections the visual system in a particular orientation. An oblique imaging plane (turquoise) can be aligned to the retinotopic plane of the visual system. The axons of the cell type L2 lie in layer 2 of the medulla. For simplicity, the chiasm between lamina and medulla is omitted in this schematic. c) Horizontal plane of L2 expressing GCaMP6f. d) Oblique, retinotopic plane of axon terminals of L2 in the medulla layer 2 expressing GCaMP6f.

compatible with existing hardware (Botcherby et al., 2012; Colon and Lim, 2015), or have solely been used to jump planes (Sofroniew et al., 2016), our module can be added to standard two-photon microscopes and oblique planes can be imaged with no tradeoff in resolution or frame rate.

With this technique, we image over 100 axon terminals in the retinotopic plane of the cell type L2 and its responses to natural scenes, and thus for the first time reconstitute the 2D representation of the visual surround as represented by this cell type. L2 receives direct input from photoreceptors (Rivera-Alba et al., 2011) and codes for local light decrements when probed with simple stimuli (Clark et al., 2011; Freifeld et al., 2013; Reiff et al., 2010). Therefore, we expect that L2 codes for the inverse of local luminance of natural scenes.

RESULTS

Imaging oblique planes with remote-focusing two photon microscopy
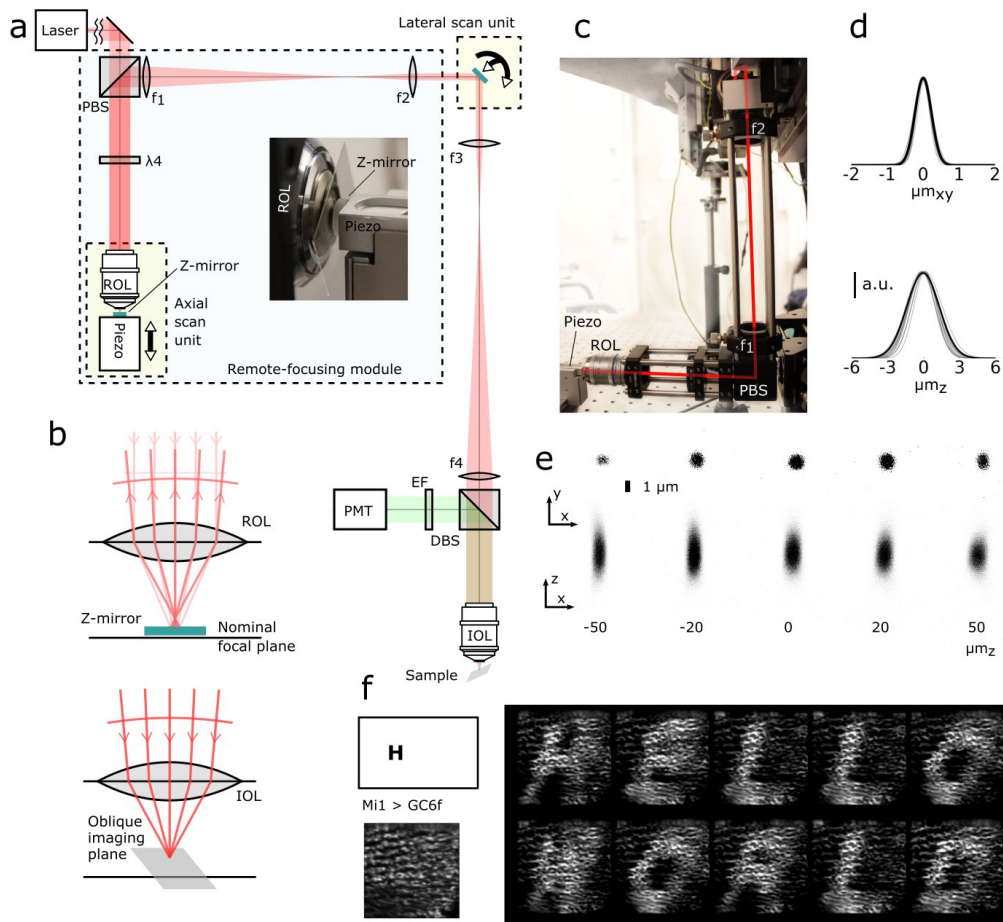
Figure 2 | Remote-focusing two-photon microscope. a) The remote-focusing module in a standard two-photon microscope. A polarization-based beam splitter and λ/4 plate guide the beam to the axial scan unit, which is comprised of the remote objective lens (ROL) and the motorized Z-mirror. Two lenses (f1 and f2) in 4f configuration image the beam from the back aperture of the ROL to the lateral scan unit. Inset: Image of the axial scan unit. b) Schematic of the beam waveform and focal point at the ROL and imaging objective lens (IOL). A collimated beam entering the ROL is reflected by the Z-mirror. Depending on its position, the beam at the back focal plane of the ROL will have non-zero vergence. The beam is imaged to the IOL, where it results in a focal spot outside the nominal focal plane of the lens. c) Image of the remote-focusing module. d) Fit of a Gaussian to the point spread functions in e) in xy and z. Grey represents the individual measurements, black the average. e) Point spread function of fluorescent beads at five different positions in z and from different angles. f) Calcium imaging of the retinotopic plane of Mi1 axon terminals in layer 10 of the medulla. Top left: A letter is displayed on the stimulus screen close to the receptive fields of the cells in view. Bottom left: Calcium baseline level without a stimulus. Right: Fluorescence increase in response to different letters.

We first set out to design a remote-focusing module that can be easily integrated into a standard two-photon microscope and that provides homogenous resolution throughout a large volume (Fig. 2a). Remote focusing was first devised by Botcherby and colleagues (2008). The principle is based on creating a beam with a convergent or divergent waveform, which, upon passing through a lens, will focus before or beyond the lens' focal plane, respectively (Fig. 2b, bottom). To create such a beam and to easily modulate the beam vergence, we introduced a second objective lens, the remote lens, and a mirror (z-mirror, Fig. 2a inset, Fig. 2b, top), moveable with a piezo motor. The collimated laser beam passes through the remote lens, is reflected by the z-mirror and passes back through the lens again. Depending on the position of the z-mirror with respect to the focal plane of the lens, the resulting waveform is convergent or divergent (Fig. 2b, top). To image the back aperture of the remote lens onto the back focal plane of the imaging objective lens, we used two

telescopes in 4f configuration (Fig. 2a and c), with the intermediate conjugate plane at the lateral scan unit. The magnifications of the telescopes were chosen such that the overall magnification between the remote and imaging lens was one. This way, aberrations introduced by the remote lens are cancelled out by the imaging lens (Botcherby et al., 2007). As a result, the optical resolution is diffraction-limited throughout the imaging volume, and the relationship between mirror position and position of the laser focal spot in z is linear (Botcherby et al., 2007).

To confirm the optical resolution across a large range, we measured the point spread function of 1 μm fluorescent beads at different positions in a 100x100x100 μm volume. As expected from the theoretical estimate (Details see Methods) given the wavelength and numerical aperture, the lateral resolution limit lies below 1 micrometer at all positions in the volume (Fig. 2d-e and Fig S1). The axial resolution limit, defined as the full width at half-maximum (FWHM) of the bead point spread function, lies at around 3 μm at all positions in the imaging volume (Fig. 2d-e and Fig. S1), which is close to the theoretical estimate of 2.7 μm.

To utilize the remote focusing module for imaging oblique planes, we extended the functionality of the open source software ScanImage to control the movement of the motorized z-mirror in addition to controlling the lateral scan mirrors (x- and y-mirrors). In ScanImage, the imaging plane is represented as a sequence of two-dimensional vectors, each dimension corresponding to the movement of one mirror. We extended this representation to three dimensions, the third dimension reflecting the movement of the z-mirror. With simple matrix operations, the imaging plane can then be rotated and translated arbitrarily in the 3D imaging space. User controls in the GUI enabled *on-line* plane rotation, such that the plane can easily be aligned to cellular structures.

To show the practical applicability of our method, we next performed calcium imaging of oblique planes in the *Drosophila* visual system. We expressed the calcium indicator GCaMP6f in the medulla interneuron Mi1, which is a unicolumnar cell type with its axons in the medulla layer 9 and 10 (Fischbach and Dittrich, 1989). We aligned the oblique imaging plane with the retinotopic map spanned by its axon terminals, viewing the 2D array of axon terminals in the imaging plane (Fig. 2f). Mi1 responds to brightness increments at its receptive field location (Arenz et al., 2017; Behnia et al., 2014). We presented different bright letters in the fly's field of view, each for 1 s duration (Fig. 1f). Cells with a receptive field within the pixels of the letter responded with an increase in calcium levels. As a result, the activity map of Mi1 clearly reflects the respective letter shown to the fly.

Linear response properties of L2

We next imaged the retinotopic array of the lamina neuron L2, and first set out to map its linear response properties. To this end, we expressed the calcium sensor GCaMP6f cell type specific in L2. Aligning the imaging plane with the retinotopic plane spanned by the L2 axon terminals, we measured the linear spatial receptive fields of all cells with using white-noise stimulation and reverse correlation (Dayan and Abbott, 2005). A white noise stimulus consisting of pixels that changed their luminance value randomly was shown to the fly for four minutes. Each cell in the imaging
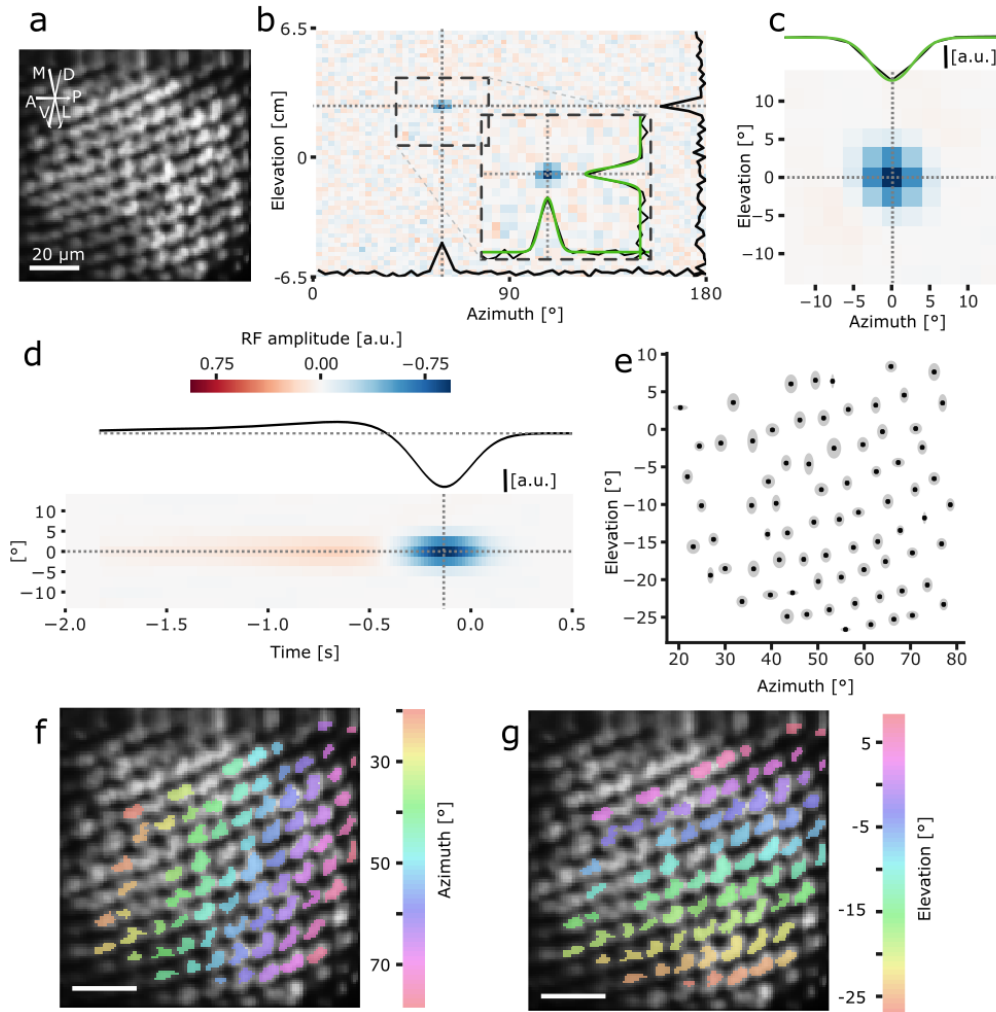
Figure 3 | Receptive field analysis of L2 with white noise. a) Mean image of a recording of L2 axon terminals in layer 2 of the medulla expressing GCaMP6f. b) Reverse correlation of the white noise stimulus with the calcium signal of an example cell, with 1D profiles (black) through the dashed lines. Red corresponds to a positive correlation with the stimulus luminance, blue to a negative correlation. Inset: Magnification of receptive field (RF). 1D profiles (black) were fitted with a Gaussian function (green). c) Average RF of all cells (N = 3 flies, n = 109 cells). The profile in azimuth (black line) was fitted with a Gaussian (green). d) 1D and 2D temporal profiles of the RF. e) RF locations (black dots) with RF size (grey shaded areas) of all cells in an example recording. f) Axon terminals colored by their RF location in azimuth. g) Axon terminals colored by their RF location in elevation.

field of view was easily separable due to its compact axon terminals (Fig. 3a). From each terminal, we then extracted the calcium signal and reverse correlated it to the stimulus, resulting in a spatio-temporal receptive field (Fig. 3b). Consistent with previous studies that probed L2 response properties with flashes and gratings (Freifeld et al., 2013; Reiff et al., 2010), the cells responded to light decreases with an increase in calcium levels. Accordingly, the receptive field was anticorrelated with the luminance of the stimulus. The RFs showed a weak, but anisotropic, antagonistic surround (Fig. 3c and Fig. S2, R^2 of 2D symmetric Gaussian fit to mean = 0.99). To determine the size and location of the receptive fields, we fitted a Gaussian to the azimuth and elevation axes of the spatial receptive field (Fig. 3b inset and Fig. 3c). We termed the mean μ of the Gaussian the receptive field location of each cell. The receptive field size was given by the standard deviation σ. In the temporal domain, L2 showed band-pass characteristics (Fig. 3d). The regular, hexagonal and retinotopic
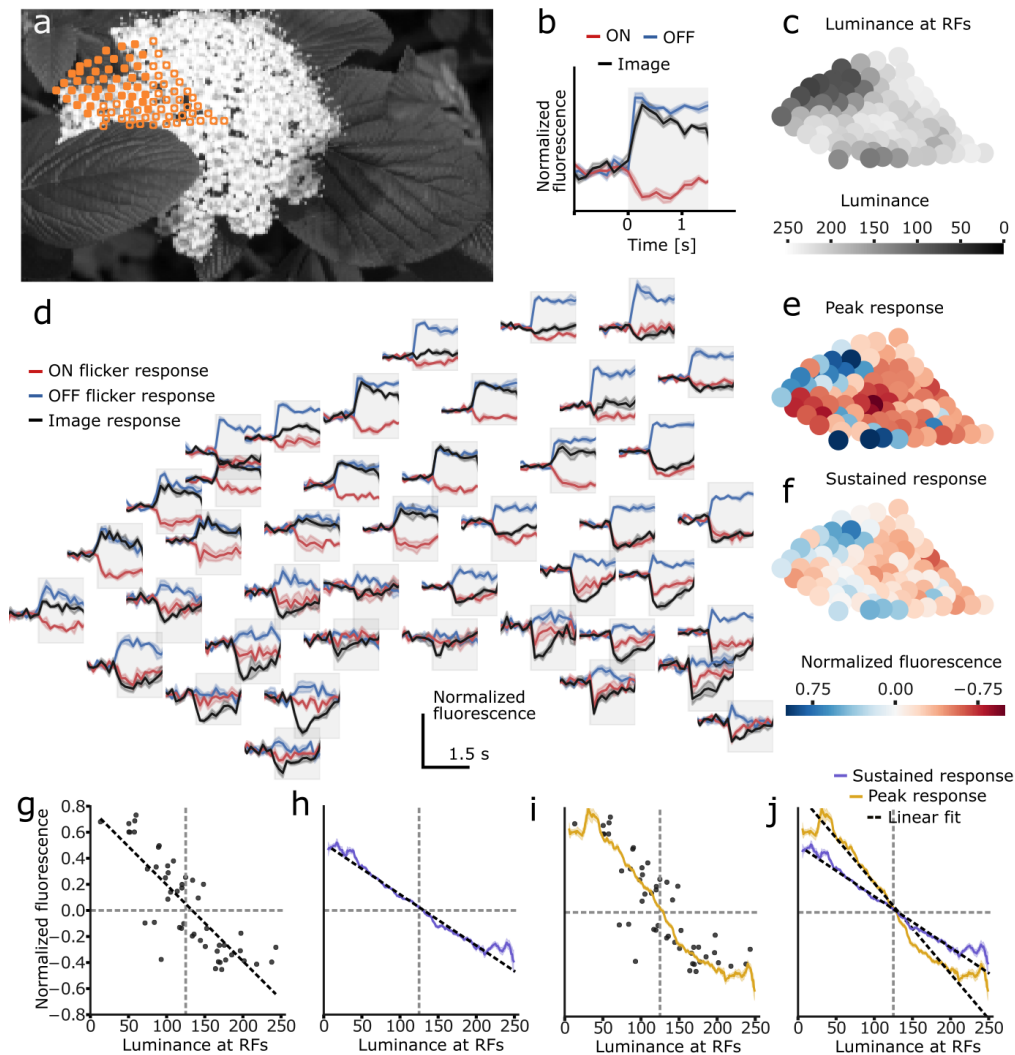
Figure 4 | L2 responses to natural scenes. a) Example image with the receptive field (RF) location of all cells in an example recording. b) Example response of a cell to ON and OFF flicker stimuli and an image. Average over three trials. Shaded areas indicate standard deviation. The grey shaded area indicates the time during which the image was shown. c) Luminance at the RF locations of the example recording, filtered with the average RF size. d) Example responses of a selection of cells (filled circles in a). e) Peak response of all cells in the example recording to the example image. The peak response was defined as the largest absolute response during the time the image was shown. f) Sustained responses like in e. Sustained responses were defined as the average response over the last second during which the image was shown. g) Sustained response of an example cell plotted against the luminance at its RF. Dashed line represents the linear fit. h) Average sustained response (N=6, n=80). Dashed line represents linear fit; shaded area indicates 95 % confidence interval. i) Peak response of an example cell (black dots) and average peak response as in h. j) Comparison of peak and sustained response.

coverage of visual space by L2 is visualized by the receptive field locations (Fig. 3e-g). On average, the receptive field size was 2.5 ± 0.5 ° (σ of the Gaussian fit, corresponding to 5.8 ± 1.2 ° full width at half maximum).

L2 responses to natural scenes encode luminance

We next asked how L2 represents natural scenes. To this end, we showed natural images to the fly while recording from the retinotopic plane of L2 axon terminals. The stimulus set consisted of 50 images from the van Hateren Image Database (van Hateren and van der Schaaf, 1998), showing a variety of natural environments (Fig. S3). Each image was displayed for 1.5 seconds replacing a
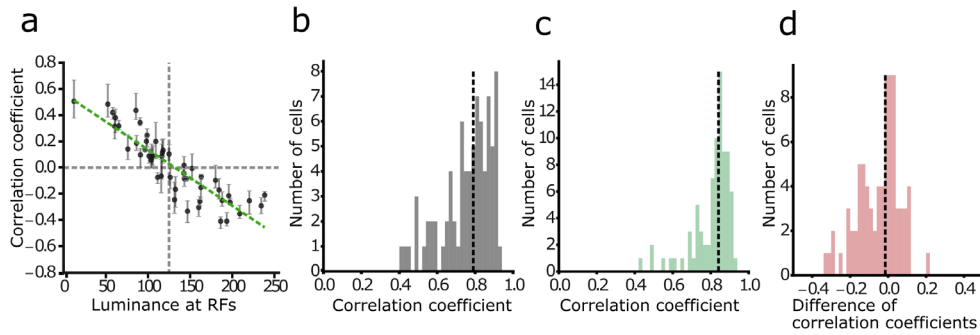
Figure 5 | Comparison of linear RF with natural scene responses. a) Example cell sustained responses (black dots) and trial-by-trial variance. Error bars denote the largest and smallest response out of three trials. Dashed green line is the linear least squares fit. b) Trial-to-trial variance expressed as the between-trial correlation. Details see Methods. c) Correlation of linear prediction to natural scene responses. d) Difference of between-trials correlation and correlation with linear fit.

grey screen. Additionally, an ON and an OFF flicker stimulus – an entirely white or black image, respectively - were included in the stimulus set. Fig. 4a shows one of the images with the receptive field locations of all cells in a representative recording. The calcium signals from all cells were baseline subtracted and normalized (Details see Methods). Responses from three trials were averaged (Fig. 4b). Illustrated by a selection of cells in an example recording to the image in Fig. 4a, the retinotopic array of L2 represented local image luminance (Fig. 4d) with a calcium increase to dark areas and a decrease to bright areas.

To quantitatively compare the cellular response to the luminance of the image at its RF, we filtered the images with a 2D Gaussian function, the standard deviation of the filter given by the RF size from the white noise analysis (Fig. 4c). We then defined two response phases: The peak response was defined as the largest absolute response during the time the image was shown (Fig. 4e). The sustained response was defined as the average response over the last second during which the image was shown (Fig. 4f). Consequently, we compared the responses of each cell to the luminance at its RF location for all 50 images. L2 exhibited luminance-dependent responses (representative cell: Fig. 4g; Bonferroni corrected t-test: $p<0.05$ for 82/83 cells), encoding brightness increments with a calcium decrease and brightness decrements with a calcium increase. On average, both the sustained and peak response component linearly encode the luminance at the RF location of a cell (Fig. 4h-j, $R^2 = 0.67$ for peak responses $R^2 = 0.65$ for sustained responses, $p<0.0001$). The correspondence was linear over most of the luminance range with a slight saturation at low and high luminance values.

L2 responses to natural scenes can be predicted by linear response properties

We next set out to test how well the L2 response to natural scenes can be predicted with their linear response properties. Filtering the natural scenes with the Gaussian kernel obtained from the white noise analysis yields the linear prediction for the L2 response to natural scenes. We fitted a linear function to each individual cell's correspondence between its image responses (Fig. 5a) and the linear prediction, and computed the correlation coefficient between them. For the example cell in Fig. 5a, the correlation coefficient was 0.904. Because of measurement and cellular noise, this

correlation coefficient cannot reach the theoretical maximum of one. We thus determined an upper bound for the correlation coefficient from the trial-to-trial variability. To this end, we computed the correlation coefficient between trials (Details see Methods). For the example cell in Fig. 5a, the between-trials correlation, and thus the upper bound of predictability, was 0.903. On average, for the sustained response component, the between-trials correlation was 0.8 (Fig. 5b) and the correlation with the linear fit was similar (Fig. 5c). The difference between them was close to zero (Fig. 5d). Thus, L2 linearly encodes the local luminance of natural images and the prediction from the L2 receptive field, which in essence is the local luminance, can fully explain all predictable variance of the responses. Since we cannot separate the residual variance into cellular or measurement noise, it is possible that a small contribution of the surround receptive field or other small, nonlinear contributions are masked by measurement noise. In summary, our findings demonstrate that for lamina neuron L2, the response properties as determined by an artificial stimulus reflect its response properties to naturalistic stimuli.

DISCUSSION

Here, we presented a remote-focusing module to image oblique planes with uncompromised resolution and frame rate, and which can be added to typical two-photon microscopes. We then used the technique to simultaneously image over 100 cells in the retinotopic plane of the *Drosophila* visual system, a significant proportion of its visual field. Showing natural scenes to the cell type L2, we could demonstrate that that the linear receptive field of L2 can fully predict its responses to natural scenes.

In the original remote-focusing system and in some subsequently adapted, the remote focusing unit follows the lateral scan unit in the beam path, or is embedded between scan and tube lens (Botcherby et al., 2012; Colon and Lim, 2015; Rupprecht et al., 2016). Typical two-photon microscopes do not allow for additional optical elements between lateral scan unit and imaging objective. In this paper, we presented a system where the remote-focusing module is placed ahead of the lateral scan unit in the beam path (see also: Sofroniew et al., 2016). This way, standard two-photon microscopes can be equipped with this module without changes in the arrangement and design of the main optical elements. The distance between two columns in the visual system of the fruit fly is 4-5 μm. Therefore, we aimed for a diffraction-limited focal spot size of around 0.5x0.5x3 μm throughout the volume. A typical sine-corrected objective lens forms a diffraction-limited focal spot in the focal plane, but aberrations are introduced when focusing above or below the focal plane (Botcherby et al., 2008, 2007; Conrady, 1905). By introducing a second objective lens (the remote lens), aberrations are introduced on purpose to be then cancelled out by the imaging lens. Full cancellation is achieved with a 4f telecentric system and an overall magnification of one between the two objectives (Botcherby et al., 2007). A similar system omitted either the telecentric relay or used non-unity magnification (Rupprecht et al., 2016). The axial resolution then depends on the axial position of the mirror, reaching 7 μm, and the relationship between mirror displacement and focus displacement is no longer linear. Overall, for small biological structures a telecentric relay with unity magnification as we present here is preferable.

Most other studies have used remote focusing to jump to different planes in z. Our approach of tilting the imaging plane provides a simple and efficient solution to maximize information content in one, oblique plane (see also: Colon and Lim, 2015) without a slower imaging rate. A comparable result could be attained by volume-imaging approaches using acousto-optic deflector systems (e.g. Grewe et al., 2010; Katona et al., 2012) or electro-tunable lenses (Grewe et al., 2011). Acousto-optic deflectors require beam stabilization and dispersion control (Iyer et al., 2003) and both would involve substantial modifications or an entirely new setup. Thus, our system provides a simpler, low maintenance alternative that utilizes existing two-photon hardware. While we are here only using the module to image oblique planes, arbitrary line scans following individual cells, or volume imaging, albeit at a slower frame rate, would be possible by software extensions.

We demonstrated our technique by imaging the retinotopic plane spanned by the cell type L2 in the visual system of *Drosophila*, while showing natural images. The response properties of L2 to moving gratings, edges and flashes, have been characterized in detail (Clark et al., 2011; Freifeld et al., 2013; Reiff et al., 2010). L2 responds with depolarization to OFF flicker in the center, but with slight hyperpolarization to OFF flicker in the surround, and vice versa to ON flicker (Freifeld et al., 2013). Responses to moving gratings differ slightly for vertical and horizontal directions, suggesting a spatial anisotropy in the receptive field (Freifeld et al., 2013). Moreover, L2 is necessary for the downstream computation of visual motion to dark edges (Clark et al., 2011; Joesch et al., 2013, 2010; Silies et al., 2013). Consistent with these data, the spatial receptive field we obtained with reverse correlation to white noise showed an OFF center and an anisotropic ON surround.

Our results show that the natural scene responses are predictable from the linear receptive field of the cells. It is possible, however, that the following limitations resulted in this predictability: firstly, only static images were shown, and only time-averaged response measures were predicted with the spatial receptive field. Thus, the dynamic properties of L2 might exhibit properties that we did not account for. Secondly, the luminance range in our stimulus arena spans about two orders of magnitude and cannot reproduce the large luminance range in the natural environment. While photoreceptors have adaptive properties and compress luminance information (Juusola and Hardie, 2001) the signals L2 receives in our stimulus arena still likely differ from actual natural conditions. Finally, some studies observed a rectification of L2 responses for brightness increments with flash and edges stimuli (Freifeld et al., 2013; Reiff et al., 2010, but see: Clark et al., 2011). In our study, responses to natural scenes did not show a strong rectification. This discrepancy might be due to the smaller luminance changes in our experimental design, showing all images from an intermediate grey background.

In the blowfly, the dynamic range of lamina monopolar cells (L1-L3), spans about 1.5 log units (Laughlin and Hardie, 1978). Within this range, the cells' response shows a log-linear dependency on luminance. While the dynamic range of L2 in our study is comparable, we find the relationship between response and luminance to be linear rather than log-linear. This difference could potentially arise from differences in calcium and voltage signals, or the non-linear properties of the calcium indicator. Furthermore, responses to time-series of natural scenes have been measured in

lamina monopolar cells of the blowfly (Van Hateren, 1997). The study showed that the photoreceptors as well as the lamina monopolar cells compress the large range of luminances in natural scenes, removing temporal redundancies to form an efficient representation. While here, we did not directly measure, whether L2 is adapting over time to different luminance conditions, as expected, the dynamic range of L2 was centered on the intermediate grey value shown before the appearance of the natural images. This is also consistent with the large body of literature showing that sensory systems have evolved to match the particular statistics of natural scenes (*Drosophila*: Clark et al., 2014; Fitzgerald and Clark, 2015; Leonhardt et al., 2016; Straw et al., 2008; general review: Simoncelli and Olshausen, 2001), and that short-term adaptation mechanisms are in place to adjust the visual system to current circumstances (e.g.: Arenz et al., 2017; Juusola and Hardie, 2001; Strother et al., 2018).

Several studies have compared responses to artificial and naturalistic stimuli (Kayser et al., 2004). For example, complex cells of V1 in cats respond to natural scenes as predicted from their linear receptive field (Touryan et al., 2005). Moreover, natural scenes have been used for reverse correlation analysis. This results in receptive fields comparable to white noise stimuli, and can be a more efficient method in some cases (Ringach et al., 2002; Theunissen et al., 2001; Touryan et al., 2005; Vance et al., 2016; Willmore et al., 2010). Overall, these studies and our finding corroborate the hypothesis that cellular response properties as determined with simple stimuli have predictive value for more complex, naturalistic scenarios.

In conclusion, with remote focusing microscopy we could for the first time access the retinotopic maps in the *Drosophila* visual system, as it was previously only possible in the mammalian retina. Our method will provide new perspectives on spatial relationships of the cells in the visual system of the fly in general, and in particular cells downstream to the lamina, where spatial surround information may play a critical role.

METHODS

EXPERIMENTAL ANIMALS

Drosophila melanogaster were kept on standard cornmeal-agar medium at 25 ° and at 60 % relative humidity in a 12/12 h light/dark cycle. The Gal4-UAS system (Brand and Perrimon, 1993) was used to express GCaMP6f (Chen et al., 2013) in the lamina cell type L2 and the medulla cell type Mi1. The resulting genotype of the L2 experimental flies was w$^+$/w$^-$; R53G02-AD/UAS-GCaMP6f; R29G11-DBD/UAS-GCaMP6f, and of the Mi1 experimental flies was: w$^+$/w$^-$; R19F01-AD/UAS-GCaMP6f; R71D01-DBD/UAS-GCaMP6f. Female flies 1-5 days after eclosion were used for the experiments.

REMOTE-FOCUSING TWO-PHOTON MICROSCOPY

We designed a remote-focusing two-photon microscope based on the two-photon microscope described in Euler et al. (2009) and the principle of remote focusing described in Botcherby et al. (2008).

BASIC TWO-PHOTON SYSTEM

A MaiTai eHP DS Ti:Sapphire oscillator was used as laser source. A beam attenuator (Newport VA-BB-2-CONEX) was used to control the laser intensity, and a telescope initially collimated the beam and widened it approx. 4-fold. The lateral scan unit comprised of two galvanometer optical scanners (Cambridge Technologies 6215H). A telescope in 4f configuration then magnified the beam with magnification $M_1 = 4$ and imaged it to the back aperture of the sample objective lens. The water immersion Olympus Lumplfln 40x w with an NA of 0.8, back aperture of 9 mm and working distance of 3.3 mm was used as sample objective lens. A Photomultiplier Tube (Hamamatsu H10770-PB40) collected the emitted fluorescence signal, which was then digitized with a NI PCI-6110 data acquisition board. A band-pass filter (BrightLine 514/30) and a short-pass filter ensured that light from the visual stimulation arena and the laser could not reach the photomultiplier tube.

REMOTE FOCUSING MODULE

The remote focusing module was positioned in the beam path after the attenuator and the initial telescope, and before the lateral scan unit. The beam enters the module at a polarization-based beam splitter (PBS). It then passes through a $\lambda/4$ plate, enters the remote objective lens, is reflected by the z-mirror, and travels back through the remote objective lens and the $\lambda/4$ plate. It is in turn reflected by the PBS and is imaged with a telescope to the lateral scan unit.

The position of the z-mirror with respect to the focal plane of the remote objective lens was controlled by a piezoelectric actuator and a closed-loop DSP controller. The actuator was manufactured customized (nanoFaktur) to have a fast settling time (1 ms at 1 μm precision) at a large traveling range (280 μm) and a position sensor. A DSP Controller (nPoint LC.402 with extended current) was used to control the actuator. The z-mirror was glued on the actuator. We chose a lightweight mirror (Thorlabs PF05-03-M01, 12mm diameter) that can tolerate the laser intensities needed for calcium imaging. Since the laser beam diameter on the mirror is very small (down to <1 μm if the mirror position is exactly in the focal plane), the energy per area is considerable. In addition, for mode-locked laser sources with femtosecond pulse length, damage thresholds are not usually determined by the manufacturer. Therefore, we tested several different mirrors for applicability in this specific setting. Metal-coated mirrors were generally found to be superior to dielectric mirrors. The actuator together with the z-mirror were mounted on a XYZ translation stage (Thorlabs PT3/M). This enabled a simple manual translation of the z-mirror and actuator in the case of local mirror damage.

As the remote objective lens, we chose a Leica HC PLAN APO objective with 40x magnification, 0.85 NA, 0.21 mm free working distance and transmission of about 70% at 920 nm. Since the sample objective lens is a water immersion objective, we could not use the exact same model for the remote focusing lens. To account for the cover slip correction of this lens, we mounted a cover slip to its front.

Two telescopes in the telecentric 4f configuration imaged the back aperture of the remote objective lens onto the back aperture of the imaging objective lens, with an intermediate conjugate plane at

the lateral scan unit. The magnification of the telescope in the remote focusing module was $M_2 = 3$. Considering the refractive indices of air and water, the overall magnification of the system is then at unity and therefore free of spherical aberrations (Botcherby et al., 2007).

VALIDATION

To validate the resolution and scaling along the z-axis, we imaged 1 μm diameter fluorescent beads (Polysciences Fluoresbrite) embedded in agarose. The approximate theoretical resolution limit of an imaging system, i.e. the minimum distance between two points to be resolvable, is given by the Rayleigh criterion (adapted for two-photon microscopy, see Yang and Yuste, 2017):

$$R_{xy} = \frac{0.4 \, \lambda}{NA} = 0.46 \, nm$$

$$R_z = \frac{1.4 \, \lambda \, n}{NA^2} = 2.7 \, nm$$

Where NA is the numerical aperture of the objective lens, λ is the wavelength and n the refractive index of the imaging medium. The cross-section of a bead was measured in x, y and z at different points in a 100x100x100 μm volume. A Gaussian was fitted to each cross-section, and the full width at half maximum (FWHM) was defined as the resolution of the system.

SOFTWARE

A customized version of the open-source software ScanImage 5.1 (Pologruto et al., 2003) was used to acquire the data. NI data acquisition boards were used to generate the output voltage for the mirrors (PCIe-6353 with two BNC 2110 accessories) and to collect the signal from the photomultiplier tube. In a second input channel, the trigger signal for synchronization of the visual stimulus with the imaging data was recorded. We extended the functionality of ScanImage (Matlab2013b) to control the movement of the piezo actuator for scanning in z in the same way as it controls the galvanometer motors for lateral scanning. A third output channel generated the voltage that defined the position of the z-mirror. A scaling parameter (0.0565 V/μm$_z$, determined empirically) ensured the appropriate transformation of voltage to the axial displacement in the sample space. ScanImage represents an imaging plane as a time series of 2D vectors, the positions of the focal point in a virtual space. The first and second dimension correspond to the movement of the x and y scan mirrors, respectively. We added a third dimension to this representation. Simple matrix transformations applied to the 3D vectors can then arbitrarily rotate the imaging plane in 3D space. With additional rotational controls in the graphical user interface, the user can rotate the imaging plane in the 3D imaging volume in all directions during imaging.

A full list of parts for the remote focusing arm as well as the customized version of ScanImage 5.1 are available from the authors upon request.

VISUAL STIMULATION ARENA

For the display of visual stimuli to the fly, we used a custom-build projector-based arena as described previously (Arenz et al., 2017). Briefly, visual stimuli were projected onto a semi-

transparent cylinder that was positioned in front of the fly, spanning 180 ° in azimuth and 105 ° in elevation of the fly visual field. Stimuli were projected with a frame rate of 180 Hz in 255 brightness steps. The brightest absolute luminance of the arena reached approximately 250 candela/m2, and the wavelength was restricted to > 550 nm, to separate it from the GCaMP emission wavelength. Stimuli were written in Python 2.7 using the Panda3D software. To synchronize the visual stimulation with data acquisition, a small corner of the arena displayed a trigger signal that was recorded together with the imaging data.

VISUAL STIMULI AND NATURAL IMAGES DATASET

To determine the linear receptive field of the neurons, a white noise stimulus was shown for 4 minutes. The stimulus consisted of a grid of 64x54 pixels, each pixel corresponding to approximately 2.8x1.9 ° in visual space. For a subset of acquisitions, the luminance of each pixel changed in every third frame according to values drawn from a normal distribution. The luminance values were then slightly low-pass filtered in space and time ($\tau = 0.5$ s; termed continuous, Gaussian noise). For a second subset, the luminance of each pixel could either take the value 0 or 255, set by a pseudo-random number generator every 0.5 s (termed binary, discrete noise).

We used images from the Van Hateren natural image dataset (van Hateren and van der Schaaf, 1998) for the natural scenes stimulus. Of the 4000 images in the dataset, we selected 50 images that represented the image diversity and showed variable luminance and contrast characteristics (Fig. S3). The images were then cropped, downsampled to 180x260 px (corresponding to approx. 1x0.4 ° in the fly's visual space) and gamma corrected ($\gamma=1.5$) to compress the luminance range to the 8 bit depth of the stimulation arena.

At the beginning of each trial, following a gray screen (luminance = 125), an image was displayed on the full screen. The image stayed for 1.5 s before the screen turned gray again. To compare the image responses to luminance steps, we added two additional stimulus conditions. Following a gray background, either a black screen (luminance = 0, OFF flicker stimulus) or a white screen (luminance = 255, ON flicker stimulus) were shown for 1.5. This resulted in a total of 52 stimulus conditions in the natural scenes stimulus. Each stimulus condition was repeated three times and trials were randomized for each acquisition.

CALCIUM IMAGING PROCEDURE

Flies were prepared for imaging following a standard procedure (Maisak et al., 2013). Briefly, flies were anesthetized on ice and then glued to a holder with the posterior side of the head exposed to the objective lens through a small opening. Ringer's solution was applied, and the cuticle, fat and trachea at the back of the head were removed for optical access.

Initially, a moving grating was shown to locate areas of axon terminals with large calcium responses. The imaging plane was then rotated with the rotational controls in the ScanImage GUI until the plane spanned by the 2D array of axon terminals was visible. Then, the white noise stimulus was shown, followed by the natural scenes stimulus, the latter of which lasted about 20 minutes. For a subset of acquisitions, we showed a white noise stimulus again after the natural images stimulus, to

verify that the receptive field locations of the cells had not changed. Recordings were terminated prematurely if cells stopped showing calcium responses, and data from these recordings were not used for further analysis. All data were recorded at a frame rate of 7.5 Hz and a zoom of 3, equivalent to a field of view size of about 90 μm.

DATA PROCESSING AND ANALYSIS

Data analysis was performed in Python 2.7 with custom-written software.

PREPROCESSING

Image sequences were registered to the mean image with a phase correlation algorithm described in detail in Arenz et al., (2017). Individual cells' axon terminals were selected based on initial thresholding of the high-pass filtered mean image followed by manual inspection and corrections.

WHITE NOISE ANALYSIS

For baseline subtraction, a dynamic baseline fluorescence level was determined by low-pass filtering (τ = 5 s) the raw fluorescence signal for each cell. The baseline was then shifted by in time by $\tau * framerate/2$ and subtracted from the raw fluorescence signal, resulting in the calcium signal ΔF. For each cell, reverse correlation of the calcium signal to the white noise stimulus was then performed, described in detail in (Arenz et al., 2017), resulting in the spatio-temporal linear receptive field (RF). RFs were normalized and cells with a z-score < 7 and recordings with the binary noise stimulus were not included in the white noise analysis (Fig. 3). Cells with RFs with a z-score < 5.5 or a standard deviation of trial-to-trial correlation coefficients of > 0.04 were not included in the natural scenes analysis (Fig. 4 and 5). A 1D Gaussian function was fitted to the azimuth and the elevation axes of the spatial RF. The center of the fit was consequently used as the RF location for each cell.

NATURAL SCENES ANALYSIS

For each cell, the raw fluorescence signal in the second before the start of a trial was used as baseline and subtracted from the fluorescence signal during that trial. Subsequently, responses to three trials were averaged to obtain the mean response for each stimulus condition. Cells that exhibited very small responses to ON and OFF flicker stimuli (< 200 difference in baseline-subtracted raw fluorescence units) were excluded from further analysis. The peak and sustained responses, R_{peak} and R_{sust}, were quantified as follows:

$$R_{peak} = \begin{cases} \min(\Delta F_s) & if \ |\min(\Delta F_s)| > |\max(\Delta F_s)| \\ \max(\Delta F_s) & if \ |\min(\Delta F_s)| < |\max(\Delta F_s)| \end{cases}$$

$$R_{sust} = mean(\Delta F_{s,t>1})$$

With ΔF_s being the calcium signal during the stimulus presentation, and $\Delta F_{s,t>1}$ the calcium signal during the last 0.5 s of stimulus presentation. The calcium signals and the response quantifications

were then divided by a normalization factor. The factor was defined for each cell as 4 times the standard deviation of R$_{sust}$ over all stimulus conditions.

LINEAR REGRESSION AND CORRELATIONS

For each cell, we fitted a linear function to its mean responses and the image luminance at its RF (Fig. 4 and 5) using the least squares method. How well the fit could predict the data was expressed as the Pearson's correlation coefficient ϱ between fit and mean data (average over three trials). ϱ^2 is then equivalent to the variance explained by the fit.

The correlation coefficient between trials was determined as the mean of the correlation coefficient of each pair of trials, determined independently for each cell:

$$\rho_{trials} = \frac{\rho_{0,1} + \rho_{0,2} + \rho_{1,2}}{3}$$

Where $\varrho_{0,1}$ denotes the Pearson's correlation coefficient between the first and second trial of all images. This measure defines an upper bound for the variance predictable by the fit.

REFERENCES

Akerboom J, Carreras Calderón N, Tian L, Wabnig S, Prigge M, Tolö J, Gordus A, Orger MB, Severi KE, Macklin JJ, Patel R, Pulver SR, Wardill TJ, Fischer E, Schüler C, Chen T-W, Sarkisyan KS, Marvin JS, Bargmann CI, Kim DS, Kügler S, Lagnado L, Hegemann P, Gottschalk A, Schreiter ER, Looger LL. 2013. Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Front Mol Neurosci* **6**.

Aptekar JW, Frye MA. 2013. Higher-order figure discrimination in fly and human vision. *Curr Biol* **23**:R694–R700.

Arenz A, Drews MS, Richter FG, Ammer G, Borst A. 2017. The temporal tuning of the Drosophila motion detectors is determined by the dynamics of their input elements. *Curr Biol* **27**:929–944.

Baden T, Berens P, Franke K, Roman-Roson M, Bethge M, Euler, Román Rosón M, Bethge M, Euler T. 2016. The functional diversity of retinal ganglion cells in the mouse. *Nature* **529**:345–350.

Behnia R, Clark DA, Carter AG, Clandinin TR, Desplan C. 2014. Processing properties of ON and OFF pathways for Drosophila motion detection. *Nature* **512**:427–430.

Borst A, Helmstaedter M. 2015. Common circuit design in fly and mammalian motion vision. *Nat Neurosci* **18**:1067–1076.

Botcherby EJ, Juškaitis R, Booth MJ, Wilson T. 2008. An optical technique for remote focusing in microscopy. *Opt Commun* **281**:880–887.

Botcherby EJ, Juškaitis R, Booth MJ, Wilson T, Juskaitis R, Booth MJ, Wilson T. 2007. Aberration-free optical refocusing in high numerical aperture microscopy. *Opt Lett* **32**:2007–2009.

Botcherby EJ, Smith CW, Kohl MM, Débarre D, Booth MJ, Juškaitis R, Paulsen O, Wilson T. 2012. Aberration-free three-dimensional multiphoton imaging of neuronal activity at kHz rates. *Proc Natl Acad Sci USA* **109**:2919–24.

Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **415**:401–415.

Chen T-W, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger

MB, Jayaraman V, Looger LL, Svoboda K, Kim DS. 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**:295–300.

Chklovskii DB, Koulakov AA. 2004. Maps in the brain: What can we earn from them? *Annu Rev Neurosci* **27**:369–392.

Clark DA, Bursztyn L, Horowitz MA, Schnitzer MJ, Clandinin TR. 2011. Defining the computational structure of the motion detector in Drosophila. *Neuron* **70**:1165–1177.

Clark DA, Fitzgerald JE, Ales JM, Gohl DM, Silies M a, Norcia AM, Clandinin TR. 2014. Flies and humans share a motion estimation strategy that exploits natural scene statistics. *Nat Neurosci* **17**:296–303.

Colon J, Lim H. 2015. Shaping field for 3D laser scanning microscopy. *Opt Lett* **40**:3300–3303.

Conrady AE. 1905. The optical sine-condition. *R Astron Soc* **515**:501–509.

Dayan P, Abbott LF. 2005. Theoretical Neuroscience. MIT Press.

Demb JB, Singer JH. 2015. Functional circuitry of the retina. *Annu Rev Vis Sci* **1**:263–289.

Denk W, Detwiler PB. 1999. Optical recording of light-evoked calcium signals in the functionally intact retina. *Proc Natl Acad Sci* **96**:7035–7040.

Diamond JS. 2017. Inhibitory interneurons in the retina: types, circuitry, and function. *Annu Rev Vis Sci* **3**:1–24.

Dowling J. 2012. The Retina. An approachable part of the brain, 2nd ed. Belknap Press.

Euler T, Hausselt SE, Margolis DJ, Breuninger T, Castell X, Detwiler PB, Denk W. 2009. Eyecup scope - optical recordings of light stimulus-evoked fluorescence signals in the retina. *Eur J Physiol* **457**:1393–1414.

Fischbach K-F, Dittrich APM. 1989. The optic lobe of Drosophila melanogaster. I . A Golgi analysis of wild-type structure. *Cell Tissue* **258**:441–475.

Fitzgerald JE, Clark DA. 2015. Nonlinear circuits for naturalistic visual motion estimation. *Elife* **4**:e09123.

Freifeld L, Clark DA, Schnitzer MJ, Horowitz MA, Clandinin TR. 2013. GABAergic lateral interactions tune the early stages of visual processing in Drosophila. *Neuron* **78**:1075–1089.

Grewe BF, Langer D, Kasper H, Kampa BM, Helmchen F. 2010. High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nat Methods* **7**:399–405.

Grewe BF, Voigt FF, van 't Hoff M, Helmchen F. 2011. Fast two-layer two-photon imaging of neuronal cell populations using an electrically tunable lens. *Biomed Opt Express* **2**:2035–2046.

Iyer V, Losavio BE, Saggau P. 2003. Compensation of spatial and temporal dispersion for acousto-optic multiphoton laser-scanning microscopy. *J Biomed Opt* **8**:460–471.

Ji N, Freeman J, Smith SL. 2016. Technologies for imaging neural activity in large volumes. *Nat Neurosci* **19**:1154–1164.

Joesch M, Schnell B, Raghu SV, Reiff DF, Borst A. 2010. ON and OFF pathways in Drosophila motion vision. *Nature* **468**:300–304.

Joesch M, Weber F, Eichner H, Borst A. 2013. Functional specialization of parallel motion detection circuits in the fly. *J Neurosci* **33**:902–905.

Juusola M, Hardie RC. 2001. Light adaptation in Drosophila photoreceptors: I. Response dynamics and signaling efficiency at 25°C. *J Gen Physiol* **117**:3–25.

Katona G, Szalay G, Maák P, Kaszás A, Veress M, Hillier D, Chiovini B, Vizi ES, Roska B, Rózsa B. 2012. Fast two-photon in vivo imaging with three-dimensional random-access scanning in large tissue volumes. *Nat Methods* **9**:201–208.

Kayser C, Körding KP, König P. 2004. Processing of complex stimuli and natural scenes in the visual cortex. *Curr Opin Neurobiol* **14**:468–473.

Land M, Fernald R. 1992. The evolution of eyes. *Ann Rev Neurosci* **15**:1—29.

Land MF. 1997. Visual Acuity in Insects. *Annu Rev Entomol* **42**:147–177.

Laughlin SB, Hardie RC. 1978. Common strategies for light adaptation in the peripheral visual systems of fly and dragonfly. *J Comp Physiol A* **128**:319–340.

Leonhardt A, Ammer G, Meier M, Serbe E, Bahl A, Borst A. 2016. Asymmetry of Drosophila ON and OFF motion detectors enhances real-world velocity estimation. *Nat Neurosci* **19**:706–715.

Maisak MS, Haag J, Ammer G, Serbe E, Meier M, Leonhardt A, Schilling T, Bahl A, Rubin GM, Nern A, Dickson BJ, Reiff DF, Hopp E, Borst A. 2013. A directional tuning map of Drosophila elementary motion detectors. *Nature* **500**:212–216.

Mauss AS, Vlasits A, Borst A, Feller M. 2017. Visual circuits for direction selectivity. *Annu Rev Neurosci* **40**:211–230.

Meister M, Pine J, Baylor DA. 1994. Multi-neuronal signals from the retina: acquisition and analysis. *J Neurosci* **51**:95–106.

Pfeiffer BD, Jenett A, Hammonds AS, Ngo T-TB, Misra S, Murphy C, Scully A, Carlson JW, Wan KH, Laverty TR, Mungall C, Svirskas R, Kadonaga JT, Doe CQ, Eisen MB, Celniker SE, Rubin GM. 2008. Tools for neuroanatomy and neurogenetics in Drosophila. *Proc Natl Acad Sci USA* **105**:9715–9720.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP. 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**:995–999.

Pologruto TA, Sabatini BL, Svoboda K. 2003. ScanImage: Flexible software for operating laser scanning microscopes. *Biomed Eng Online* **9**:13.

Ready DF, Hanson TE, Benzer S. 1976. Development of the Drosophila retina, a neurocrystalline lattice. *Dev Biol* **53**:217–240.

Reiff DF, Plett J, Mank M, Griesbeck O, Borst A. 2010. Visualizing retinotopic half-wave rectified input to the motion detection circuitry of Drosophila. *Nat Neurosci* **13**:973–978.

Ringach DL, Hawken MJ, Shapley R. 2002. Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *J Vis* **2**:12–24.

Rivera-Alba M, Vitaladevuni SN, Mischenko Y, Lu Z, Takemura S, Scheffer L, Meinertzhagen IA, Chklovskii DB, De Polavieja GG, Gonzalo G. 2011. Wiring economy and volume exclusion determine neuronal placement in the Drosophila brain. *Curr Biol* **21**:2000–2005.

Ronzitti E, Emiliani V, Papagiakoumou E. 2018. Methods for three-dimensional all-optical manipulation of neural circuits. *Front Cell Neurosci* **12**.

Rupprecht P, Prendergast A, Wyart C, Friedrich RW. 2016. Remote z-scanning with a macroscopic voice coil motor for fast 3D multiphoton laser scanning microscopy. *Biomed Opt Express* **7**:1656–1671.

Sanes JR, Zipursky SL. 2010. Design principles of insect and vertebrate visual systems. *Neuron* **66**:15–36.

Segev R, Goodhouse J, Puchalla J, Berry MJ. 2004. Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nat Neurosci* **7**:1154–1161.

Silies M, Gohl D, Fisher Y, Freifeld L, Clark DA, Clandinin TR. 2013. Modular use of peripheral input channels tunes motion-detecting circuitry. *Neuron* **79**:111–127.

Simoncelli EP, Olshausen BA. 2001. Natural image statistics and neural representation. *Annu Rev Neurosci* **24**:1193–1216.

Sofroniew NJ, Flickinger D, King J, Svoboda K. 2016. A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *Elife* **5**:e14472.

Song B-M, Lee C-H. 2018. Toward a mechanistic understanding of color vision in insects. *Front Neural Circuits* **12**.

Straw AD, Rainsford T, O'Carroll DC. 2008. Contrast sensitivity of insect motion detectors to natural images. *J Vis* **8**:1–9.

Strother JA, Wu S-T, Rogers EM, Eliason JLM, Wong AM, Nern A, Reiser MB. 2018. Behavioral state modulates the ON visual motion pathway of Drosophila. *Proc Natl Acad Sci USA* **115**:E102–E111.

Theunissen FE, David S V., Singh NC, Hsu A, Vinje WE, Gallant JL. 2001. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Netw Comput Neural Syst* **12**:289–316.

Touryan J, Felsen G, Dan Y. 2005. Spatial structure of complex cell receptive fields measured with natural images. *Neuron* **45**:781–791.

Van Hateren JH. 1997. Processing of natural time series of intensities by the visual system of the blowfly. *Vision Res* **37**:3407–3416.

van Hateren JH, van der Schaaf A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc Biol Sci* **265**:359–366.

Vance P, Das GP, Kerr D, Coleman SA, Mcginnity TM. 2016. Refining receptive field estimates using natural images for retinal ganglion cells.International Conference on Advanced Cognitive Technologies and Applications. pp. 77–82.

Willmore BDB, Prenger RJ, Gallant JL. 2010. Neural representation of natural images in visual area V2. *J Neurosci* **30**:2102–2114.

Yang W, Yuste R. 2017. In vivo imaging of neural activity. *Nat Methods* **14**:349–359.

SUPPLEMENTARY FIGURES
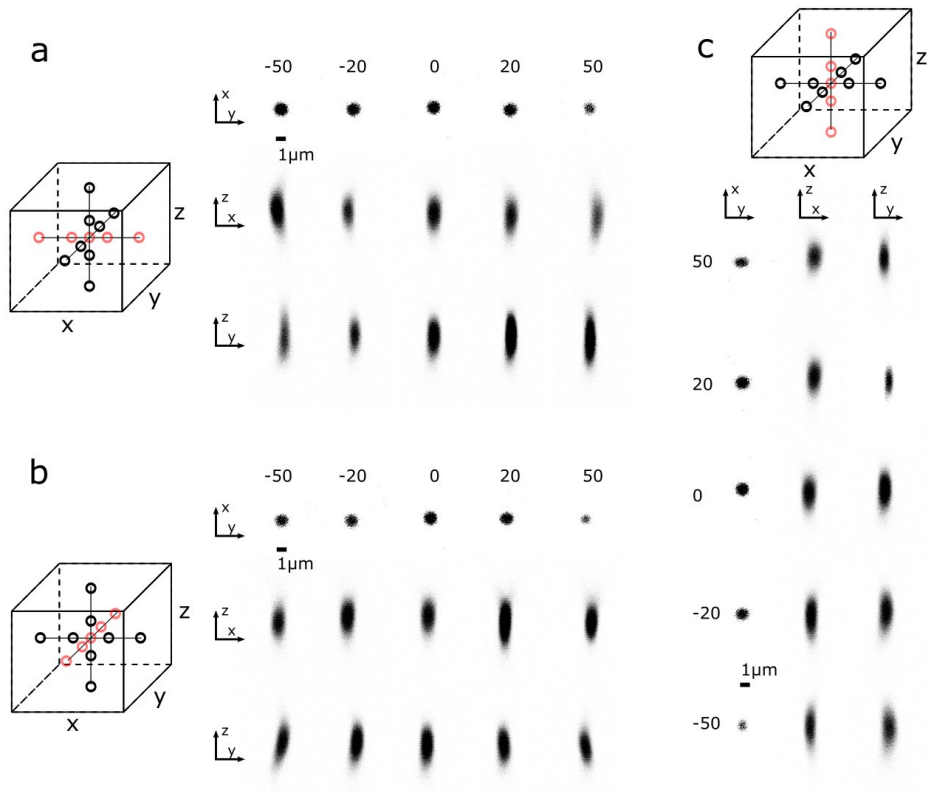
Figure S 1 | Point spread function measurements. a) Point spread function of fluorescent beads at five positions along the first lateral axis (x). Fluorescent beads of 1 μm diameter were imaged in three planes with different orientations: The xy, xz and yz plane. The recorded bead shape is depicted in inversed colors. b) Same as a, but for the second lateral axis (y). c) Same as a and b, but for the z-axis.
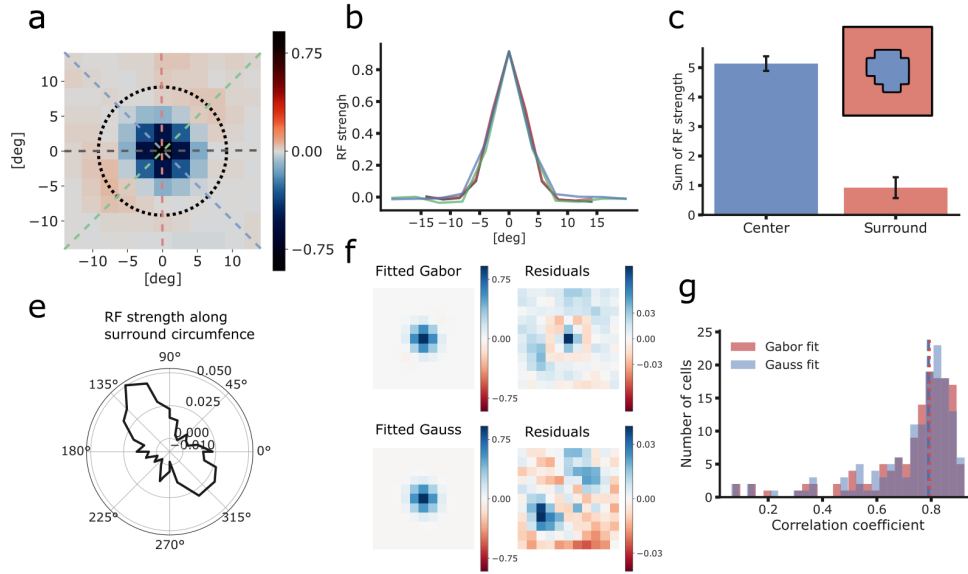
Figure S 2 | Anisotropic receptive field of L2. a) Mean L2 receptive field (RF) with scaled color map, showing the anisotropic surround. The differently colored cuts through the RF are plotted in b and e. b) 1D slices through different axes of the RF, as indicated in a. c) Relative RF strength of the center to the surround. All values of the RF center (blue area in inset) summed up result in the center RF strength (absolute value). All value of the RF surround (red area in inset) summed up result in the surround RF strength. The surround accounts for $1/6$th of the RF strength. e) RF strength along the circumference of the RF indicated in a. A clear anisotropy is observed. f) Fitted Gabor and Gauss filter to the spatial RF, and the corresponding residuals. g) Correlation coefficient of Gabor and Gauss filters to the RFs of all cells. Median correlation for Gauss: 0.790. Median correlation coefficient for Gabor: 0.793. Both filters predict the data equally well.

Figure S 3 | Natural images. All 50 natural images from the van Hateren Image Database (van Hateren and van der Schaaf, 1998). The images spanned 180 ° in azimuth and 105 ° in elevation (the full extent of the stimulation arena) and were compressed with gamma correction to a 255-valued luminance range.

MANUSCRIPT 2

Dynamic signal compression for robust motion vision in flies

Michael S. Drews[1,2,4], Aljoscha Leonhardt[1,4], Nadezhda Pirogova[1,2], Florian G. Richter[1,2], Anna Schuetzenberger[1,2], Lukas Braun[3], Etienne Serbe[1], and Alexander Borst[1,2]

1 Department Circuits-Computation-Models, Max-Planck-Institute of Neurobiology, Martinsried, Germany
2 Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität, Munich, Germany
3 Bernstein Center for Computational Neuroscience, Berlin, Germany
4 These authors contributed equally: Michael S. Drews, Aljoscha Leonhardt

ABSTRACT

Biological and artificial sensory systems need to reliably extract information from highly variable natural signals. Flies, for instance, use optic flow to guide their course and are remarkably adept at estimating image velocity regardless of visual statistics in any given environment. It remains unclear, however, how this robustness is achieved. Current models of fly motion detection explain responses to artificial stimuli in great detail but are vulnerable to the extreme contrast fluctuations pervasive in realistic images. Here, we demonstrate that the Drosophila visual system reduces input variability by continuously adjusting its sensitivity to current contrast conditions. We comprehensively map functional properties of neurons in the motion detection circuit and find that local responses are nonlinearly compressed by a signal that estimates surround contrast. The compressive signal is fast, integrates spatially, and derives from neural feedback. This mechanism resembles divisive normalization as commonly found in vertebrate visual processing, emphasizing that evolutionarily distant neural systems often converge on similar algorithmic solutions. We train fly-like convolutional neural networks8 on estimating the velocity of natural stimuli and show that introducing dynamic signal compression closes the performance gap between model and organism. This may provide a building block for the efficient implementation of low-power machine vision systems. Overall, our work represents a mechanistic end-to-end account of how neural systems attain the robustness required to carry out behaviorally relevant tasks in challenging real-world environments.

AUTHOR CONTRIBUTIONS

M.S.D., A.L., and A.B. jointly conceived the study. M.S.D. and A.L. designed all experiments. A.L. and L.B. conducted behavioral experiments. E.S. recorded electrophysiological responses. M.S.D., N.P., F.G.R., and A.S. performed calcium imaging. A.L. designed and analyzed the convolutional model. M.S.D. and A.L. analyzed data, performed modelling, and wrote the manuscript. All authors participated in editing the manuscript.

Dynamic signal compression for robust motion vision in flies

Michael S. Drews[1,2,4], **Aljoscha Leonhardt**[1,4], **Nadezhda Pirogova**[1,2], **Florian G. Richter**[1,2], **Anna Schuetzenberger**[1,2], **Lukas Braun**[3], **Etienne Serbe**[1], and **Alexander Borst**[1]

[1]Department Circuits-Computation-Models, Max-Planck-Institute of Neurobiology, Martinsried, Germany
[2]Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität, Munich, Germany
[3]Bernstein Center for Computational Neuroscience, Berlin, Germany
[4]These authors contributed equally: Michael S. Drews, Aljoscha Leonhardt

Biological and artificial sensory systems need to reliably extract information from highly variable natural signals[1,2]. Flies, for instance, use optic flow to guide their course[3] and are remarkably adept at estimating image velocity regardless of visual statistics in any given environment[4,5]. It remains unclear, however, how this robustness is achieved. Current models of fly motion detection explain responses to artificial stimuli in great detail but are vulnerable to the extreme contrast fluctuations pervasive in realistic images. Here, we demonstrate that the Drosophila visual system reduces input variability by continuously adjusting its sensitivity to current contrast conditions. We comprehensively map functional properties of neurons in the motion detection circuit and find that local responses are nonlinearly compressed by a signal that estimates surround contrast. The compressive signal is fast, integrates spatially, and derives from neural feedback. This mechanism resembles divisive normalization as commonly found in vertebrate visual processing[6,7], emphasizing that evolutionarily distant neural systems often converge on similar algorithmic solutions. We train fly-like convolutional neural networks[8] on estimating the velocity of natural stimuli and show that introducing dynamic signal compression closes the performance gap between model and organism. This may provide a building block for the efficient implementation of low-power machine vision systems. Overall, our work represents a mechanistic end-to-end account of how neural systems attain the robustness required to carry out behaviorally relevant tasks in challenging real-world environments.

Correspondence: *leonhardt@neuro.mpg.de, drews@neuro.mpg.de*

To rigorously assess the robustness of Drosophila motion processing we took advantage of a widespread visual reflex, the optomotor response. Flies react to whole-field retinal motion by turning in the same direction as their surround which allows them to maintain a straight path under perturbations[3]. For this reflex to work effectively, flies need to respond reliably and independently of the particular visual statistics of their environment. We measured walking responses to a diverse set of moving naturalistic panoramas on a treadmill setup (Fig. 1a). Fly turning was highly consistent across images and velocity tuning curves showed virtually no variation over different scenes, matching previous findings[4] (Fig. 1b, Fig. S1a–c). To quantify reliability at the neural level, we recorded the membrane potential of motion-sensitive lobula plate tangential cells (LPTC) which detect optic flow fields corresponding to yaw rotation and control turning[3] (Fig. 1c). LPTCs pool the output of a retinotopic map of locally motion-sensitive cell types T4 and T5 that
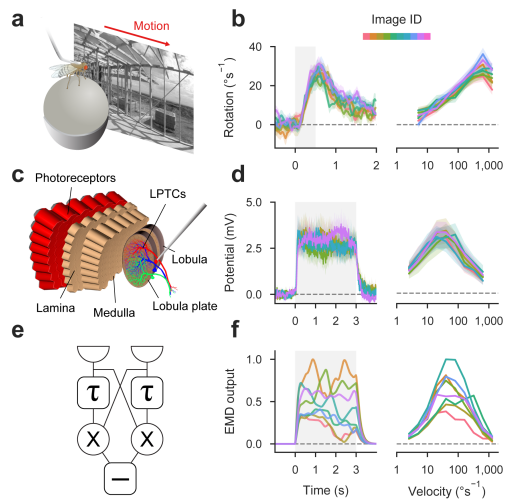
Figure 1. Flies respond more robustly to natural scene variability than predicted by correlation-based motion detectors. a, Illustration of behavioral set-up. Tethered wild-type Drosophila were stimulated with translating natural images. **b,** Left, turning responses for images moving at $80°\,s^{-1}$ (N=16 flies). Each color indicates a distinct scene. Images moved during gray-shaded period. Right, velocity tuning curves for all measured scenes (averaged between 0 and 1s after motion onset). **c,** Illustration of fly visual system. Photoreceptor signals are processed in five retinotopically arranged neuropils. Wide-field lobula plate tangential cells (LPTC) respond to particular optic flow fields. **d,** Left, membrane potential of horizontal system LPTCs in response to images moving at $20°\,s^{-1}$ (N=11 cells from 9 flies). Right, velocity tuning curves (averaged between 0 and 3s after motion onset). **e,** Schematic of an individual correlation-based elementary motion detector (EMD; τ denotes delay line, X multiplication, – subtraction). **f,** Left, responses of an array of EMDs to stimulation with natural images moving at $20°\,s^{-1}$. Right, velocity tuning curves of EMD array (evaluated like LPTC output). Note that in contrast to experiments, mean responses were averaged across many different starting phases. Shaded areas around curves in b,d indicate bootstrapped 68% confidence intervals.

are sensitive to ON and OFF motion[9], respectively, and necessary for the optomotor response[4,10]. Voltage was tuned to scene velocity but again exhibited little image-dependent variation and additionally proved highly stable across time, consistent with earlier work in hoverflies[5] (Fig. 1d).

Fly motion responses are well explained by correlation-based detector models that rely on multiplication of spatially adjacent, asymmetrically filtered luminance signals[3,11] (Fig. 1e). These elementary motion detectors (EMD) account for subtle features of optomotor responses, LPTCs, and T4/T5 cells

1

such as pattern dependency of velocity tuning, reverse-phi sensitivity, or velocity gain adaptation[3, 12, 13]. We tested the robustness of EMDs on the same set of stimuli as in behavior and electrophysiology. As anticipated from similar studies[4, 14, 15, 16], responses were remarkably unreliable across time and images (Fig. 1f). EMD output strongly depends on local contrast as defined by the average difference between light and dark[11]. However, contrast within natural images varies substantially[2]. Correlation-based models invariably confound velocity with contrast and consequently, output from individual EMDs is sparse and fluctuates heavily under naturalistic conditions (Fig. S1d–f). This stands in stark contrast to the observed robustness of motion responses and leads to the central question: How does the fly visual system compensate for natural contrast variability?

Various general mechanisms for adaptation to naturalistic signals have been described in the fly visual system including gain control in photoreceptors or LPTCs[17, 18, 19], redundancy reduction through lateral inhibition[20], subtractive enhancement of flow field selectivity[3], and tailoring of processing to fundamental natural scene statistics[4, 14]. However, none effectively addresses the problem of contrast fluctuations.

We designed an optomotor stimulus to establish whether Drosophila dynamically adapt the sensitivity of motion-induced turning to image contrast, which could serve to normalize variation within natural scenes. The stimulus segregates the visual field into a background pattern containing random luminance fluctuations at a controlled contrast level and a random foreground pattern whose motion triggers turning (Fig. 2a). We confirmed that background produced no net activity in EMDs (Fig. 2b). At zero background contrast, foreground motion induced a reliable optomotor response (Fig. 2c). Turning was fully suppressed at maximum background contrast, proving that turning gain is controlled by surround contrast. Average field luminance was constant for all conditions, so linear processing could not account for the phenomenon. A full measurement of contrast tuning curves for foreground motion revealed a smooth shift of the dynamic range of the optomotor response toward surround contrast (Fig. 2d,e).

To efficiently map features of contrast gain control in a single stimulus condition, we sinusoidally modulated background contrast over time which resulted in oscillations around mean turning (Fig. 2f). Whenever background contrast was high, syndirectional rotation in response to motion was transiently suppressed. Evaluating oscillation amplitude thus allowed a read-out of the level of contrast-induced gain adjustment. We determined the spatial scale by varying spacing between fore- and background. Modulation fell with distance and dropped to baseline at approximately 40° of separation, so contrast estimation was non-local but spatially limited (Fig. 2g, Fig. S2a,b). When we varied oscillation frequency, suppression followed contrast changes up to fast time scales beyond 3 Hz (Fig. 2h, Fig. S2c–f). However, modulation decreased at lower frequencies than for equivalent foreground oscillations which is indicative of temporal integration. Silencing T4 and T5 cells abolished all contrast-guided oscillatory turn-
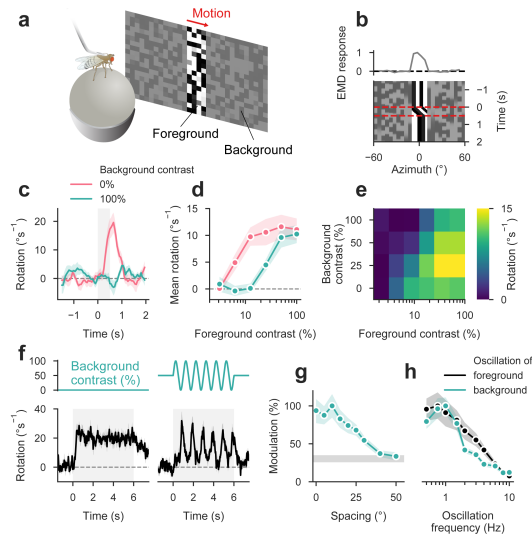
Figure 2. Sensitivity of Drosophila optomotor response is controlled by surround contrast. a, Experimental set-up. Visual display is separated into two areas whose contrast can be set independently. **b,** Bottom, space-time plot of base stimulus. Foreground pattern moved during time span indicated by dashed lines; background is dynamic but contains no coherent motion. Top, time-averaged response of EMD array along azimuth. Only foreground produced net activity. **c,** Turning responses for extreme background contrast conditions (N=16 wild-type flies). Gray-shaded area indicates motion. **d,** Mean rotation (averaged between 0 and 1s after stimulus onset) as function of foreground contrast for two background conditions (N=16). **e,** Heatmap of mean rotation for multiple background conditions. With increasing background contrast optomotor sensitivity shifted rightward (N=16). **f,** Stimulus for mapping magnitude of sensitivity shift (N=16). Left, baseline turning response in the absence of background contrast (foreground contrast 25%). Right, turning response for sinusoidal change in background contrast. During high-contrast phase, optomotor response was suppressed; turning modulation allowed read-out of background-induced changes in gain. **g,** Turning response modulation as a function of spacing between motion stimulus and background (N=16). Gray-shaded bar indicates 68% confidence interval around baseline modulation in the absence of background. **h,** Turning response modulation as a function of carrier frequency for either foreground (N=13) or background (N=13). Shaded area around curves indicates bootstrapped 68% confidence interval.

ing (Fig. S2g–i), suggesting that contrast adaptation is not mediated by a system parallel to motion detection[10]. Our experiments thus point to a rapid, spatially distributed gain control mechanism that emerges in early visual processing.

We next used two-photon calcium imaging to locate the neural origin of contrast adaptation. Fly motion vision circuitry is arranged in cartridges processing visual input retinotopically. In various combinations, lamina cells L1–5 feed into at least four medulla cell types per ON and OFF pathway[3, 21]. Medulla units fall into two classes characterized either by transient temporal filtering and moderate center-surround antagonism in their spatial receptive field (Mi1, Tm3 for ON; Tm1, Tm2, and Tm4 for OFF) or by tonic responses and strong antagonistic surround (Mi4 and Mi9 for ON; Tm9 for OFF)[22]. T4 and T5 cells then compute direction by comparing medulla signals with different dynamics across neighboring cartridges[3, 9, 21]. Here, we additionally estimated linear receptive fields for L1–5 using stochastic stimuli as before[22]
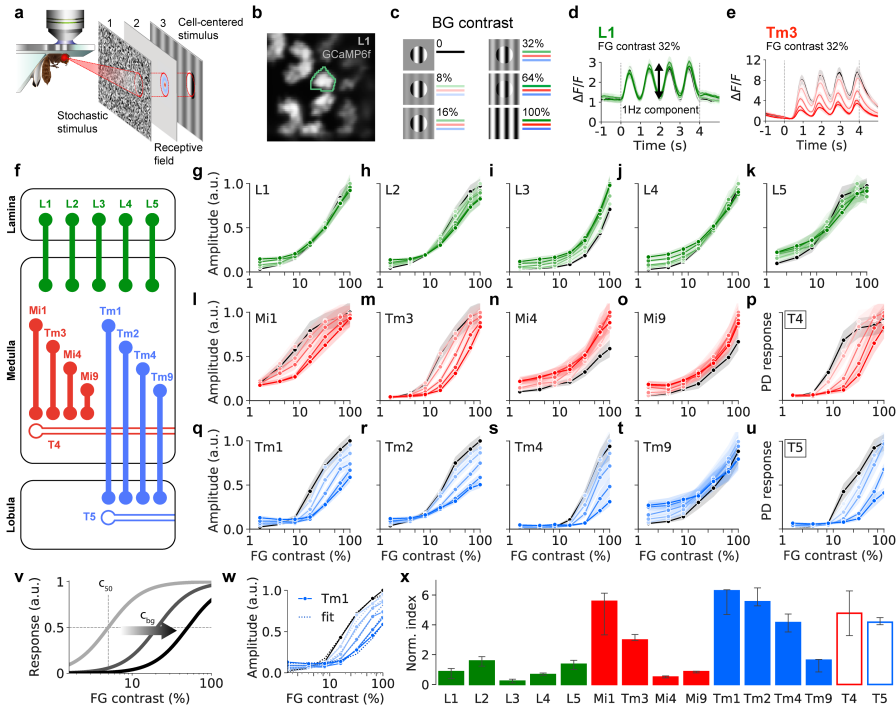
Figure 3. Contrast normalization emerges in transient medulla neurons. a, Schematic of experimental procedure: 1. White noise stimulus. 2. Receptive field reconstruction from single neuron calcium signals. 3. Drifting grating with different contrasts in foreground and background. **b,** Two-photon image of L1 axon terminals expressing GCaMP6f. Green line indicates example region of interest. **c,** Experimental protocol. Darker color shade corresponds to higher background contrast as used in panels g–u. Zero background contrast condition is shown in black. **d,e,** Average calcium responses of L1 and Tm3 for fixed foreground and various background contrasts. **f,** Schematic of the motion circuit including all neurons measured. **g–k,** Contrast tuning curves measured as amplitude of calcium signals at stimulus frequency for L1–L5. Shaded areas show bootstrapped 68% confidence intervals around the mean (L1: 21/7 cells/flies, L2: 26/8, L3: 23/6, L4: 19/6, L5: 18/9). **l–p,** Contrast tuning curves for ON pathway neurons (Mi1: 20/5, Tm3: 21/8, Mi4: 20/13, Mi9: 21/9, T4: 23/10). **q–u,** Contrast tuning curves for OFF pathway neurons. (Tm1: 21/7, Tm2: 20/6, Tm4: 20/13, Tm9: 19/6, T5: 21/9) **v,** Illustration of divisive normalization model for tuning curves. Increasing background contrast c_{bg} shifts the sigmoidal tuning curve from baseline sensitivity c_{50} to higher contrasts. **w,** Example fit of model for Tm1. **x,** Normalization index for all neurons shown as median with 68% bootstrapped confidence intervals. Transient medulla neurons Mi1, Tm3, Tm1, Tm2, and Tm4 as well as T4/T5 exhibited strongest degree of normalization.

(Fig. S3a–t). Consistent with previous functional work[13, 23], spatiotemporal filters grouped into tonic (L3) or transient units (L1, L2, L4, and L5) like in the medulla.

To precisely map context-dependent changes in contrast sensitivity for these cell types, we targeted visual stimuli to individual neurons by determining receptive field coordinates of single axon terminals through a combination of stochastic stimuli and online reverse-correlation (Fig. 3a,b; see Methods). Analogously to behavioral experiments, we then presented drifting sine gratings with independently controlled contrast in foreground (as defined by a 25° circular window centered on the receptive field) and background (Fig. 3c). At fixed foreground contrast, L1 activity followed local grating luminance regardless of background (Fig. 3d). Responses in downstream synaptic partner Tm3, however, showed the signature of gain control as signal amplitude was increasingly suppressed by growing surround contrast (Fig. 3e).

We evaluated calcium modulation at the stimulus frequency to obtain contrast response curves for all columnar cell types and T4/T5 cells (Fig. 3f). Lamina units tracked foreground contrast approximately linearly but were weakly if at all modulated by the surround except for a vertical shift at low lev-

els, likely due to background leaking into the receptive fields (Fig. 3g–k, Fig. S4). In the medulla (Fig. 3l–u), tonic Mi4, Mi9, and Tm9 showed similar tuning as L1–5 and again little surround-dependency. For all transient cells (Mi1 and Tm3 for ON; Tm1, Tm2, and Tm4 for OFF), however, tuning deviated from linearity and increasing background contrast had a strongly suppressive effect. Curves were shifted rightward on the logarithmic axis which corresponds to divisive stretching in linear contrast space. Sensitivity to foreground contrast was generally higher in ON than in OFF units. Importantly, preferred direction responses in T4 and T5 were strongly background-dependent (Fig. 3p,u) even though not all their medulla inputs are subject to gain control.

To quantify tuning curves in detail, we fit a closed-form model based on a common neural computation, divisive normalization (Fig. 3v; see Methods). Here, responses are effectively divided by an inhibitory signal that estimates surround contrast as the average activity of a pool of nearby neurons[6, 7]. Divisive normalization thus compresses signals of varying contrast into a fixed range by dynamically adjusting gain to current conditions[1, 6]. The model accurately reproduced tuning curves for each cell type (Fig. 3w, Table S2).
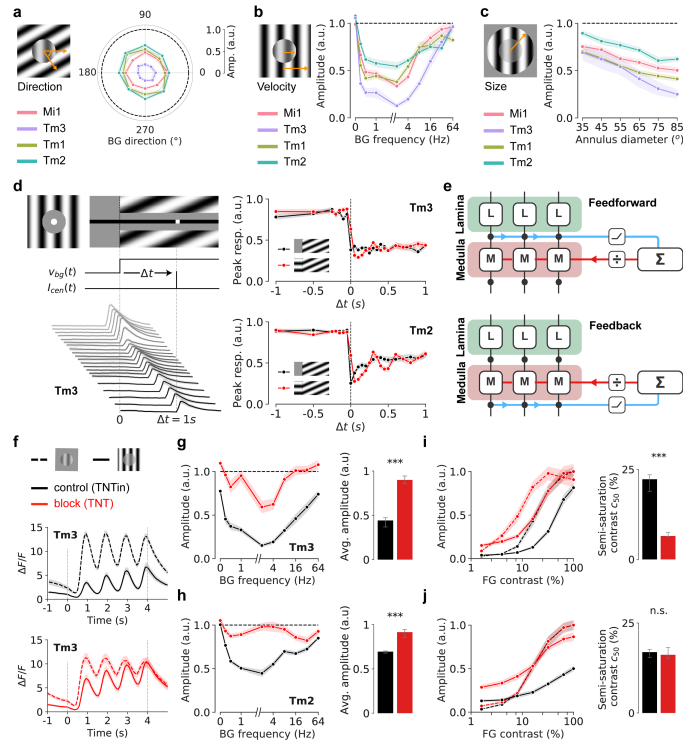
3

Figure 4. Neural contrast normalization relies on rapid feedback from a pool of transient units. a, Polar plot of response amplitude for different directions of background motion. Black dashed line represents reference measured with background contrast of 0%. For each neuron, foreground contrast was chosen to maximize possible background suppression (Mi1: 16%, Tm3: 32%, Tm1: 64%, Tm2: 100%). **b,** Responses for different background contrast frequencies, revealing transient tuning of suppression. **c,** Suppression strength increased with outer diameter of background annulus (Mi1: 21/9 cells/flies, Tm3: 20/6, Tm1: 18/6, Tm2: 21/4 in a–c). **d,** Top left, x-y and x-t plots of contrast-step stimulus for Tm3 (ON center). Background contrast frequency was 3 Hz. Center left, velocity function $v_{bg}(t)$ of background and intensity function $I_{cen}(t)$ of center pulse. Bottom left, mean responses of Tm3 for different time intervals Δt. Right, mean peak amplitude for Tm3 and Tm2 (Tm3: 19/6, Tm2: 20/5). **e,** Schematic of feedforward and feedback model for surround suppression. **f,** Mean responses of Tm3 for TNT block (red) and TNTin controls at background frequency 16 Hz (black; dashed line indicates reference response and solid line the response at full background contrast). **g,** Left, frequency tuning for block experiments (as in b). Right, average amplitude over all frequencies was higher for Tm3 block flies (Tm3 block: 21/5, Tm3 control: 20/5; Mann-Whitney U: 8, ***P<0.001). **h,** Blocking results for Tm2 (as in g; Tm2 block: 20/5, Tm2 control: 25/6; Mann-Whitney U: 17, ***P<0.001). **i,** Left, foreground contrast tuning for block experiments at 0 and 100% background contrast. Right, contrast sensitivity as measured by semi-saturation constant c_{50} was increased for Tm3 block flies (Mann-Whitney U: 39, ***P<0.001). **j,** Blocking results for Tm2 (as in i; Mann-Whitney U: 239, NS P=0.49). Semi-saturation constant at 0% background contrast did not change for Tm2 block flies. Shaded areas show bootstrapped 68% confidence intervals around the mean. Error bars show bootstrapped 68% confidence intervals around the median.

Critically, it accounted for vertical shifts through linear background leakage as well as sigmoidal tuning curves and context-dependent changes in contrast sensitivity through the divisive term. We computed a normalization index from estimated model parameters that quantified the factor by which tuning curves would shift when background contrast was increased from 0% to 100% (see Methods). The index was substantially higher in transient medulla cells (Mi1, Tm3, Tm1, Tm2, and Tm4) and direction-selective T4/T5 cells than in L1–5 or tonic medulla units (Mi4, Mi9, and Tm9; Fig. 3x). Curiously, L2 and L5 exhibited mildly elevated normalization indices which may be due to particular connectivity with parallel neurons[3] or the proposed complex receptive field structure of L2[24].

Thus, fly contrast gain control appeared to be based on divisive normalization that predominantly originates in medulla units with transient response dynamics. We focused on these neurons to investigate the mechanism in detail. Responses in

Mi1, Tm1, Tm2, and Tm3 were equally suppressed for all background grating directions relative to a reference stimulus with zero contrast (Fig. 4a). Contrast frequency tunings for suppression resembled band-pass filters with a peak at 2 Hz (Fig. 4b). Crucially, static backgrounds did not have a suppressive effect. Both isotropy and frequency tunings were strikingly similar to filter properties of transient lamina and medulla units (Fig. S3u,v). Suppression steadily increased with the outer diameter of an annulus containing the background pattern, which again indicated an extended integration area (Fig. 4c). Spatiotemporal features of neural gain control thus matched our findings from behavior (Fig. 2).

To determine the temporal scale of normalization, we designed a contrast-step stimulus in which the foreground was replaced by a single light pulse matching each cell's polarity (Fig. 4d). By varying time between background grating onset and pulse, we scanned the temporal profile of the suppressive signal. For the tested neurons Tm3 and Tm2, we found virtu-

4

ally immediate response reduction within a measurement precision of 50ms. We observed transient ringing of suppression strength at the background frequency. Ringing was stronger when the grating was present before motion onset compared to when it was masked. A similar effect has been described in LPTCs[11] where it results from neural integration of multiple transient, out-of-phase inputs. In sum, these findings suggest that surround suppression derives from a pool of direction-agnostic transient neurons. We built a time-resolved, data-driven model to determine whether a mechanism that pools transient units across space to divisively suppress local responses could reproduce our findings. The model faithfully predicted direction, frequency, and size tunings as well as contrast-step ringing, T4/T5 responses, and LPTC output for our behavioral stimuli (Fig. S5a–k).

Spatial pooling, however, could occur over either feedforward signals from the lamina or feedback from the medulla (Fig. 4e). In vertebrate systems it has proven difficult to distinguish the two[6]. Fly transient units in lamina or medulla have similar temporal properties (Fig. S3u,v) and both implementations produce equivalent steady-state output[7], so we used genetic silencing to pinpoint the source. We co-expressed a calcium indicator and the tetanus toxin light chain (TNT; see Methods) in different cell types, blocking synaptic output and thus feedback from the entire neuron array but leaving feedforward input and calcium signals intact. Compared to controls with inactive TNT, we observed significantly reduced suppression across background frequencies for ON cells Mi1 and Tm3 as well as OFF-sensitive Tm2 but not Tm1 (Fig. 4f,g,h and Fig. S6a,b). In Mi1 and Tm3, baseline contrast sensitivity as measured by the semi-saturation constant of model fits was significantly increased (Fig. 4i and Fig. S6c), suggesting that cells were disinhibited due to reduced pool signal. For Tm2, linear background contribution at low foreground contrasts increased whereas high contrast responses were less suppressed by the surround (Fig. 4j). In the fly, contrast normalization is therefore at least partially based on feedback from a combination of medulla neurons.

Does this type of response normalization account for the robustness of fly motion detection? Previous work on EMDs and natural scenes has exploited compressive transforms but did so heuristically or without surround-dependent gain control[15, 16]. We evaluated natural image responses in the data-driven LPTC model and found moderate reduction of cross-image variability compared to a model with bypassed normalization (Fig. S5l–n). However, post-hoc ablation may specifically disadvantage the simpler model. To investigate performance limits in a principled way, we pursued a task-driven approach. Recent progress in deep artificial networks has made it feasible to use image-processing models of neural systems for rigorously assessing performance on real-world problems[8, 25]. EMD-like architectures are concisely expressed as multi-layer convolutional networks[8] and fully differentiable, rendering them amenable to optimization methods like gradient descent. We designed a fly-like neural network and independently trained possible types of contrast processing such that each model class could optimally adapt
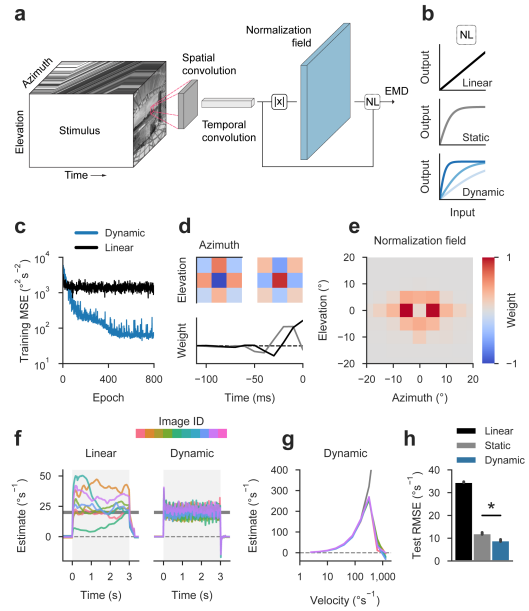
Figure 5. Contrast normalization enhances robustness to natural scene variability. a, Schematic of single convolutional input filter. Motion stimuli are sequentially processed by a spatial 3 x 3 x 1 (azimuth, elevation, time) and a temporal 1 x 1 x 30 filter. Through a transfer function, the signal is combined with a normalization signal generated by a 11 x 11 x 1 convolution operating on full-wave rectified input signal. The output of two distinct channels is processed analogously to multiplicative EMDs. **b,** Input-output relationships for linear, static, and dynamic models. In the dynamic model, response sensitivity is a function of normalization field activity. **c,** Training mean squared error (MSE) for two example models during stochastic gradient descent. **d,** Spatial and temporal receptive fields for the two channels of typical dynamic model. Depicted are normalized filter weights. **e,** Spatial receptive field of normalization pool for the model from d. **f,** Model output for individual images moving at 20°s^{-1} during gray-shaded period. Gray line indicates target velocity. Left, example model without non-linearity. Right, example model with dynamic non-linearity. **g,** Velocity tuning curves of example dynamic model for individual images (averaged between 0 and 3s after motion onset). Gray line indicates true velocity. **h,** Mean performance of trained models on held-out test set, estimated as root mean square error (RMSE; N=22/23/16 for linear/static/dynamic; *P<0.001, t=9.01, Student's t-test with assumed equal variance; only difference between static and dynamic was tested). Error bars indicate bootstrapped 68% confidence intervals.

to a specific, behaviorally relevant task.

All models featured linear, spatiotemporally separable input convolutions (Fig. 5a). We evaluated three alternatives for contrast transformation: a linear stage where output was transmitted unchanged, a statically compressive stage that limited signal range independently of context, and a dynamic compression stage with adaptive gain depending on the output of a contrast-sensitive surround filter (Fig. 5a,b; see Methods). Resulting output from two distinct channels was then processed according to a multiplicative EMD scheme. Through backpropagation and stochastic gradient descent, models were trained to estimate the true velocity of natural images translating at random speeds. All models successfully learned the task on the training set (Fig. 5c). We initialized convolutions randomly but after training observed antagonistic spatial filters and transient temporal filters where one channel was phase-delayed with respect to the other (Fig. 5d,

5

Fig. S7a–c). Models thus made extensive use of redundancy reduction through center-surround configurations[20] and discovered the EMD strategy of delay-and-compare[11]. Normalization fields for the dynamic model spanned approximately 30° in azimuth and invariably excluded information from the center of the filter (Fig. 5e). Interestingly, dynamic models exploited normalization in both channels and switched normalization strategies during training, transitioning from purely static to purely context-dependent compression (Fig. S7d,e). Overall, normalized networks acquired representations that matched filtering and gain control properties of the fly medulla.

When tested on previous experimental stimuli (Fig. 1), linear models exhibited improved velocity tuning curves compared to a standard EMD (Fig. 1f and Fig. 5f, Fig. S5f) but estimates still varied substantially across time. Dynamic models, on the other hand, proved extremely robust at extracting scene motion across time, images, and velocities (Fig. 5f,g). We compared average estimation error on a held-out test set and found both types of non-linear compression to vastly outperform the linear stage (Fig. 5h). The performance of static compression indicates that simple response saturation already enhances robustness to contrast fluctuations in natural scenes. However, fly-like context-sensitivity consistently decreased test error over the static non-linearity (error reduction 22.0–29.2%; bootstrapped 95% CI). Finally, we benchmarked generalization on a fully independent image set (Fig. S7g) where linear models failed catastrophically while both compressive stages retained performance. This was particularly pronounced when testing images with high dynamic range (see Methods). Critically, on all datasets dynamic compression resulted in substantial error reduction with respect to both linear transfer and static compression.

In summary, our work represents the first demonstration that divisive contrast normalization occurs in the fly visual system and offers a comprehensive look at non-linear response properties in a virtually complete motion vision circuit. Normalization has been described as a generic mechanism for removing higher-order correlations from natural signals[1, 26]. Here, we close the loop between neural mechanism and an ecologically critical behavior, the optomotor response, and demonstrate how contrast gain control renders motion detection resilient to challenges imposed by natural scene statistics. Our convolutional network solves the task of estimating velocity across diverse environments and at little computational cost, particularly compared to standard optic flow algorithms. Present findings may thus aid the design of low-power, low-latency machine vision systems suitable for autonomous vehicles[27]. Gain control in the Drosophila optic lobe bears striking resemblance to normalization in other modalities like the fly olfactory system[28] or mammalian auditory cortex[29] as well as processing in vertebrate visual areas from retina to V1[6]. The tuning of non-linear surround suppression in lateral geniculate nucleus, in particular, matches that of transient units in the fly medulla[30]. This provides further proof for evolutionary convergence on canonical algorithms[6].

Author Contributions. M.S.D., A.L., and A.B. jointly conceived the study. M.S.D. and A.L. designed all experiments. A.L. and L.B. conducted behavioral experiments. E.S. recorded electrophysiological responses. M.S.D., N.P., F.G.R., and A.S. performed calcium imaging. A.L. designed and analyzed the convolutional model. M.S.D. and A.L. analyzed data, performed modelling, and wrote the manuscript. All authors participated in editing the manuscript.

Author Information. The authors declare no competing interests. Correspondence and requests for materials should be addressed to A.L. (leonhardt@neuro.mpg.de) or M.S.D. (drews@neuro.mpg.de).

References

1. F. Rieke and M. E. Rudd, "The challenges natural images pose for visual adaptation," *Neuron*, vol. 64, no. 5, pp. 605–616, 2009.
2. W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annual Review of Psychology*, vol. 59, pp. 167–192, 2008.
3. A. S. Mauss, A. Vlasits, A. Borst, and M. Feller, "Visual circuits for direction selectivity," *Annual Review of Neuroscience*, vol. 40, pp. 211–230, July 2017.
4. A. Leonhardt, G. Ammer, M. Meier, E. Serbe, A. Bahl, and A. Borst, "Asymmetry of *Drosophila* ON and OFF motion detectors enhances real-world velocity estimation," *Nature Neuroscience*, vol. 19, pp. 706–715, May 2016.
5. A. D. Straw, T. Rainsford, and D. C. O'Carroll, "Contrast sensitivity of insect motion detectors to natural images," *Journal of Vision*, vol. 8, no. 3, pp. 32–32, 2008.
6. M. Carandini and D. J. Heeger, "Normalization as a canonical neural computation," *Nature Reviews Neuroscience*, vol. 13, pp. 51–62, Nov. 2011.
7. D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual neuroscience*, vol. 9, pp. 181–197, Aug. 1992.
8. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
9. M. S. Maisak, J. Haag, G. Ammer, E. Serbe, M. Meier, A. Leonhardt, T. Schilling, A. Bahl, G. M. Rubin, A. Nern, B. J. Dickson, D. F. Reiff, E. Hopp, and A. Borst, "A directional tuning map of *Drosophila* elementary motion detectors," *Nature*, vol. 500, pp. 212–216, Aug. 2013.
10. A. Bahl, E. Serbe, M. Meier, G. Ammer, and A. Borst, "Neural mechanisms for *Drosophila* contrast vision," *Neuron*, vol. 88, no. 6, pp. 1240–1252, 2015.
11. A. Borst, C. Reisenman, and J. Haag, "Adaptation of response transients in fly motion vision. II: Model studies," *Vision Research*, vol. 43, pp. 1311–1324, May 2003.
12. A. Borst, V. L. Flanagin, and H. Sompolinsky, "Adaptation without parameter change: dynamic gain control in motion detection," *Proceedings of the National Academy of Sciences* , vol. 102, pp. 6172–6176, Apr. 2005.
13. D. A. Clark, L. Bursztyn, M. A. Horowitz, M. J. Schnitzer, and T. R. Clandinin, "Defining the computational structure of the motion detector in *Drosophila*," *Neuron*, vol. 70, pp. 1165–1177, June 2011.
14. J. E. Fitzgerald and D. A. Clark, "Nonlinear circuits for naturalistic visual motion estimation," *Elife*, vol. 4, p. e09123, Oct. 2015.
15. R. O. Dror, D. C. O'Carroll, and S. B. Laughlin, "Accuracy of velocity estimation by reichardt correlators," *Journal of the Optical Society of America. A, Optics and Image Science*, vol. 18, pp. 241–252, Feb. 2001.
16. R. S. A. Brinkworth and D. C. O'Carroll, "Robust models for optic flow coding in natural scenes inspired by insect biology," *PLoS computational biology*, vol. 5, p. e1000555, Nov. 2009.
17. R. A. Harris, D. C. O'Carroll, and S. B. Laughlin, "Contrast gain reduction in fly motion adaptation," *Neuron*, vol. 28, pp. 595–606, Nov. 2000.
18. W. Reichardt, T. Poggio, and K. Hausen, "Figure-ground discrimination by relative movement in the visual system of the fly," *Biological Cybernetics*, vol. 46, pp. 1–30, Jan. 1983.
19. J. H. van Hateren, "Processing of natural time series of intensities by the visual system of the blowfly," *Vision Research*, vol. 37, pp. 3407–3416, Dec. 1997.
20. M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive coding: a fresh view of inhibition in the retina," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 216, pp. 427–459, Nov. 1982.
21. K. Shinomiya, G. Huang, Z. Lu, T. Parag, C. S. Xu, R. Aniceto, N. Ansari, N. Cheatham, S. Lauchie, E. Neace, O. Ogundeyi, C. Ordish, D. Peel, A. Shinomiya, C. Smith, S. Takemura, I. Talebi, P. K. Rivlin, A. Nern, L. K. Scheffer, S. M. Plaza, and I. A. Meinertzhagen, "Comparisons between the ON- and OFF-edge motion pathways in the *Drosophila* brain," *eLife*, vol. 8, p. e40025, 2019.

6

22. A. Arenz, M. S. Drews, F. G. Richter, G. Ammer, and A. Borst, "The temporal tuning of the *Drosophila* motion detectors is determined by the dynamics of their input elements," *Current Biology*, vol. 27, pp. 929–944, Apr. 2017.

23. M. Silies, D. M. Gohl, Y. E. Fisher, L. Freifeld, D. A. Clark, and T. R. Clandinin, "Modular use of peripheral input channels tunes motion-detecting circuitry," *Neuron*, vol. 79, pp. 111–127, July 2013.

24. L. Freifeld, D. A. Clark, M. J. Schnitzer, M. A. Horowitz, and T. R. Clandinin, "GABAergic lateral interactions tune the early stages of visual processing in *Drosophila*," *Neuron*, vol. 78, pp. 1075–1089, June 2013.

25. D. L. Yamins and J. J. DiCarlo, "Using goal-driven deep learning models to understand sensory cortex," *Nature neuroscience*, vol. 19, no. 3, p. 356, 2016.

26. O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," *Nature Neuroscience*, vol. 4, pp. 819–825, Aug. 2001.

27. J. Plett, A. Bahl, M. Buss, K. Kühnlenz, and A. Borst, "Bio-inspired visual ego-rotation sensor for mavs," *Biological cybernetics*, vol. 106, no. 1, pp. 51–63, 2012.

28. S. R. Olsen, V. Bhandawat, and R. I. Wilson, "Divisive normalization in olfactory population codes," *Neuron*, vol. 66, pp. 287–299, Apr. 2010.

29. N. C. Rabinowitz, B. D. B. Willmore, J. W. H. Schnupp, and A. J. King, "Contrast gain control in auditory cortex," *Neuron*, vol. 70, pp. 1178–1191, June 2011.

30. V. Bonin, V. Mante, and M. Carandini, "The suppressive field of neurons in lateral geniculate nucleus," *Journal of Neuroscience*, vol. 25, pp. 10844–10856, Nov. 2005.
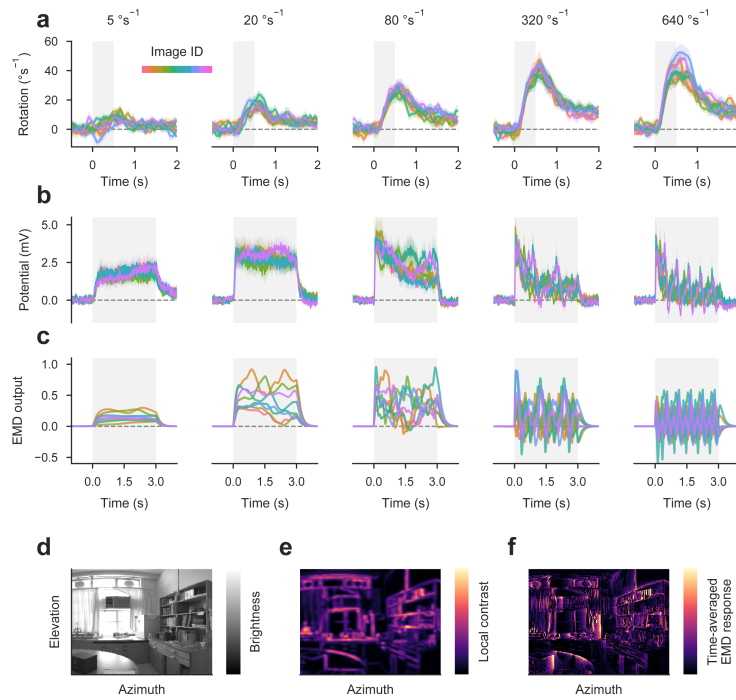
Figure S1. Behavioral, neural, and model responses to natural scenes (related to Fig. 1). a, Turning responses for 8 images (indicated by trace color) and 5 velocities (indicated by panel title; N=16 wild-type flies; data as in Fig. 1b). Gray shaded area indicates duration of motion stimulus. **a,** Membrane potential for 6 images and 5 velocities (N=11 HS cells from 9 flies; data as in Fig. 1d). **c,** Output of model LPTC for same images and velocities as b (data as in Fig. 1f). **d,** Natural image patch as seen through the field of view of model LPTC. **e,** Estimate of local contrast in natural image patch. RMS contrast was estimated by filtering the image with a Gaussian (σ = 0.5°), subtracting the filtered image from the original, squaring the mean-subtracted image, filtering it with a Gaussian (σ = 0.5°), and taking the square root. **f,** Spatially reconstructed output of simulated LPTC for same image patch as before, plotted as the square root of the time-averaged response. A horizontally motion-sensitive LPTC was constructed using the same parameters as in c (see Methods) with the exception of more fine-grained sampling at exactly the image resolution. The depicted panorama was moved for 16s at a velocity of 22.5° s^{-1}, resulting in a single complete rotation. Responses at each pixel location were then averaged across the full stimulus period. This demonstrates that the response of the EMD array depends strongly on squared local image contrast.
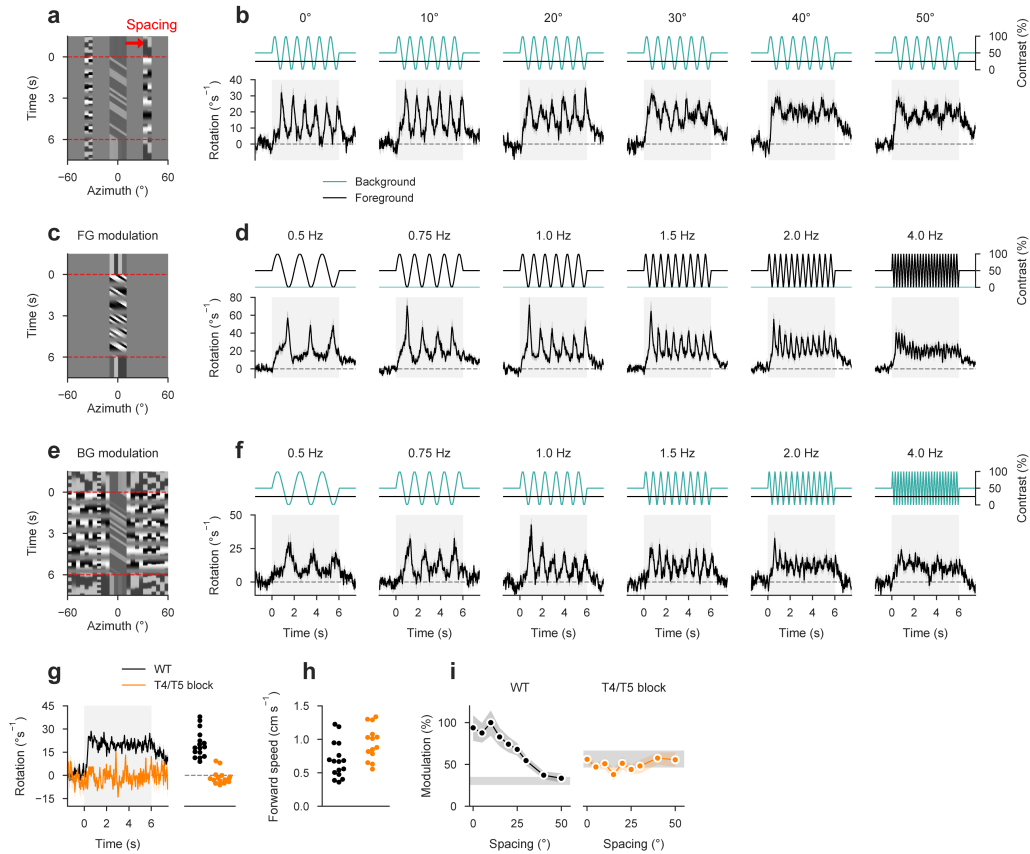
Figure S2. Detailed behavioral responses to contrast stimuli (related to Fig. 2). a, Illustration of spatial oscillation experiment. Background was restricted to 10° wide stripes flanking the foreground motion stimulus at the spacing indicated by red arrow. Dashed lines indicate period during which foreground pattern moved at $50°\,s^{-1}$. This arrangement was repeated at plus and minus 90° from the frontal axis of the fly; 0° in this plot indicates the center of the foreground. **b,** Contrast traces and turning responses for five spacing conditions (indicated above each panel). Top, instantaneous contrast (25% in foreground, oscillating at 1 Hz in background). Bottom, turning response of the fly (N=16 wild-type flies). Modulation was reduced as spacing between foreground and background increased. **c,** Illustration of temporal foreground modulation stimulus at 1 Hz frequency. **d,** Contrast traces and turning responses for five foreground oscillation frequencies (N=13; background contrast was 0%). Modulation decreased as frequency increased. **e,** Illustration of temporal background modulation stimulus at 1 Hz frequency. **f,** Contrast traces and turning responses for five background oscillation frequencies (N=13; foreground contrast was 25%). Modulation again decreased with frequency. **g,** Left, comparison of turning responses between wild-type flies and flies in which T4/T5 cells were silenced using TNT (see Methods; N=16/14 for WT/block flies). Right, turning responses averaged between 0 and 6s following motion onset. Syndirectional turning was abolished in T4/T5-silenced flies. **h,** Average forward speed throughout full experiment. T4/T5 block flies did not exhibit locomotion deficiencies. **i,** Comparison of spatial oscillation tuning. T4/T5 block flies did not show modulation at the contrast oscillation frequency of 1 Hz and a generally increased level of baseline fluctuation.

Figure S3. Lamina and T4/T5 receptive field mapping. a–e, Averaged 2D spatial receptive fields (RF) of L1–L5 from reverse correlation using white noise stimulation (L1: 21/7 cells/flies, L2: 34/5, L3: 34/5, L4: 17/6, L5: 18/9). **f–j,** 1D projection (averaged over all orientations) of the RFs in a–e. All cell types possessed linear RFs with antagonistic center-surround structure. **k–o,** Temporal RFs measured in the center of the spatial RFs. **p–t,** Frequency-space representations of temporal RFs. **u,** Frequency representations of lamina transient/band-pass cells (all lamina cells except for L3) after deconvolution with a putative linear GCaMP6f low-pass filter with time constant 350ms as performed previously[22]. **v,** Deconvolved frequency responses of medulla bandpass filter cells (replotted from previous work[22]). **w,** Spatial integral of the 2D RFs in a–e. For L3, the strong antagonistic ON surround exactly counterbalanced the OFF center contribution. **x,** x-y plot of the stochastic motion noise stimulus used for localizing T4/T5 RFs. **y,** Example RF of a T4 cell from reverse correlation with the motion noise stimulus. **z,** Average responses of T4/T5 to 25° windowed drifting gratings probing different positions around the estimated RF center. This validated the RF coordinates obtained from the stochastic motion noise stimulus. All data are shown as mean ± s.d.

Figure S4. Raw calcium responses for basic contrast stimuli (related to Fig. 3). a, Shown is only a subset of the data evaluated in Fig. 3. Background contrast of 0% is indicated by black lines, background contrast of 100% is depicted in magenta. Responses are shown only for 3 out of 7 foreground contrasts. **b–p,** Average calcium responses of all neurons to combinations of different foreground and background contrasts.

11
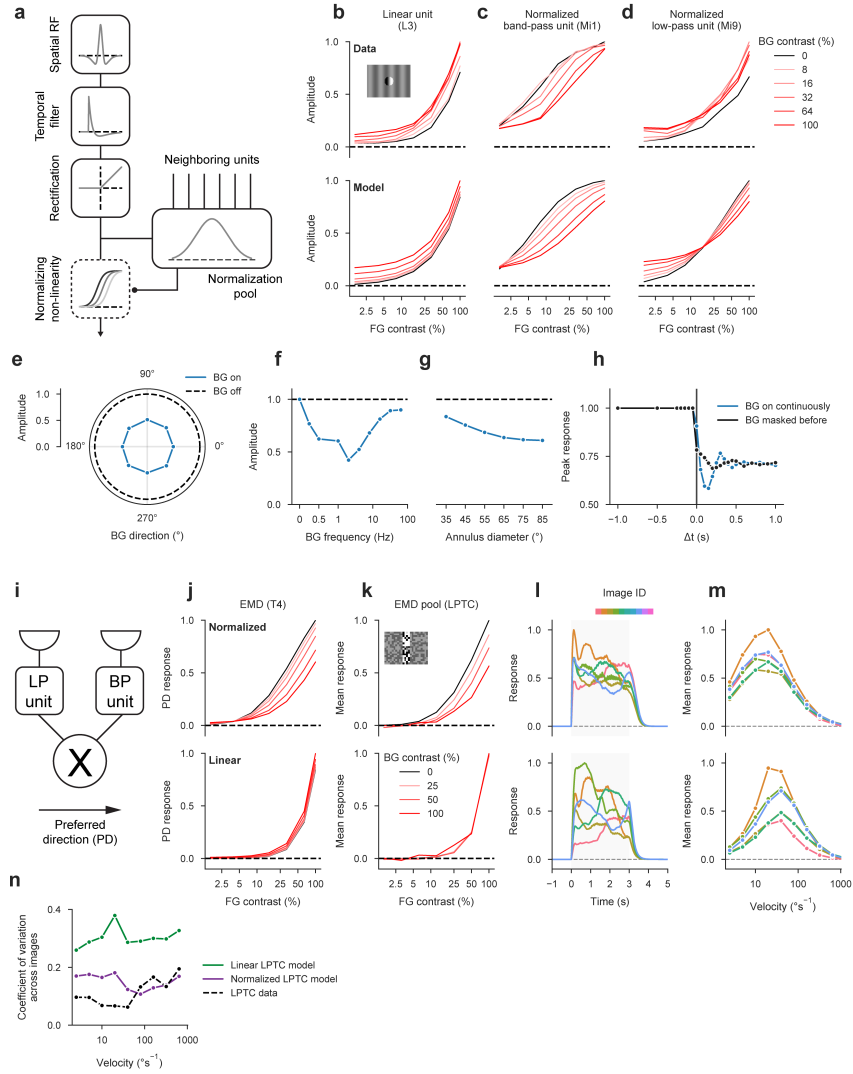
Figure S5. Data-driven functional model of normalization circuit. a, Illustration of signal cascade for data-driven cell model (see Methods for details). Filter elements are sketched for an ON band-pass cell with normalization. **b–d,** Contrast tuning curves for three model cells, estimated using the same protocol as during calcium imaging (FG = foreground, BG = background). Top, empirical data for L3, Mi1, and Mi9 (see Fig. 3). Inset depicts a single frame from stimulus centered on recorded cell with background contrast 25% and foreground contrast 100%. Bottom, tuning curves from models manually tuned to resemble their empirical counterparts (see Methods for parameters). **e,** Responses of normalized ON band-pass cell model to orientation tuning stimulus (see Fig. 4a; dashed line marks reference stimulus without background). Stimuli and evaluation were exactly matched to the experiment. **f,** Responses of the same model to background frequency tuning experiment (see Fig. 4b; dashed line marks reference stimulus without background). **g,** Responses of the same model to background size stimulus (see Fig. 4c; dashed line marks reference stimulus without background). **h,** Responses of the same model to contrast-step protocol (see Fig. 4d). **i,** Illustration of T4 or T5 model. Signals from a strongly normalized band-pass and a weakly normalized low-pass unit covering adjacent areas of the visual field are multiplied, yielding a direction-selective signal. **j–m,** Top, responses from motion detector models with normalization. Bottom, responses from motion detector models in which normalization was switched off for both input arms. **j,** Foreground contrast tuning for simulated T4 cell (see Fig. 3). **k,** Responses to behavioral contrast stimulus for a LPTC model composed of T4 and T5 models (see Methods). **l,** Responses to various natural scenes moving at $20°\,s^{-1}$ (modelled and evaluated as in Fig. 1). **m,** Velocity tuning curves for natural scenes (modelled and evaluated as in Fig. 1). **n,** Coefficient of variation across images for individual image velocities (derived from velocity tuning curves in m and Fig. 1f; see Methods for details). A model including input normalization outperformed the linear model and approximated the variability of LPTC responses.

12
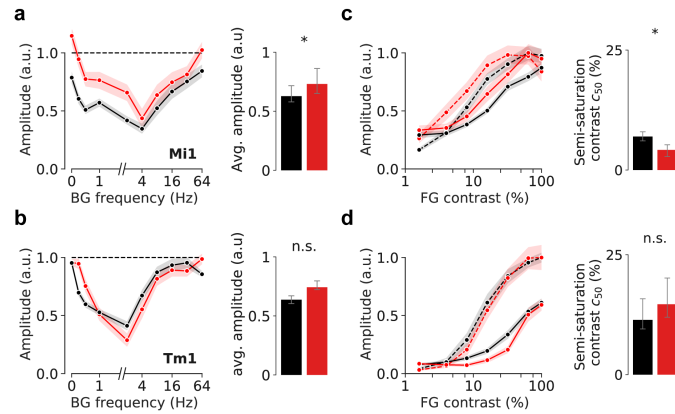
Figure S6. Additional silencing data for contrast tuning experiment (related to Fig. 4). a, Left, frequency tuning of suppression as in Fig. 4b,g,h. Black dashed line represents reference stimulus with background contrast of 0%. Right, average amplitude over all frequencies was higher for Mi1 block flies which indicates reduced suppression (Mi1 block: 20/5 cells/flies, Mi1 control: 21/6; Mann-Whitney U: 143, *P=0.04). **b,** Tm1 data as in a. There was no significant effect when blocking Tm1 cells (Tm1 block: 20/5, Tm1 control: 19/5; Mann-Whitney U: 169, NS P=0.28). **c,** Left, foreground contrast tuning for 0 (dashed line) and 100% (solid line) background contrast as in Fig. 4i,j. Right, contrast sensitivity as measured by semi-saturation constant c_{50} of model fit to data was increased for Mi1 block flies (Mann-Whitney U: 128, *P=0.02). **d,** Tm1 data as in c. There was no significant effect when blocking Tm1 cells (Mann-Whitney U: 158, NS P=0.19). Error bars show bootstrapped 68% confidence intervals around the median.

13
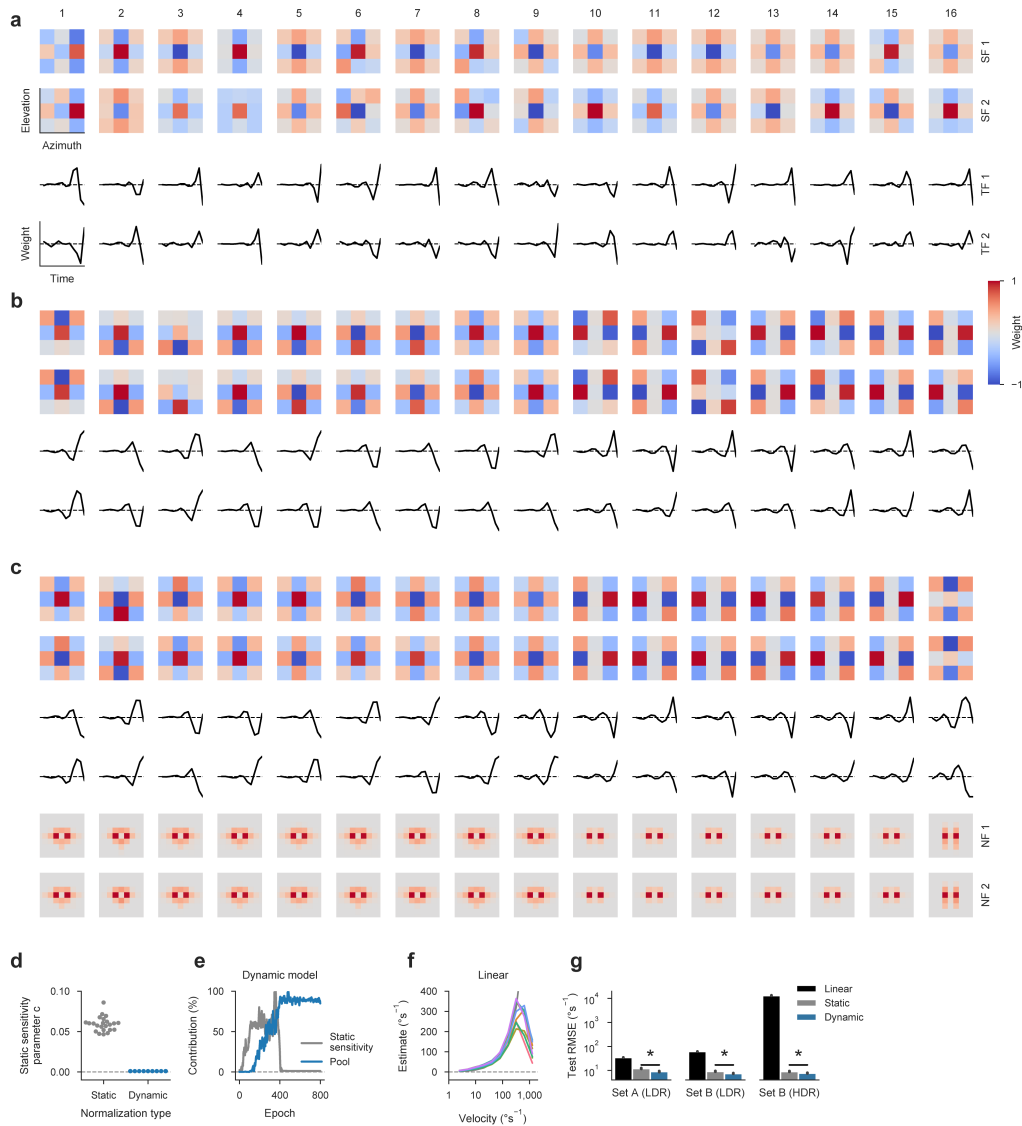
Figure S7. Detailed receptive fields and performance data for task-driven model (related to Fig. 5). a–c, Receptive fields and temporal filters for 16 models of each non-linearity configuration (a, linear; b, static; c, dynamic). Models were sorted by test set error (increasing from left to right). Each pair of spatial and temporal filters was normalized to the maximum absolute weight across both channels (SF = spatial filter, TF = temporal filter, NF = normalization filter). Axis limits are the same as in Fig. 5. **d,** Values of sensitivity parameter c for all static (N=23) and dynamic (N=16) normalization models. **e,** Evolution of weights for a single dynamic model. Both curves were independently normalized to their maximum across epochs. Pool contribution was quantified as the sum of weights across both 11 x 11 x 1 normalization filters. **f,** Velocity tuning curves of best-performing linear model for various images (analogously to Fig. 5g). Gray curve indicates true scene velocity on logarithmic axis. **g,** Quantification of average model performance for all tested data sets (analogously to Fig. 5h; LDR = low dynamic range, HDR = high dynamic range). See Methods for details on how data sets were generated. Note that performance is plotted on a logarithmic axis. N=22/23/16 for linear/static/dynamic; *P<0.001; t=9.01/7.51/7.72 for set A/set B (LDR)/set B (HDR); Student's t-test with assumed equal variance; only difference between static and dynamic was tested.

14

Short name	Full genotype	Used in
WT	w+/w+; +/+; +/+	Fig. 1, Fig. 2
T4/T5 block	w+/w-; R59E08-AD/UAS-TNT; R42F06-DBD/+	Fig. S2
L1-GCaMP6f	w+/w-; VT027316-AD/UAS-GCaMP6f; R40F12-DBD/UAS-GCaMP6f	Fig. 3, Fig. S3, Fig. S4
L2-GCaMP6f	w+/w-; R53G02-AD/UAS-GCaMP6f; R29G11-DBD/UAS-GCaMP6f	Fig. 3, Fig. S3, Fig. S4
L3-GCaMP6f	w+/w-; R59A05-AD/UAS-GCaMP6f; R57H07-DBD/UAS-GCaMP6f	Fig. 3, Fig. S3, Fig. S4
L4-GCaMP6f	w+/w-; R20A03-AD/UAS-GCaMP6f; R31C06-DBD/UAS-GCaMP6f	Fig. 3, Fig. S3, Fig. S4
L5-GCaMP6f	w+/w-; R21A05-AD/UAS-GCaMP6f; R31H09-DBD/UAS-GCaMP6f	Fig. 3, Fig. S3, Fig. S4
Mi1-GCaMP6f	w+/w-; R19F01-AD/UAS-GCaMP6f; R71D01-DBD/UAS-GCaMP6f	Fig. 3, Fig. S4, Fig. S4
Mi4-GCaMP6f	w+/w-; R48A07-AD/UAS-GCaMP6f; R13F11-DBD/UAS-GCaMP6f	Fig. 3, Fig. S4
Mi9-GCaMP6f	w+/w-; R48A07-AD/UAS-GCaMP6f; VT046779-DBD/UAS-GCaMP6f	Fig. 3, Fig. S4
Tm1-GCaMP6f	w+/w-; R41G07-AD/UAS-GCaMP6f; R47G01-DBD/UAS-GCaMP6f	Fig. 3, Fig. S4, Fig. S4
Tm2-GCaMP6f	w+/w-; +/UAS-GCaMP6f; VT12282/UAS-GCaMP6f	Fig. 3, Fig. 4a-c, Fig. S4
Tm2split-GCaMP6f	w+/w-; R28D05-AD/UAS-GCaMP6f; R82F12-DBD/UAS-GCaMP6f	Fig. 4d
Tm4-GCaMP6f	w+/w-; +/UAS-GCaMP6f; R35H01/UAS-GCaMP6f	Fig. 3, Fig. S4
Tm9-GCaMP6f	w+/w-; +/UAS-GCaMP6f; VT65303/UAS-GCaMP6f	Fig. 3, Fig. S4
T4-GCaMP6f	w+/w-; VT16255-AD/UAS-GCaMP6f; VT12314-DBD/UAS-GCaMP6f	Fig. 3, Fig. S3, Fig. S4
T5-GCaMP6f	w+/w-; VT13975-AD/UAS-GCaMP6f; R42H07-DBD/UAS-GCaMP6f	Fig. 3, Fig. S3, Fig. S4
Mi1-GCaMP6f, TNT-E	w+/w-; R19F01-AD/UAS-TNT-E; R71D01-DBD/UAS-GCaMP6f	Fig. S6
Mi1-GCaMP6f, TNTin	w+/w-; R19F01-AD/UAS-TNTin; R71D01-DBD/UAS-GCaMP6f	Fig. S6
Tm3-GCaMP6f, TNT-E	w+/w-; R13E12-AD/UAS-TNT-E; R59C10-DBD/UAS-GCaMP6f	Fig. 4
Tm3-GCaMP6f, TNTin	w+/w-; R13E12-AD/UAS-TNTin; R59C10-DBD/UAS-GCaMP6f	Fig. 4
Tm1-GCaMP6f, TNT-E	w+/w-; R41G07-AD/UAS-TNT-E; R47G01-DBD/UAS-GCaMP6f	Fig. S6
Tm1-GCaMP6f, TNTin	w+/w-; R41G07-AD/UAS-TNTin; R47G01-DBD/UAS-GCaMP6f	Fig. S6
Tm2split-GCaMP6f, TNT-E	w+/w-; R28D05-AD/UAS-TNT-E; R82F12-DBD/UAS-GCaMP6f	Fig. 4
Tm2split-GCaMP6f, TNTin	w+/w-; R28D05-AD/UAS-TNTin; R82F12-DBD/UAS-GCaMP6f	Fig. 4

Table S1. Genotypes and abbreviations

15

Cell type	L_{fg}	L_{bg}	p	c_{50}	w_{pool}	q	Norm. index	$R^2_{DivisiveNorm}$	R^2_{linear}
L1	1.47	0.07	1.10	0.53	0.22	0.97	0.42	98.39 ± 0.10	92.55 ± 0.14
L2	1.10	0.05	1.37	0.23	0.36	0.77	1.58	99.29 ± 0.03	85.17 ± 0.13
L3	1.68	0.16	1.46	1.00	0.00	1.27	0.00	95.90 ± 0.08	97.17 ± 0.07
L4	1.41	0.12	1.23	0.53	0.32	1.09	0.61	98.94 ± 0.04	93.71 ± 0.07
L5	1.04	0.05	1.29	0.14	0.19	1.10	1.36	94.51 ± 0.23	69.34 ± 0.24
Mi1	1.03	0.03	1.21	0.06	0.25	1.05	4.33	97.37 ± 0.14	56.26 ± 0.41
Mi4	1.61	0.33	0.90	1.00	0.31	5.92	0.31	90.08 ± 0.26	87.50 ± 0.35
Mi9	1.69	0.23	0.99	1.00	0.40	2.87	0.40	92.40 ± 0.24	89.61 ± 0.32
T4	0.96	0.01	2.47	0.11	0.49	0.74	4.45	96.78 ± 0.15	74.17 ± 0.35
T5	1.08	0.07	1.97	0.26	1.17	0.92	4.55	97.02 ± 0.13	77.27 ± 0.27
Tm1	0.98	0.09	1.87	0.18	0.86	0.71	4.75	97.53 ± 0.11	78.67 ± 0.29
Tm2	1.08	0.17	1.36	0.20	1.14	0.91	5.76	97.58 ± 0.08	73.09 ± 0.32
Tm3	1.02	0.01	1.97	0.16	0.53	0.72	3.39	97.97 ± 0.12	82.33 ± 0.20
Tm4	1.06	0.11	2.33	0.40	1.44	0.81	3.61	96.77 ± 0.16	76.96 ± 0.37
Tm9	1.83	0.50	0.92	0.98	1.01	1.65	1.03	96.37 ± 0.14	87.42 ± 0.25

Table S2. Fits for divisive normalization model (related to Fig. 3)

16

Methods

Flies and genetics. Drosophila melanogaster were kept on a 12 h light/12 h dark cycle at 25°C and 60% humidity on standard cornmeal-agar medium. Genetic expression of effectors was targeted through the Gal4-UAS system[31]. Resulting genotypes and their shorthands are listed in Table S1.

Unless stated otherwise, locomotion and tangential cell responses were recorded in wild-type Canton S flies 1 to 5 days after eclosion (Fig. 1 and Fig. 2). We used the genetically encoded calcium indicator GCaMP6f[32] to determine the functional properties of individual cell types (Fig. 3 and Fig. 4). Throughout silencing experiments (Fig. 4, Fig. S2 and Fig. S5), we expressed tetanus toxin light chain (TNT) or an inactive version (TNTin) in the cell type of interest[33]. For calcium imaging experiments involving silencing (Fig. 4 and Fig. S6), one day old flies were collected and put on 29°C for 3 days to boost expression of TNT or TNTin.

Natural image sets. For electrophysiology, behavioral, and modelling experiments, we used images from a published set of 20 natural panoramic scenes[16] termed data set A. All images were independently processed as follows: We averaged across color channels and downsampled the scene to a resolution of 1,600 x 320 pixels (covering 360° sampled at 0.225 pixels per degree along the azimuth) using linear interpolation. To be able to render 12 bit images on conventional screens with 8 bits of dynamic range, we first performed standard gamma correction by raising raw pixel values to a power of 0.45 and then clipped the top percent of pixel intensities. The resulting image was scaled to fill the range between 0 and 255. For optomotor experiments (Fig. 1), we selected a subset of 8 images that covered different types of terrain. From this set, we again selected a subset of 6 images to determine tangential cell responses. We used all 20 images to build the convolutional network (Fig. 5), randomly assigning 15 scenes to the training and 5 to the test set. Finally, we validated the trained convolutional model with images from an independent panoramic scene collection[34] consisting of 421 images (Fig. S7f). These scenes were kept at their native resolution of 927 x 251 pixels (corresponding to an azimuthal sampling rate of 0.39 pixels per degree) and processed as above, yielding data set B. We then generated two test sets: One had gamma correction applied to limit the images' bit depth ("low dynamic range" or LDR) and the other one was left at 12 bit depth to produce a data set with high dynamic range (HDR).

Behavioral experiments.

Locomotion recording. Experiments on the treadmill setup were conducted as described before[4, 10, 35]. Briefly, we tethered flies to a thin metal rod and placed them on air-cushioned polyurethane balls whose movement was tracked at 4 kHz, allowing for direct read-out of rotational motion along all three axes. Temperature within the vicinity of the fly was 25°C at the start of each experiment. Using a closed-loop thermoregulation system, we linearly increased it to 34°C within 15 min to encourage locomotion.

Visual stimuli. We used three identically calibrated computer screens that were placed in a rectangle surrounding the fly. To simulate a cylindrical display, all stimuli were rendered onto a virtual cylinder and distorted accordingly before projection onto screens. Our setup covered approximately 270° in azimuth and 120° in elevation of the visual field. All stimuli were displayed at 144 Hz and at a spatial resolution greatly exceeding that of the fly eye. Screens had a maximum luminance of approximately 100 cd m-2 and a luminance depth of 8 bit; for all descriptions below, we assume pixel brightness to range from 0 to a maximum of 1. Patterns were generated in real-time and programmed in Python 2.7 using the game engine Panda3D.

We measured velocity tuning curves (Fig. 1) for 8 distinct natural images at 6 logarithmically spaced velocities ranging from 5 to $1,280° s^{-1}$. Initial image phase was randomized on each trial. Scenes were displayed at their native gamma-corrected mean luminance and contrast (see above). On each trial images stood still for 1.5s, then were rotated at the chosen velocity for 0.5s, and remained fixed for another 1.5s.

The optomotor contrast stimulus separated the visual field into two areas (see Fig. 2a and Fig. S2). For the so-called background, we tiled the visual field with pixels of size 5° x 5°. At each pixel location we drew a temporal frequency f from a normal distribution ($\mu = 0$ Hz, $\sigma = 1$ Hz) and a starting phase λ from a uniform distribution covering 0 to 360°. Instantaneous luminance of each pixel i was then determined by a random sinusoid of the form

$$I_i(t) = 0.5 + 0.5 \cdot c_{bg} \cdot g(sin(2\pi f_i t + \lambda_i))$$

where the experimental parameter c_{bg} runs from 0 to 100% and controls the effective contrast of the background. To increase average contrast in the visual field, we applied the compressive transform

$$g(x) = \sqrt{\frac{1+\alpha^2}{1+\alpha^2 x^2}}$$

where $\alpha = 5$ determined the degree of curve flattening. Using this method, we generated stochastic and dynamic visual input at a controllable contrast level without introducing coherent motion (see Fig. 2b).

17

The so-called foreground delivered a coherent motion stimulus driving the optomotor response. It consisted of two vertical stripes that were placed at plus and minus $90°$ from the frontal axis of the fly, each spanning $20°$ in azimuth and the full screen elevation. We again tiled each stripe with pixels covering an area of approximately $5°$ x $5°$. For each pixel i, luminance was fixed over time and determined by

$$I_i(t) = 0.5 + 0.5 \cdot c_{fg} \cdot g(sin(\lambda_i))$$

where the experimental parameter c_{fg} controls the effective motion contrast and λ was independently drawn from a uniform distribution covering 0 to $360°$. The pixelated noise pattern smoothly wrapped around the azimuthal borders when moving. Note that for all instantiations of the stimulus, mean luminance across the visual field was 0.5.

For the basic contrast tuning experiment (Fig. 2a–e; see Supplementary Video 1), we exhaustively measured combinations of logarithmically spaced values for c_{fg} (1.6, 3.1, 6.3, 12.5, 25, 50, and 100%) and c_{bg} (0, 25, 50, and 100%). At the beginning of each trial we simultaneously presented the dynamic background and the static foreground pattern. Between 1.5 and 2.0s following stimulus onset, the foreground pattern moved at a fixed velocity of $50°s^{-1}$. For oscillation experiments (Fig. 2f–h), the motion period was extended to 6 s. While the foreground pattern was moving, we sinusoidally modulated the contrast of either fore- or background between 0 and 100% around a mean value of 50% and at the specified temporal frequency (see Fig. 2f, Fig. S2a,c,e, and Supplementary Video 2). When mapping the spatial extent of the contrast-induced modulation, we set the modulation frequency to 1 Hz and restricted the background pattern to two stripes of $10°$ width flanking each foreground pattern (see Fig. S2a and Supplementary Video 3). The spacing (0, 5, 10, 15, 20, 25, 30, 40, or $50°$) determined the distance between foreground and background. In this experiment, we additionally measured a zero-contrast background condition to obtain an appropriate modulation baseline. Here, the motion stimulus had a contrast of 25% and luminance in the rest of the field was set to a uniform 0.5. For the temporal experiments, we measured oscillation frequencies of 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, 8, and 10 Hz (Fig. 2h). Background contrast was zero when measuring foreground tuning; for background tuning, foreground contrast was set to 25%.

All stimulus patterns were displayed twice throughout optomotor experiments, once in clockwise and once in counterclockwise direction of motion. We recorded multiple trials to obtain robust turning responses for each fly (15 trials for natural image stimuli, 20 for contrast tuning, 25 for oscillation stimuli). Presentation order was shuffled across conditions within any trial to mitigate adaptation effects. Individual experiments lasted between 60 and 120 min.

Data evaluation. To ensure data quality, we excluded all flies whose average forward velocity during the experiment was below $0.25\,\text{cm}s^{-1}$ and whose average turning tendency was either slowly drifting or far from $0°s^{-1}$. Fewer than 20% of all experiments failed these criteria. Measurements of ball movement were downsampled via linear interpolation for further processing (to 50 Hz for natural image stimuli, Fig. 1; 20 Hz for contrast tuning, Fig. 2; 100 Hz for oscillation stimuli, Fig. 2). Trials were averaged.

Responses for clockwise and counterclockwise motion were subtracted and divided by two to minimize residual deviations from straight forward walking. Traces for natural image and contrast tuning stimuli were filtered using a first-order low-pass with a time constant of 100ms. For the contrast oscillation experiments, we evaluated modulation at the relevant carrier frequency by calculating the zero-padded Fourier Transform of the turning trace and averaging the amplitude spectrum in a window of width 0.2 Hz centered on the target frequency. These values were normalized per experiment such that the modulation peak after averaging was 100%. We applied a Savitzky-Golay filter (window length 11 samples, 5th order polynomial) before plotting traces from oscillation experiments; this did not affect analysis.

All analysis for behavioral experiments was performed in custom-written software using Python 3.6, NumPy 1.15, and SciPy 1.1.

Electrophysiology.

Procedure. Our patch-clamp recordings from tangential cells followed established protocols[36]. Cell bodies of horizontal system (HS) units were targeted visually through a microscope. We confirmed their preferred direction by stimulation with oriented moving sine wave gratings before each experiment.

Visual stimulus. Visual stimulation was delivered using a cylindrical projector-based arena as previously described[22]. Briefly, the screen of the arena covered a viewing angle of the fly of $180°$ in azimuth and $105°$ in elevation. Stimuli were generated at a framerate of 180 Hz using green light spanning approximately 500 nm to 600 nm in wavelength. The maximum luminance this arena achieved was $276\pm48\text{cd m}^{-2}$ (mean±s.d. across devices). All visual stimuli were rendered using custom software written in Python 2.7 and the Panda3D framework. Membrane potential was recorded using custom software written in MATLAB (MathWorks, MA).

We measured tuning curves for 6 distinct natural image panoramas at 9 logarithmically spaced velocities ranging from 2.5 to $640°s^{-1}$ (Fig. 1). On each presentation, the scene was displayed at a fixed phase, stayed still for 1 s, and then rotated horizontally for 3s at the chosen constant velocity. Image movement was always in the preferred direction of the HS unit.

18

We showed images at their native gamma-corrected mean luminance and contrast (see above). Each condition was repeated 5 times. Conditions and trials were randomly interleaved to exclude adaptation effects along any stimulus dimension.

Data evaluation. Voltage data were digitized at 1,000 Hz. To account for slow drift in potential, we subtracted the average voltage in a 1s window before stimulus onset from each trace per stimulus condition and trial. Signals were then low-pass filtered (8th order Chebyshev Type 1) and resampled at 100 Hz. Finally, we averaged cell responses across trials. Cells whose mean depolarization during full-contrast sine grating presentation in preferred direction remained below 5mV were discarded before further analysis. All analysis for electrophysiological experiments was performed in custom-written software using Python 3.6, NumPy 1.15, and SciPy 1.1.

Calcium Imaging.

Procedure. Calcium imaging experiments were performed using custom-built two-photon microscopes as described before[22]. The imaging acquisition rate was 11.8 Hz for all experiments, or 23.7 Hz for the experiment in Fig. 4d, with imaging resolutions ranging from 32 x 32 to 64 x 128 pixels. Image acquisition was controlled using the ScanImage software (version 3.8)[37]. We prepared flies as previously described[9, 22]. Briefly, Drosophila were anesthetized on ice and glued onto a plexiglas holder with the back of their head exposed to a perfusion chamber filled with Ringer's solution. Then the cuticula was surgically opened to allow optical access.

Visual stimuli. Stimuli were presented using the same projector system as in electrophysiological experiments, with additional long-pass filters (cut-off wavelength of 550 nm) in front of projectors to spectrally separate visual stimulation from GCaMP fluorescence signals.
To identify receptive field (RF) positions of individual neurons, white noise stimuli of 3 min length were used (except for T4/T5 cells, see below). The stimuli were pre-rendered at 60 Hz and generated as previously described[22]. Briefly, the spatial resolution of all white noise stimuli was 2.8° of visual angle corresponding to 64 pixels across the 180° screen. For all lamina cells, the same stimulus was used in order to provide a systematic description of their spatiotemporal filtering properties (Fig. S3). This stimulus had a Gaussian autocorrelation with a standard deviation of approximately 45ms in time and a contrast of 25% around a mean intensity value of 50 on an 8 bit grayscale. For some cell types, variants of this stimulus with higher contrast or longer time constants were used if necessary to reliably locate their RFs on the arena. Specifically, we mapped RFs for Tm4, Mi4, Mi9 and Tm9 with a binary stimulus at 100% contrast and a temporal cut-off frequency of 1 Hz. For Mi9, we chose a 1D version of this stimulus, consisting of horizontal (1.5 min) and vertical bars (1.5 min) instead of pixels.
For T4/T5, we relied on a novel stochastic motion noise stimulus to determine RF coordinates. First, we determined the preferred direction of a ROI using drifting gratings. Then we displayed a stimulus consisting of 20 randomly distributed 15° wide circular windows. Inside of each window, a 30° wavelength sine grating drifted at $30°s^{-1}$ in the preferred direction (Fig. S3x). The positions of these 20 windows were changed and randomly chosen every second over 4 min. Reverse correlation of T4/T5 responses with the area covered by those windows at a given time point yielded motion-sensitive RFs which were fit with a Gaussian to determine center coordinates (Fig. S3y). These were verified by presenting 25° windows containing full contrast drifting gratings at the estimated RF center and 6 hexagonally distributed positions around the center. Cells responded only to the grating in the RF center (Fig. S3z).
For the experiments shown in Fig. 3, a 25° circular window around the RF center of a cell defined the foreground whereas the rest of the screen was defined as background. A drifting sine grating with 30° wavelength and a velocity of $30°s^{-1}$ was shown, starting with medium gray at the center of the RF and moving for 4 seconds after stimulus onset (see Supplementary Video 4). The contrast of the grating was varied independently between background and foreground. A stimulus matrix of 7 foreground contrasts (1.6, 4, 8, 16, 32, 64 and 100%) and 6 background contrasts (0, 8, 16, 32, 64 and 100%) at a constant mean luminance of 0.5 was presented.
For the experiments shown in Fig. 4a–c (see Supplementary Videos 5–7), the foreground contrast was chosen depending on the cell type as the point where the suppression elicited by 100% background contrast (as measured in Fig. 3) would be greatest. This was 16% for Mi1, 32% for Tm1, 100% for Tm2 and 64% for Tm3. The background had 100% contrast and 30° wavelength. We varied either its direction, its velocity (0, 0.25, 0.5, 1, 2, 4, 8, 16, 32 or $64°s^{-1}$), or restricted its presentation to an annulus with changing outer diameter. A reference condition with 0% background contrast was added to the stimulus protocol.
For the contrast-step stimulus experiments shown in Fig. 4d (see Supplementary Videos 8–9), the background grating had 30° spatial wavelength, drifted with $90°s^{-1}$ after motion onset and its initial phase was randomized. For Tm2 it had full contrast, for Tm3 44% contrast. The 25° foreground window was 50% gray and we placed a 5° wide dot in the center. For Tm3, the dot was initially black and set to white for a duration of 50ms at a given time interval after motion onset of the background grating. For Tm2, the dot was initially white and then set to black. The time interval was varied in steps of 50ms from –250ms to 500ms and then in steps of 100ms. Negative values indicate that the surround grating started to move after the dot changed its intensity. Additional time intervals were –500ms and –1s. The block experiments in Fig. 4f–j and Fig. S6 were performed

19

with the same frequency tuning stimuli as before (Fig. 4b). For the contrast tunings, the same stimuli as in Fig. 3 were used but with background contrast of either 0 or 100% only.

All stimuli were repeated three times in randomized condition order to prevent adaptation to any stimulus features.

Data evaluation. Calcium imaging stacks were registered in order to correct for translational movement artifacts of brain tissue using custom-written software. Responses of individual neurons were extracted by manually selecting small regions of interests (ROI) encompassing individual anatomical structures. For T4 and T5 these corresponded to single or few axon terminals; for Mi and Tm cells, individual axon terminals could be identified clearly through visual inspection. For ON pathway medulla cells, signals were measured in layer 10 of the medulla, for OFF pathway medulla cells in layer 1 of the lobula. For lamina cells L1–5, signals were measured at axon terminals in corresponding layers 1–5 in the medulla. For T4/T5, signals were recorded in the lobula plate.

To reconstruct RFs, calcium signals were mean subtracted and reverse-correlated with the stimulus as previously described[22]. 1D Gaussians were fit to horizontal and vertical cross-sections of spatial receptive fields to obtain precise RF coordinates. For lamina cells (Fig. S3), all reconstructed RFs were peak-aligned and analysed as previously[22]. For 1D projections of spatial RFs (Fig. S3f–j), an average of 1D projections of 2D RFs along 3600 evenly distributed projection angles between $0°$ and $360s^{-1}$ was calculated. This enhanced the visibility of the center-surround structure but neglected possible anisotropies in the spatial structure of RFs[24]. For impulse responses (Fig. S3k–o) the temporal receptive field of the 9 center pixels was averaged; frequency responses (Fig. S3p–t) are the Fourier-transformed impulse responses. Deconvolution (Fig. S3u) was performed by dividing the frequency spectra with the frequency response of a 1st order low-pass filter with time-constant 350ms as a proxy for calcium indicator dynamics[22, 38].

Relative fluorescence changes ($\Delta F/F$) from raw calcium traces were obtained by adapting an automatic baseline detection algorithm[39]. Briefly, raw data were first smoothed with a Gaussian window (full-width at half maximum, $FWHM = 1s$). Then, minima within a 90s long sliding window were extracted and the resulting trace smoothed with a Gaussian window ($FWHM = 4min$). The result was used as a dynamic baseline $F0$ and $\Delta F/F$ values were computed as $\Delta F/F = (F - F0)/F0$.

For further evaluation only recordings with good signal-to-noise ratio (SNR) were taken. The criterion was that the standard deviation of the mean signal averaged over trials had to be at least 120% of the mean standard deviation over trials. This criterion filtered out cells with an inter-trial variance larger than the typical cell response (caused by movement artifacts or photobleaching). In addition, the standard deviation of the mean signal had to be larger than 25% $\Delta F/F$. On average, 90% of all cells measured passed these criteria with slight variations due to different levels of GCaMP expression depending on the genotype.

For experiments with drifting gratings, the driving foreground contrast frequency was 1 Hz. For these experiments, we evaluated the amplitude of the 1 Hz component of the signal. This was achieved by computing the Fourier coefficient at that frequency, using the equation

$$F = \left| \frac{1}{T} \int_0^T dt s(t) e^{-2\pi i \cdot 1Hz \cdot t} \right|$$

where $s(t)$ denotes the signal and T the stimulation time. For experiments in Fig. 4d, we evaluated the peak response of the calcium signal.

Amplitudes were averaged over trials and normalized to the maximum, then averaged over cells and normalized to the maximum. For Fig. 4a–c,g,i and Fig. S6a,b, amplitudes were normalized to the response amplitude for the reference stimulus.

Modeling.

Natural motion stimuli. To evaluate the performance of our models under naturalistic conditions, we generated a synthetic set of motion sequences that closely mimicked the experimental stimuli described above. For each sequence we translated $360°$ images at a fixed horizontal velocity through a virtual window spanning $100°$ in azimuth. Given their panoramic nature, scenes wrapped around seamlessly at each border. Movies were generated at a time resolution of 100 Hz. To reduce jitter for small velocities, we linearly interpolated non-integer pixel shifts. Fly eye optics were simulated ahead of time. We blurred each frame with a Gaussian filter (full width at half-maximum of $4°$) to approximate the acceptance angle of each photoreceptor[11] and then sampled individual signals from a rectangular grid with isotropic spacing of $4°$ (yielding 23 x 17 receptor signals per frame for data set A and 23 x 23 for data set B, as described above).

For the comparison in Fig. 1, we modelled the exact stimulus parameters of the electrophysiological experiment including an approximation of the image's starting phase on the arena. We generated sequences for our convolutional detector models (Fig. 5) as follows: The set of 20 panoramic images was randomly split into a training group consisting of 15 scenes and a test group consisting of 5 scenes. For each sequence, a random image was drawn from the appropriate set. The stimulus lasted 5 s. Between 1 and 4 s, scene velocity stepped from zero to a fixed value drawn from a Gaussian distribution with $s.d. = 100°s^{-1}$. The initial window phase followed a uniform distribution spanning $360°$. To further augment the data set, we flipped the

20

underlying image along the horizontal and vertical axes with a probability of 50%. We generated 8,192 such sequences for the training set and 512 for the test set.

Experimental stimuli. For all modelling experiments in Fig. S5, we replicated the experimental protocols described above as precisely as feasible. All stimuli were projected onto a field of view that spanned $120°$ in azimuth and $90°$ in elevation at a spatial resolution of $1°$ for calcium imaging experiments and $0.5°$ for behavioral experiments. Frames were then blurred and sampled as described for natural image stimuli. Brightness values for all stimuli ran from 0 to 1 and we fixed the mean level for contrast stimuli at 0.5. For calcium imaging stimuli, we always placed the foreground disk at the center of the field of view. Patterns were rendered and processed at 100 Hz.

Tuning curves for the basic contrast experiment (Fig. S5b–d), the frequency experiment (Fig. S5f), and the background diameter experiment (Fig. S5g) were estimated from a single trial per parameter setting. For the background orientation experiment (Fig. S5e) and the step interval experiment (Fig. S5h) we averaged 100 trials with randomized background pattern phases to approximate the experimental phase stochasticity that results from individual cell receptive fields being located in different parts of the visual field. We averaged 200 trials for the behavioral stimuli (Fig. S5k) to account for the intrinsic stochasticity of the stimulus and to generate reliable model responses. Throughout Fig. S5, we calculated point estimates for all tuning curves exactly as described for the behavioral and calcium data.

Tuning curve normalization model. The analytical model for divisive normalization (Fig. 3v–x) resembles previous formulations in the literature[7, 40, 41]. The steady-state response R of a neuron is given by

$$R(c_{fg}, c_{bg}) = \frac{L_{fg}c_{fg}^p + L_{bg}c_{bg}^p}{c_{50}^p + c_{fg}^p + S^p}$$

where c_{fg} and c_{bg} are foreground and background contrast and L_{fg} and L_{bg} are weight factors defining the respective amount of linear contribution of foreground and background to the response. The semi-saturation constant c_{50} determines the contrast at which the cell responds with 50% strength and the parameter p defines the steepness of the saturation curve.
The normalization term

$$S = w_{pool} \cdot c_{bg}^q$$

gives the amount of divisive surround suppression which is proportional to background contrast to a power of q, which accounts for possible non-linear scaling behaviour, with a proportionality weight constant w_{pool}. In this model, the normalization index w_{pool}/c_{50} quantifies how much the sigmoidal tuning curve shifts to the right when c_{bg} is increased from 0 to 1 (full contrast), in relation to the semi-saturation constant. It thus describes the fold decrease in contrast sensitivity between no background contrast and full background contrast.

For evaluation of the normalization index (Fig. 3x), this model was fit individually for each cell. Parameter fits to the average tuning curve per cell type are listed in Table S2. Since tuning curves from individual cells are subject to measuring inaccuracies, we cross-validated fit quality. We optimized model parameters for the average tuning curve of 50% of all measured cells per type and evaluated variance explained for the other 50%. This was repeated 100 times with shuffled training and validation sets. For all cell types, cross-validated variance explained was more than 90% (see $R^2_{DivisiveNorm}$ in Table S2). When we repeated this procedure with a fully linear model

$$R(c_{fg}, c_{bg}) = L_{fg}c_{fg} + L_{bg}c_{bg}$$

variance explained dropped substantially for all units except L3 (see R^2_{linear} in Table S2).
This analysis was implemented using Python 2.7 and NumPy 1.11.3. Optimization of model parameters was performed using the L-BFGS-B algorithm in SciPy 0.19.0.

Data-driven detector model. The reference model in Fig. 1 was based on a standard implementation of the Reichardt-type correlational motion detector[11]. Briefly, all receptor signals of the two-dimensional input grid (see above) were filtered with a first-order high-pass ($\tau = 150ms$). We then multiplied each local signal with the delayed horizontal neighbor (first-order low-pass, $\tau = 50ms$). This was done twice in a mirror-symmetrical fashion and resulting output was subtracted. Finally, we summed across all local detectors to derive a model of tangential cell output. For the illustration in Fig. S1f, we simulated the receptor array at the full image resolution without blurring. These models were implemented in Python 3.6 using PyTorch 0.4.1.
We simulated time-resolved cell models for three basic response types: a purely linear low-pass unit (modelled after L3; Fig. S5b), a strongly normalized band-pass unit (modelled after Mi1; Fig. S5c), and a weakly normalized low-pass unit (modelled after Mi9; Fig. S5d). We hand-tuned parameters based on our and previous work[22] to qualitatively match response properties of the corresponding cell. Models were implemented as signal processing cascades (see Fig. S5a). First, signals at each location in the field of view were filtered with a spatial difference of Gaussians kernel that had a central full-width at

21

half-maximum (FWHM) of $6°$ and a FWHM of $20°$ in the surround. In accordance with results from receptive field mapping (Fig. S3), the weight ratio between surround and center was 100% for low-pass units and 50% for the band-pass model. Full-field flashes would thus produce no activation in low-pass units. This was followed by first-order temporal filters: a single low-pass filter for low-pass units ($\tau = 80ms$) or serial low- ($\tau = 50ms$) and high-pass filters ($\tau = 150ms$) for band-pass units. We then left the signal as is for ON cells or sign-inverted it for OFF cells and half-wave rectified the output by setting all negative values to zero.

For normalized cell models, we calculated local input P_i from the normalization field by pooling across rectified signals x_i with a Gaussian kernel ($FWHM = 30°$). Final output was then calculated using the divisive normalization equation

$$f(x_i) = \frac{x_i^p}{c_{50}^p + x_i^p + (w_{pool}P_i)^p}$$

where i indexes across points in space and time, c_{50} determines baseline sensitivity, exponent p regulates the static response non-linearity, and w_{pool} adjusts sensitivity to the normalization field signal. We manually tuned normalization parameters for the band-pass ($c_{50} = 0.012$, $p = 1.3$, $w_{pool} = 1.5$) and the low-pass cell ($c_{50} = 0.12$, $p = 1.1$, $w_{pool} = 3.0$) to match critical features of the empirical contrast tuning curves (Fig. S5c,d).

To generate simulated T4 responses (Fig. S5i,j), we multiplied the output of spatially adjacent low- and band-pass units. For the linear reference model we bypassed the final normalization step in both arms of the detector. We built the LPTC model (Fig. S5k) as a spatial array of T4 and T5 cells covering the full field of view, analogously to the previously described two-quadrant detector[42]. For the T5 model, we used two OFF-sensitive input units with identical parameters as for ON cells. Output from syndirectionally tuned T4 and T5 motion detectors was summed and subtracted from a mirror-symmetric, oppositely tuned array to produce LPTC model output. The same model was used to simulate natural scene responses (Fig. S5l–n). All models in Fig. S5 were implemented using Python 3.6 and NumPy 1.15.

To quantify the robustness of velocity tunings for models and LPTCs (Fig. S5n), we calculated per-velocity coefficients of variation as the ratio between response standard deviation across images and response mean across images. For neural data, we used cell-averaged mean potential to estimate these parameters.

Task-driven detector model. We implemented the trained detector model as a four-layer convolutional neural network consisting of linear input filters, a normalization stage, local multiplication, and linear spatial summation. In contrast to typical deep architectures used for object recognition, this network processed three-dimensional inputs spanning two dimensions of space as well as time.

First, receptor signals of shape 23 x 17 x 500 or 23 x 23 x 500 (azimuth, elevation, time), depending on the data set, were processed in two independent convolutional channels. The convolutions were temporally causal and spatiotemporally separable. Each of the channels was composed of a 3 x 3 x 1 spatial filter (covering 3 simulated receptors in azimuth and elevation) followed by a temporal filter of shape 1 x 1x 30 (corresponding to 300ms at the chosen time resolution of 100 Hz). Convolutions had no bias parameter. In contrast to standard Reichardt detectors, each filter weight was allowed to vary freely during optimization.

Second, we passed local output signals x_i (where i indexes points in space and time) through one of three types of local normalization: a simple pass-through (termed "linear")

$$f(x_i) = x_i$$

a static and contrast-independent compression stage (termed "static")

$$f(x_i) = tanh(x_i/c)$$

where the trained parameter c determines the sensitivity of the saturating function, or an adaptive saturation stage (termed "dynamic")

$$f(x_i) = tanh(x_i/(c + P_i))$$

where c again determines the baseline sensitivity and P_i is the instantaneous output of a 11 x 11 x 1 spatial filter (centered on the location of x_i and operating on full-wave rectified output signals $|x_i|$; see Fig. 5a). This models the fast and spatially distributed normalization we observed during experiments. We chose the hyperbolic tangent because it generalizes to positive and negative input values, the transformation closely resembles the normalization model described above, and it is more commonly used in the field of deep learning. Spatiotemporal filters were optimized independently for each of the two channels while the sensitivity parameter c was shared.

Third, we then combined signals from both channels in a EMD-type scheme where adjacent signals were multiplied and output from two mirror-symmetric pairs was subtracted. This stage was parameter-free. Finally, resulting signals were summed across space and multiplied by a trained scalar amplification factor to generate the final time-resolved output of the model. The base

22

model without normalization had 79 trainable parameters; static normalization added one parameter and dynamic normalization another 242.

We trained each model architecture to estimate the true velocity of translation stimuli using automatic differentiation, back-propagation, and stochastic gradient descent. The loss function we applied was the mean squared error (MSE) between model output and current velocity of the scene. Weights were updated using the Adam optimizer[43], with parameters set to standard values ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$). Models were trained over 800 epochs with a batch size of 128; no early stopping was used. We set the initial learning rate to 0.025 and divided it by a factor of 4 after 400, 500, and 600 steps. Input convolutional layers were initialized to random values drawn from a uniform distribution. For the pooling receptive field, we initialized each weight with 0.0001 and the sensitivity factor c with 1.0. Static sensitivity as well as pooling weights were constrained to be positive. In the dynamic normalization model, we applied a L2 penalty of 400.0 to the spatial weights of the pooling stage. Hyperparameters were determined in preliminary experiments with an independent image set. We optimized each architecture 16 to 23 times with different random number generator seeds to assess reliability and did not select models post hoc.

We implemented all architectures in Python 3.6 using PyTorch 0.4.1 for automatic differentiation. Depending on model type, a single optimization run took between 6 and 14 hours on an NVIDIA Titan Xp GPU.

Statistics. Unless indicated otherwise, error bars show bootstrapped 68% confidence intervals around the mean (estimated as corresponding distribution percentiles after resampling the data 1,000 times). All statistical tests were two-tailed. Normality of data distributions was assessed visually but not tested formally. Sample sizes were not based on power analysis but predetermined in line with standards in the field. We did not blind experimenters to genotypes or conditions during data gathering and analysis.

Code availability. Code for analysis and modelling is available upon request to the authors.

Data availability. Data from behavior, physiology, and modelling are available upon request to the authors.

Methods References.

31. A. H. Brand and N. Perrimon, "Targeted gene expression as a means of altering cell fates and generating dominant phenotypes," *Development*, vol. 118, pp. 401–415, June 1993.
32. T.-W. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreiter, R. A. Kerr, M. B. Orger, V. Jayaraman, L. L. Looger, K. Svoboda, and D. S. Kim, "Ultrasensitive fluorescent proteins for imaging neuronal activity," *Nature*, vol. 499, pp. 295–300, July 2013.
33. S. T. Sweeney, K. Broadie, J. Keane, H. Niemann, and C. J. O'Kane, "Targeted expression of tetanus toxin light chain in *Drosophila* specifically eliminates synaptic transmission and causes behavioral defects," *Neuron*, vol. 14, pp. 341–351, Feb. 1995.
34. H. G. Meyer, A. Schwegmann, J. P. Lindemann, and M. Egelhaaf, "Panoramic high dynamic range images in diverse environments," 2014.
35. A. Bahl, G. Ammer, T. Schilling, and A. Borst, "Object tracking in motion-blind flies," *Nature neuroscience*, vol. 16, no. 6, p. 730, 2013.
36. E. Serbe, M. Meier, A. Leonhardt, and A. Borst, "Comprehensive characterization of the major presynaptic elements to the *Drosophila* OFF motion detector," *Neuron*, vol. 89, pp. 829–841, Feb. 2016.
37. T. A. Pologruto, B. L. Sabatini, and K. Svoboda, "ScanImage: flexible software for operating laser scanning microscopes," *Biomed. Eng. Online*, vol. 2, p. 13, May 2003.
38. F. G. Richter, S. Fendl, J. Haag, M. S. Drews, and A. Borst, "Glutamate signaling in the fly visual system," *iScience*, vol. 7, pp. 85–95, Sept. 2018.
39. H. Jia, N. L. Rochefort, X. Chen, and A. Konnerth, "In vivo two-photon imaging of sensory-evoked dendritic calcium signals in cortical neurons," *Nat. Protoc.*, vol. 6, pp. 28–35, Jan. 2011.
40. M. Carandini, D. J. Heeger, and J. Anthony Movshon, "Linearity and normalization in simple cells of the macaque primary visual cortex," *Journal of Neuroscience*, vol. 17, pp. 8621–8644, Nov. 1997.
41. T. C. B. Freeman, S. Durand, D. C. Kiper, and M. Carandini, "Suppression without inhibition in visual cortex," *Neuron*, vol. 35, pp. 759–771, Aug. 2002.
42. H. Eichner, M. Joesch, B. Schnell, D. F. Reiff, and A. Borst, "Internal structure of the fly elementary motion detector," *Neuron*, vol. 70, no. 6, pp. 1155–1164, 2011.
43. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

23

MANUSCRIPT 3

Surround motion shapes local motion computation on the dendrites of T4 and T5 cells in *Drosophila*

Anna Schuetzenberger[1,2] and Alexander Borst[1,2]

1 Department Circuits-Computation-Models, Max-Planck-Institute of Neurobiology, Martinsried, Germany

2 Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität, Munich, Germany

ABSTRACT

Seeing the direction of motion is important for many animals, including humans. Accordingly, motion sensitive cells have been found in the visual systems of a large range of organisms. In the fruit fly *Drosophila melanogaster*, the first motion selective cells are T4 and T5 cells, which preferentially respond to local motion in one direction, and respond only weakly to motion in other directions. The array of T4 and T5 cells covers the complete visual field, and thus forms a visual map of local motion. How motion signals are computed on the dendrites of T4 and T5 has been subject to extensive investigation. To date it has largely been assumed that individual T4 and T5 cells act as independent local motion detectors. However, recent electron microscopy studies showed that T4 and T5 cells of the same subtype form synaptic connections. Here, we therefore investigate whether local motion signals of T4 and T5 depend on surround motion. We find that, indeed, responses to local motion differ from responses to global motion. Secondly, we ask whether these differences arise owing to the synaptic connections between T4/T5. While this is the case for T5, for T4 our results remain inconclusive. In summary, in this study we could demonstrate that the local motion signals of T4 and T5 are shaped by surround motion signals, shedding new light on the computation of visual motion in *Drosophila*.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGEMENTS

INTRODUCTION

Information about visual motion is crucial for many animals. Moving objects may be detected only by their relative movement against the background, and moving yourself – or being moved – induces whole-field visual motion, which can be used for stabilization or navigation. However, the direction of motion is not explicitly represented at the level of the photoreceptors. Subsequent neural circuits have to compute visual motion information from the photoreceptor signals.

In the model organism *Drosophila*, motion vision has been studied in unprecedented detail. The first motion selective cells in the visual system, T4 and T5 cells, respond preferentially to local motion in a particular direction and only weakly to motion in other directions (Maisak et al., 2013). While T4 responds only to brightness increments (ON), T5 responds only to brightness decrements (OFF motion). Four subtypes (a-d) of both cell types code for the four cardinal directions of motion (Maisak et al., 2013). Lobula plate tangential cells (LPTCs) receive signals from a large number of T4 and T5 cells and thus integrate local motion signals over a large visual area (Mauss et al., 2015; Schnell et al., 2012). Blocking T4 and T5 abolishes many motion-driven behavioral responses like the optomotor, the landing, as well as the avoidance response (Bahl et al., 2013; Maisak et al., 2013; Schilling and Borst, 2015).

A large body of theoretical work, aimed to describe motion-based behaviors and properties of motion-sensitive cells, produced several algorithmic models of motion detection. Among the most influential ones are the Hassenstein-Reichardt and the Barlow-Levick detector (Barlow and Levick, 1964; Hassenstein and Reichardt, 1956). In both models, signals from two adjacent points in space, one of them delayed in time, are combined. In the Hassenstein-Reichardt model the signals are multiplied, leading to an enhanced response in a particular direction (preferred-direction enhancement or PDE). In the Barlow-Levick model, they are divided, leading to a suppressed response in a particular direction (null-direction suppression or NDS).

The visual system of *Drosophila* receives input from photoreceptors contained in 750 ommatidia. It consists of several neuropils, which each are arranged as retinotopic maps formed by different cell types (Fischbach and Dittrich, 1989; Ramon y Cajal and Sanches, 1915; Strausfeld, 1971). Recently, the cells providing synaptic input to T4 and T5 were identified in several electron microscopy studies (Shinomiya et al., 2019, 2014; Takemura et al., 2017, 2015). An individual T4 or T5 cell receives synaptic input from several different cell types. Consistent with its dendritic arbor spanning several columns, these presynaptic inputs come from several neighboring columns (Shinomiya et al., 2019; Takemura et al., 2017). Thus, T4 and T5 potentially compare signals not only from two but several adjacent points in space.

A number of studies has since investigated the cellular implementation of motion detection. For T4, the presynaptic cell types Mi1, Tm3, Mi4 and Mi9 play a role in the computation of motion on the T4 dendrite (Ammer et al., 2015; Arenz et al., 2017; Strother et al., 2017). Silencing of the presynaptic cells Tm1, Tm2, Tm4 and Tm9 each has an effect on the computation of motion by T5 (Fisher et al., 2015a; Serbe et al., 2016). Furthermore, responses to apparent motion stimuli

show the signatures of PDE in some studies (Fisher et al., 2015b; Salazar-Gatzimas et al., 2016), of NDS in other studies (Gruntman et al., 2018) or of a combination of both (Haag et al., 2017, 2016; Leong et al., 2016). Despite these studies, it is still under debate how each presynaptic cell contributes to the computation of motion.

Recently, electron microscopy studies have shown that also T4/T5 cells of the same subtype (e.g. T4c contacting neighboring T4c) are interconnected with each other synaptically (Shinomiya et al., 2019; Takemura et al., 2017). This hints at a modulation of local motion computation by surround motion. We therefore ask whether T5 and T5 cells respond differently to local than to global motion, and, if such differences occur, whether they can be attributed to synaptic interactions between T4 and T5 cells.

RESULTS

T4 and T5 responses are modulated by surround motion

We first set out to test the responses of T4 and T5 to local and global motion stimuli. To this end, we expressed the genetically encoded calcium indicator GCaMP6m specifically in T4c and T5c, which are selective to upward motion. The dendrites of T4 reside in layer 10 of the medulla neuropil. We presented visual stimuli on a cylindrical screen, and, using remote-focusing two-photon microscopy, we aligned the imaging plane with the oblique, retinotopic plane that is spanned by T4 dendrites (Fig. 1a). Analogously for T5, the imaging plane was aligned with the first layer of the lobula. Local and global edges travelling either upwards (in preferred direction, PD) or downwards (in null direction, ND), of either polarity (ON, bright edges on a dark background, or OFF, dark edges on a bright background) were used as visual stimuli (Fig. 1b). Local edges were shown in a 25 ° window placed on the receptive field of the imaged cells, while global edges covered the whole stimulation arena. Since we were able to view the retinotopic array of T4 and T5 dendrites, we could narrow down the area excited by the local edge to a circular patch of dendritic branches, and we could observe the wave of calcium signal travelling though the complete field of view during global stimulation (Fig. 1c-d). In order to precisely locate the dendritic branches corresponding to the center of the local stimulus, we extracted the pixels with the highest fluorescence as well as a similar timing in response to the local PD and ND edges (Fig. 1e).

We then compared PD and ND responses to local and global edges and observed several differences. For OFF edges in T5, the amplitude of the PD response to local or global edges was similar (Fig. 1f). However, the amplitude of the ND response was higher for local edges than for global edges (Fig. 1g). This difference is also reflected in the direction selectivity index (DSI, computed as the difference of PD and ND peak responses, divided by their sum) which is higher for global stimulation than for local stimulation (Fig. 1h). On average, the difference in DSI (ΔDSI) between local and global stimulation was -0.21±0.056 (95% confidence interval, Fig. 1h, right). Analogously, in T4 (to ON edges) ND responses were higher during local than during global stimulation, and the DSI was lower for local stimulation (Fig. i-k). Thus, local motion computation is affected by surround motion in that global motion reduces the ND response amplitude and thus
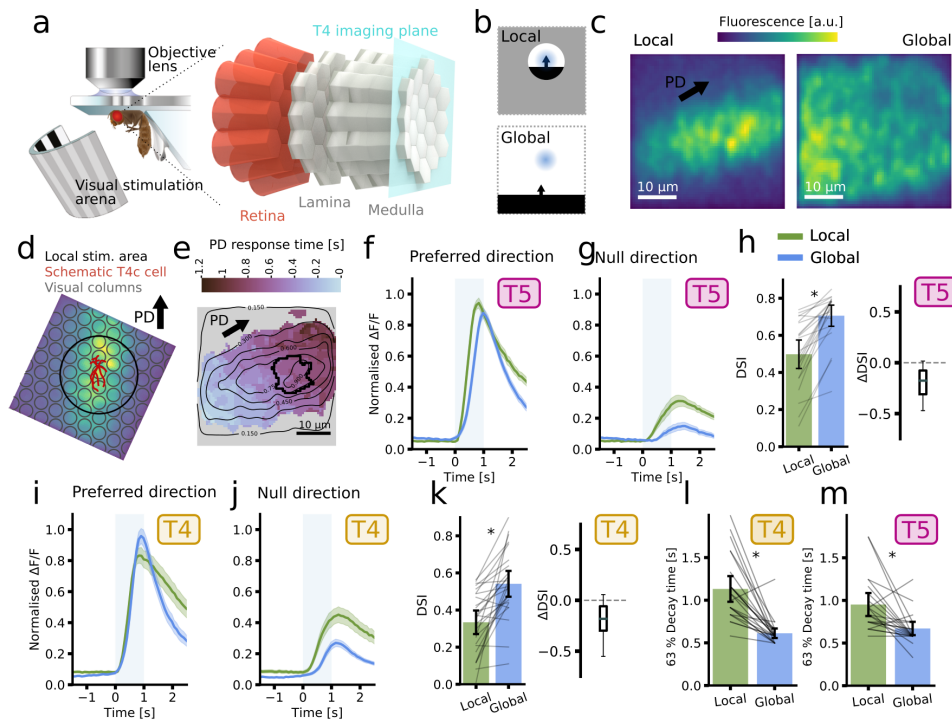
Figure 1 | T4 and T5 respond differently to local and global motion. a) Schematic showing the imaging setup and visual stimulation arena together with the fly visual system and the oblique imaging plane in the medulla. b) Schematic of the visual stimuli used. Top: a local OFF edge moving across the receptive field of the cell (blue). The area around the 25 ° local window is at intermediate contrast. Bottom: Global OFF edge. All edges moved at 30 °/s. c) Mean images (over three trials) of the responses to a local (left) or global (right) PD edge in the lobula (T5) of an example recording. The local edge stimulus elicited a response only in the central area. d) Method of ROI extraction: Peak response time to the local edge stimulus (color-coded), fluorescence intensity (gradient lines), and the resulting ROI (thick black line) for the example recording in c. e) T5 PD responses to local (green) and global (blue) edges expressed as ΔF/F. Global edge responses were shifted in time according to the receptive field of each recording to align with the local responses. f) as in e for T5 ND responses. g) Left: Direction selectivity index (DSI) for local and global edges, expressed as (PD-ND)/(PD+ND) peak amplitude. Individual lines represent individual recordings. Right: Difference in DSI between local and global stimulation. h) - j) as e – g for T4. k) Decay time expressed as the time it takes the response to drop to 63 % from the peak response, for T4 PD responses. l) as k for T5. Shaded errors represent 68 % confidence intervals; error bars 95 % confidence intervals. Box plots represent median, quartiles (box) and range (whiskers). All responses are normalized to the peak response across Fig. 1 and 2, allowing for amplitude comparisons. N = 12, n = 22 for T5; N = 15, n = 26 for T4.

increases direction selectivity.

In addition to the effect on direction selectivity, we also observed a difference in the response dynamics of the PD responses. The responses to the local edge showed a slower decay time than the responses to the global edge, expressed as the time it takes the response to drop to 63% of its maximum (Fig. 1l-m).

Local and global stimulation show similar effects in neighboring dendritic branches

Next, we set out to compare our findings in the central ROI, which predominantly comprised of cells whose complete dendritic tree was covered by the local stimulus, to cells further at the edge of circular window. To this end, we extracted two additional ROIs that included dendritic branches that either corresponded to the lower area or the upper area of the circular window from the relative response timing of PD and ND edge stimulus (Fig. 2 a-d). While the overall response dynamics
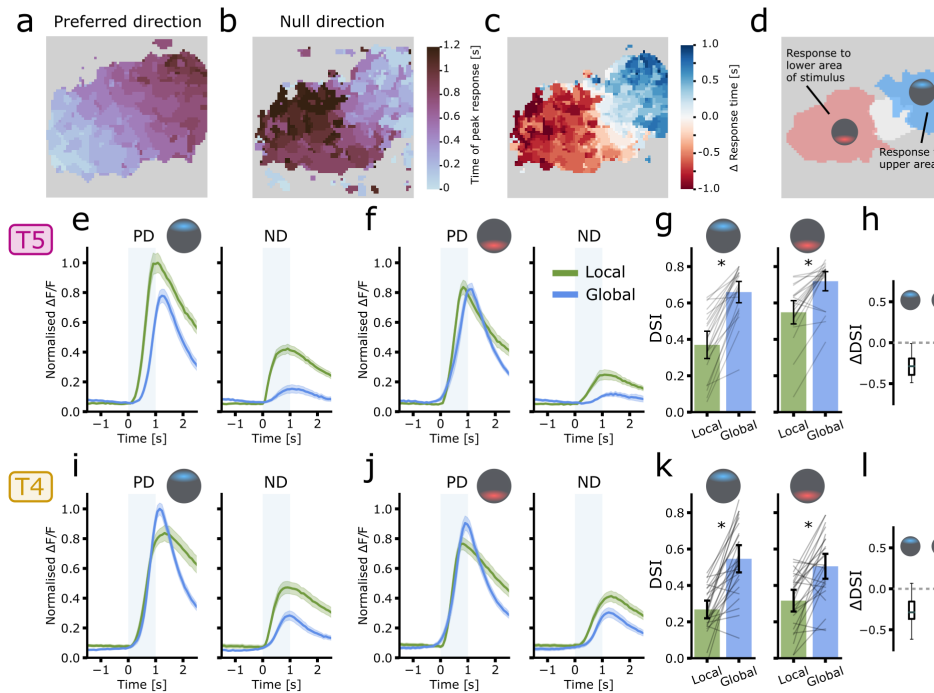
Figure 2 | Response properties of dendritic branches at the edge of the local stimulus. a) Time of peak response to a local PD stimulus for the example recording in Fig. 1c. b) Time of peak response to the local ND stimulus. c) Difference of PD and ND response time. d) Resulting ROIs of pixels of containing dendritic branches that correspond to the lower area of the circular stimulus (red), or to the upper area of the circular stimulus (blue). In the middle is the central area of Fig. 1 (white). e) T5 responses to local and global PD and ND edges for the upper area of the circular window. f) Analogous to e for the lower area. g) Resulting DSIs for upper and lower areas. Grey lines represent individual recordings. h) Difference in DSI between local and global stimulation, for upper area and lower area. i) – l) Analogous to e-h for T4. Shaded errors represent 68 % confidence intervals; error bars 95 % confidence intervals. Box plots represent median, quartiles (box) and range (whiskers). All responses are normalized to the peak response across Fig. 1 and 2, allowing for amplitude comparisons. N = 12, n = 22 for T5; N = 15, n = 26 for T4.

differed slightly, the effect we observed in the central area, both the elevated ND responses, as well as the smaller DSI for local as compared to global stimuli, remains the same for both T4 and T5 (Fig. 2 e-l). In the upper area ROI, for T5, also PD responses were elevated for local as compared to global stimulation (Fig. 2e). Overall, the dendritic branches in the close vicinity show responses similar to the central ROI. This firstly shows that our analysis is not critically dependent on the chosen pixels, and secondly indicates that cells at the edge of the circular window do not receive qualitatively different input, even though the receptive field of some of the presynaptic cells likely extends to outside the local circle.

For T5, the difference between local and global stimulation depends on T4/T5 connections

Next, we set out to test whether the differences we observed are due to interactions between T4/T5 themselves, or, alternatively, due to response differences of the cells presynaptic to T4 and T5 to the local and global stimuli. To this end, we expressed the synaptic blocking tool TNT (Sweeney et al., 1995) in all T4 and T5 cells, while expressing the GCaMP6m in T4c and T5c cells as previously. TNT affects presynaptic vesicle release, and should therefore not interfere with dendritic calcium signaling. Fig. 3a shows the local and global motion responses of T5c to PD and ND edges.
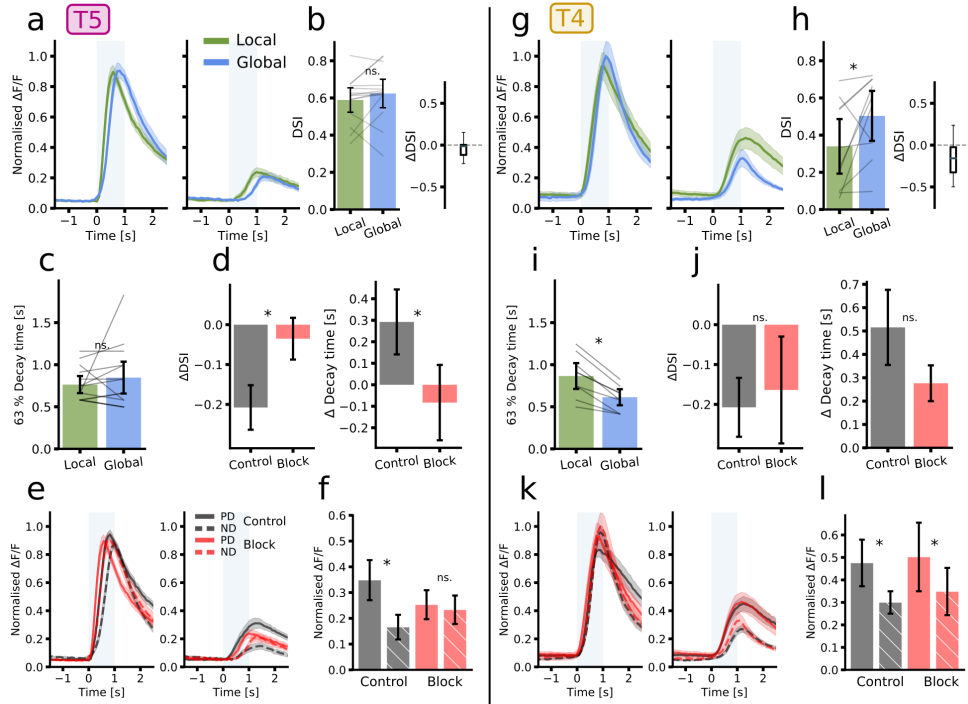
Figure 3 | T5 responses are modulated by signals of surround motion detectors. a) Responses to local and global PD and ND edges of T5 while T4 and T5 output is silenced. b) DSI of local and global responses. c) Analogous to a for T4. d) Analogous to b for T4. e) Decay time expressed as the time it takes the response to drop to 63 % from the peak response, for T5 PD responses. f) Average difference decay time for control (T5c imaging without TNT expressed) and block (T5c imaging with TNT in T4 and T5) conditions. f) Average difference in DSI for control and block. g) Analogous to e for T4. h) Analogous to f for T4. i) Comparison of control and block responses for PD (left) and ND (right) for T5. j) Maximum responses to ND responses for T5. k) Same as i for T4. l) Same as j for T4. Shaded errors represent 68 % confidence intervals; error bars 95 % confidence intervals. Box plots represent median, quartiles (box) and range (whiskers). N = 6, n = 14 for T5; N = 5, n = 10 for T4.

Strikingly, for both PD and ND, responses to local and global stimuli were indistinguishable. The amplitude of the ND responses, as well as the DSI for local and global stimulation did not show significant differences (Fig. 3b). In addition, the PD response dynamics for both local and global edges were the same (Fig. 3c). We then directly compared control and block condition by comparing the ΔDSI (difference in DSI between local and global stimulation). The ΔDSI in control flies was around -0.21±0.56 (95% confidence interval), while the ΔDSI in block flies -0.035±0.052, showing a significant difference between control and block (Fig. 3d, left). Similarly, the Δ Decay Time was significantly different between control and block (Fig. 3d, right). We next asked whether local ND responses in the block condition decrease to match the global ND response, or whether, instead, global ND responses increase. Instead, we found ND responses to be at an intermediate level (Fig. 3e-f). This measure is sensitive to small variations in the absolute response amplitudes in block and control conditions, which we cannot exclude here. Thus, we cannot conclusively say whether T4/T5 connections increase local ND responses, decrease global ND responses, or both.

In contrast to the results for T5, for T4, when blocking the T4/T5 output, the differences between local and global stimulation largely remain (Fig. 3g-i). Both the difference in DSI and the difference in decay time are still significant, albeit weaker (Fig. 3h-i), and we could not establish significant differences between control and block in ΔDSI (Fig. 3j, left). The Δ Decay Time shows a weak

reduction in block flies compared to controls that is however not significant. Furthermore, when directly comparing response amplitudes, control and block responses were similar (Fig. 3k-l). Hence, blocking T4/T5 does not have the same effect for T4 as for T5 on local motion computation.

DISCUSSION

Here, we showed that local T4 as well as T5 responses are influenced by motion in the surround, outside their receptive field, which manifests itself in reduced ND responses and a higher degree of direction selectivity for global edge stimuli. We then demonstrated that for T5 but not for T4, this modulation is mediated by (direct or indirect) synaptic connections between T4/T5 cells.

The degree of direction selectivity found in T4 and T5 varies slightly across different studies. Consistent with our findings, other studies that used global motion stimuli generally reported small ND responses and high direction selectivity (e.g. Gruntman et al., 2018; Leong et al., 2016; Maisak et al., 2013), whereas in studies using local apparent motion stimuli, higher ND responses were generally reported (Fisher et al., 2015b; Haag et al., 2017, 2016). As these studies differ in a number of parameters, it is not possible to pinpoint the differences in direction selectivity to a specific stimulus feature. However, one aspect to affect direction selectivity we identified here, the size of an edge stimulus, might have contributed to the various degrees of direction selectivity reported.

In a previous study, T4 and T5 output had been blocked while imaging full-field square-wave grating responses from both T4c and T5c axon terminals (Haag et al., 2016). This study did not observe notable differences in T4c and T5c responses between control and block. This is consistent with our results, which do not show a significant difference between control and block when comparing response amplitudes to global edges. Another study recorded dendritic responses of all subtypes of T4 while silencing T4 and T5, showing edges in a 30 ° window (Strother et al., 2017). In this study, while it could not distinguish PD from ND responses, differences in the response dynamics between block and control conditions were observed, consistent with our data.

T4 and T5 cells sample signals from several columns, and some cell types presynaptic to T4 and T5 have large receptive fields. Therefore, most likely, the synaptic input received by a T4 or T5 cell centered on the local stimulus is different for local and global stimulation. Here, we found that T4 cells respond differently to local as compared to global stimulation even when they do not receive synaptic input from other T4/T5 cells. Thus, either, the tool we used to block synaptic output was not effective, or, alternatively, the cells presynaptic to T4 respond differently to local and global edges, resulting in altered direction selectivity on T4.

T4 as well as T5 make synaptic contacts with cells of the same subtype (e.g. T5c with T5c; Shinomiya et al., 2019; Takemura et al., 2017). Moreover, these connections are directed according to the cells' preferred direction and dendrite orientation, and are cholinergic. At its tips, a T4/T5 cell receives synaptic input predominantly from cells on its preferred side, i.e. the cells that, when a PD edge moves across the visual field, would have been stimulated just before. We reason that it
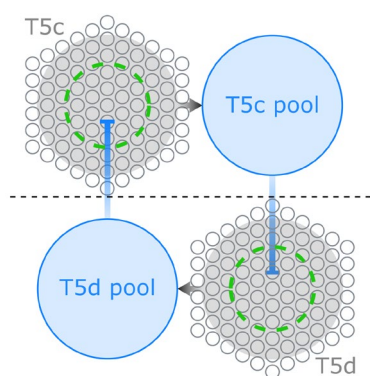
Figure 4 | Schematic of a possible circuit. The T5 signals (grey circles) of a larger spatial area (grey shaded area) is pooled by a pool cell (blue) which inhibits the T5 subtype with opposite PD. The size of the local stimulus (green dashed circle) is smaller than the area of the pool.

is unlikely that these synapses mediate the effect on the ND responses and direction selectivity. First, we observed that predominantly ND response amplitudes are affected, while PD response amplitudes are largely unaltered. Additional input from cells that also have the same preferred direction and response properties is therefore unlikely to mediate this effect. Second, these unidirectional synapses should only affect the cells upstream. Thus, the cells in the upper area of the local stimulus should receive input from its presynaptic partner (which is in the center of the local stimulus) regardless of whether the stimulus is local or global. However, also in the upper area we see differences between local and global stimulation. Electron microscopy identified also other presynaptic partners of T4 and T5 with to date unknown function (Shinomiya et al., 2019). TmY15 receives input from a number of cell types and could potentially mediate feedback connections from the lobula plate to the medulla and lobula (Shinomiya et al., 2019). Alternatively, T4 and T5 might receive additional inputs in the lobula plate, where the connectome is not resolved yet.

One possible circuit that can explain our findings involves an interneuron that pools signals from a larger area of T5d cells (T5d pool), larger than the local stimulus, and that inhibits T5c cells (Fig. 4). In this scenario, T5c receives more inhibition during a global motion stimulus than during a local motion stimulus. LPi cells in the lobula plate receive signals from one subtype of T4 and T5, and inhibit LPTCs with the opposite PD, rendering them motion opponent (Mauss et al., 2015). Thus, LPis are in a position to provide a signal to T5 (and T4) cells of the subtype with opposite PD. Moreover, their receptive field size is larger than that of T4 and T5 (Richter et al., 2018). In that, LPis constitute a candidate cell type for mediating the interaction between local motion detectors.

A recent publication on integration of T4 and T5 signals in HS cells described a non-linear combination of motion signals on HS (Barnhart et al., 2018). Moreover, they showed that T4 and T5 responses are modulated by edges moving adjacent to their receptive field. While they did not block T4/T5 output, and thus did not test whether this modulation is due to functional T4/T5 synapses, their result shows that surround motion can affect local motion responses, consistent with our findings.

In previous work, it has often been assumed that the local motion detectors act independent of each other. Here we showed that motion sensitive cells can interact (directly or indirectly), and thus response properties of T5 are dependent not only on the input signals they receive from non-direction selective cells, but also on responses of surrounding motion detectors. Therefore, in summary, in order to eventually unravel local motion computation on the dendrites of T4 and T5,

interactions between local motion detectors have to be taken into account. Our results open the doors to new questions to be answered in future studies: Which synapses or cells mediate interactions between T4 and T5? Which subtypes are interacting? And, importantly, what are the consequences for local motion coding as well as motion vision in the natural context?

METHODS

EXPERIMENTAL ANIMALS

Drosophila melanogaster were reared on standard conditions (corn-meal agar, 25 °, 60 % relative humidity) and subject to a 12h-light / 12-dark cycle. The Gal4-UAS system (Brand and Perrimon, 1993) as well as the LexA-LexAop system (Lai and Lee, 2006) were used to express transgenes. We used the genetically encoded calcium indicator GCaMP6m (Chen et al., 2013) to image calcium concentration, and TNT-E (Sweeney et al., 1995) to block synaptic output. The genotypes for all experiments used in this study are listed in the following table:

Short name	Expression	Genotype	
T4T5 block flies	TNT-E in T4+T5 GCaMP6m in T4c+T5c	$w^+ ; \dfrac{UAS - TNTE}{al\ 4 - R59E08 - AD} ;$	$\dfrac{LexA - VT50384, LexAop - CaMP6m}{al4\ - R42F06 - DBD}$
Control flies	GCaMP6m in T4c+T5c	$w^+ ; + ; \dfrac{LexA - VT50384, LexAop - CaMP\ 6m}{+}$	

REMOTE-FOCUSING TWO-PHOTON MICROSCOPY

We used a custom-built remote focusing microscope to image oblique planes, which were aligned with the retinotopic plane spanned T4 or T5 dendrites. The remote focusing module was described in detail in **Manuscript 1**. Briefly, remote focusing enables the fast, aberration-free movement of the laser focal spot along the optical axis (Botcherby et al., 2008). With additional control software based on ScanImage 5.1 (Pologruto et al., 2003), all three imaging axes can be controlled independently, and thus the imaging plane can be rotated in 3D. Rotational controls in the user interface allow the user to align the imaging plane with the retinotopic plane spanned by T4 or T5 cells.

A MaiTai eHP DS Ti:Sapphire oscillator was used as laser source, tuned to 910 nm. Fluorescence was collected via a photomultiplier tube (Hamamatsu H10770-PB40). To scan the beam laterally, we used two galvanometer optical scanners (Cambridge Technologies 6215H). To scan the beam along the optical axis, a customized piezo actuator (nanoFaktur) moved a third mirror (Thorlabs PF05-03-M01, 12mm diameter). As the remote focusing lens, we used Leica HC PLAN APO 40x, and as imaging objective lens the water immersion Olympus Lumplfln 40x with a NA of 0.8. A customized version of ScanImage 5.1 (Matlab-based) was used to collect the data. In order to synchronize the recording with the stimulus, a second channel recorded a trigger signal from the visual stimulation arena.

VISUAL STIMULATION

A custom-built, cylindrical stimulation arena was used to present visual stimuli to the fly. The arena has been described in detail elsewhere (Arenz et al., 2017). Briefly, visual stimuli were cast with two projectors onto a semi-transparent cylinder surrounding the fly. The cylinder covered 105 ° in elevation and 180 ° in azimuth of the visual field of the fly. The wavelength of the projectors was limited > 550 nm, to separate it from the GCaMP emission wavelength. Stimuli were projected with a frame rate of 180 Hz in 255 luminance steps.

The stimulus set consisted of 12 different stimuli, six of which were local stimuli and six global stimuli. Global stimuli comprised of four edge stimuli and two grating stimuli, each shown on the full extent of the stimulation arena. Edges were shown in two different directions (upwards or in preferred direction, and downwards, in null direction) and two different polarities (a bright edge, on a dark background, and dark edge on a bright background). Sine gratings were shown in two different directions. The grating data was not analyzed as part of this manuscript. Edges and gratings moved at 30 °/s. Gratings were shown for 3 s, edge stimuli for 5 s. Prior to a grating stimulus, the static grating was shown for 1 s. Prior to an edge stimulus, the screen took the contrast of the background, that is, either dark (for bright edges), or bright (for dark edges). Local stimuli were the same stimuli as global stimuli, but shown in a 25 ° window centered on the receptive field of the cells in the imaging field of view. The surrounding area on the stimulation arena was kept at an intermediate grey level. Edge stimuli were only 1 s long, since the edge covered the complete 25 ° in under 1 s. Otherwise, local stimuli were identical to global stimuli. Each stimulus was repeated three times (= trials). The trial order was randomized for every experiment, and there were 4 s of break between each trial. In the first half of the break, the static image from the end of the last trial remained on the screen. In the second half of the break, the screen switched to the background luminance of the next trial. The luminance values for bright, dark and intermediate brightness were 255, 0 and 125. We note that, due to the geometry of the arena, the height of the local, circular window depended slightly on the position in elevation, and was sometimes smaller than 25 °. The width of the circular window was 25 ° regardless of position. The results presented here showed no dependency on this factor (data not shown).

EXPERIMENTAL PROCEDURE

Flies were prepared for calcium imaging following a standard procedure (Maisak et al., 2013). Briefly, flies were anesthetized on ice and glued to a holder with UV-hardening glue. The fly's head was bent to gain access to the brain from the dorso-posterior side through a small hole in the holder. Ringer's solution was applied, and the cuticle at the back of the head, as well as fat tissue and trachea, were removed for optical access.

A grating stimulus was shown to search for areas in the medulla (T4) or lobula (T5) with large calcium responses. The imaging plane was then tilted to align it with the plane spanned by the dendrites of T4 or T5. The imaging field of view was around 40x40 μm. In the medulla, this corresponds to roughly 8x8 columns, in the lobula 8x10 columns (10 along the anterior-posterior axis, see: Shinomiya et al., 2019). Subsequently, we used an interactive, local motion stimulus to determine the receptive field location of the cells in our imaging field of view. The stimulus, an

edge moving up and down in a 7x8 ° square, could be moved interactively with the cursor across the stimulation arena. We moved the stimulus such that the center of the imaging field of view responded to the local motion. The coordinates of this stimulus were then used to position the local stimuli. Data processing (see next section) then defined the precise center of the local stimulus post-hoc. All data were recorded at a frame rate of 7.5 Hz, a zoom of 7, equivalent to a field of view size of 40 μm, and a resolution of 64x64 pixel.

DATA PROCESSING AND ANALYSIS

Data were processed using custom-written software in Python 2.7. First, a phase correlation algorithm was used to register the imaging frames to the mean image (Arenz et al., 2017). In order to select the pixels corresponding to the cells that responded to the center of the local stimulus, we used the image frames that were acquired during the presentation of the local edges moving in the preferred direction (dark edges for T5, bright edges for T4). First, we calculated a low-pass filtered mean image across all frames. The 1500 pixels with the highest fluorescence values of this mean image were chosen as a first approximation of the center, and analyzed further in the second step: For each pixel, we extracted the time of its peak fluorescence response, both to the preferred-direction and null-direction local edge. We assumed that cells responsive to the center of the local stimulus should respond to the PD and ND stimulus at approximately the same point in the trial. We therefore subtracted the PD response time from the ND response time, and we defined pixels to look at the center if the absolute value of this difference was smaller than 0.30 s. Pixels that responded first to the PD edge (PD - ND response time < -0.25 s) were interpreted as the cells responsive to the lower area of the circle, and pixels that responded first to the ND edge (PD - ND response time > 0.25 s) were interpreted as the cells responsive to the upper area of the circle. To define the lateral extent of the central area more narrowly, we used a stricter selection (500 pixels with the highest fluorescence values) on the mean fluorescence in the first step.

For the three resulting ROIs, central, lower and upper, we then computed a dynamic baseline and computed the relative fluorescence ΔF/F by subtracting the baseline from the raw fluorescence traces, and then dividing by it. For each acquisition (and all three ROIs from that acquisition), calcium signals were then normalized in the following way: For each trial, the maximum and minimum ΔF/F was computed. The six highest maxima, as well as the six lowest minima were then averaged, resulting in one maximum and one minimum, which served as the normalization factors. This way, amplitudes of PD and ND responses of the different ROIs can be compared directly.

The direction selectivity index was calculated as $DSI = \frac{PD-ND}{PD+ND}$, where PD denotes the peak response to the preferred direction edge and ND the peak response to the null direction edge. To determine the 63 % decay time, we extracted the peak amplitude and searched for the first time point after the peak where the response had dropped under 63 %. For some acquisitions with local stimuli, this point was not reached at the end of the trial. This effected 1/22 acquisition for the T4 control condition, 1/26 for T5 control and 1/10 for T4 block. These acquisitions were not included in the decay time figures and in the statistics. This exclusion criterion slightly underestimates the decay time for the local stimuli. However, even when slightly underestimating the decay time for

local stimuli, they were still significantly higher than for global stimuli. Paired Student's t-tests were used to evaluate significances in for the direction selectivity index and decay times, comparing local and global stimulation. Independent Student's t-tests were used to test significances between block and control condition. We termed significant all differences with p<0.05. Detailed p-values are found in the supplementary table 1.

REFERENCES

Ammer G, Leonhardt A, Bahl A, Dickson BJ, Borst A. 2015. Functional specialization of neural input elements to the Drosophila ON motion detector. *Curr Biol* **25**.

Arenz A, Drews MS, Richter FG, Ammer G, Borst A. 2017. The temporal tuning of the Drosophila motion detectors is determined by the dynamics of their input elements. *Curr Biol* **27**:929–944.

Bahl A, Ammer G, Schilling T, Borst A. 2013. Object tracking in motion-blind flies. *Nat Neurosci* **16**:730–738.

Barlow HB, Levick WR. 1964. The mechanism of directionally selective units in rabbit's retina. *J Physiol* **178**:477–504.

Barnhart EL, Wang IE, Wei H, Desplan C, Clandinin TR. 2018. Sequential nonlinear filtering of local motion cues by global motion circuits. *Neuron* **100**:229–243.

Botcherby EJ, Juškaitis R, Booth MJ, Wilson T. 2008. An optical technique for remote focusing in microscopy. *Opt Commun* **281**:880–887.

Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **415**:401–415.

Chen T-W, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, Looger LL, Svoboda K, Kim DS. 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**:295–300.

Fischbach K-F, Dittrich APM. 1989. The optic lobe of Drosophila melanogaster. I . A Golgi analysis of wild-type structure. *Cell Tissue* **258**:441–475.

Fisher YE, Leong JCS, Sporar K, Ketkar MD, Gohl DM, Clandinin TR, Silies M. 2015a. A class of visual neurons with wide-field properties is required for local motion detection. *Curr Biol* **25**.

Fisher YE, Silies M, Clandinin TR, Fisher YE, Silies M, Clandinin TR. 2015b. Orientation selectivity sharpens motion detection in Drosophila. *Neuron* **88**:390–402.

Gruntman E, Romani S, Reiser MB. 2018. Simple integration of fast excitation and offset, delayed inhibition computes directional selectivity in Drosophila. *Nat Neurosci* **21**:250–257.

Haag J, Arenz A, Serbe E, Gabbiani F, Borst A. 2016. Complementary mechanisms create direction selectivity in the fly. *Elife* **5**:e17421.

Haag J, Mishra A, Borst A. 2017. A common directional tuning mechanism of Drosophila motion-sensing neurons in the ON and in the OFF pathway. *Elife* **6**:e29044.

Hassenstein B, Reichardt W. 1956. Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. *Zeitschrift für Naturforsch* **11**:513–524.

Lai SL, Lee T. 2006. Genetic mosaic with dual binary transcriptional systems in Drosophila. *Nat Neurosci* **9**:703–709.

Leong JCS, Esch JJ, Poole B, Ganguli S, Clandinin TR. 2016. Direction selectivity in Drosophila

emerges from preferred-direction enhancement and null-direction suppression. *J Neurosci* **36**:8078–8092.

Maisak MS, Haag J, Ammer G, Serbe E, Meier M, Leonhardt A, Schilling T, Bahl A, Rubin GM, Nern A, Dickson BJ, Reiff DF, Hopp E, Borst A. 2013. A directional tuning map of Drosophila elementary motion detectors. *Nature* **500**:212–216.

Mauss AS, Pankova K, Arenz A, Nern A, Rubin GM, Borst A. 2015. Neural circuit to integrate opposing motions in the visual field. *Cell* **162**:351–362.

Pologruto TA, Sabatini BL, Svoboda K. 2003. ScanImage: Flexible software for operating laser scanning microscopes. *Biomed Eng Online* **9**:13.

Ramon y Cajal S, Sanches D. 1915. Contribución al conocimiento de los centros nerviosos de los insectos.

Richter FG, Fendl S, Haag J, Drews MS, Borst A. 2018. Glutamate signaling in the fly visual system. *iScience* **7**:85–95.

Salazar-Gatzimas E, Chen J, Creamer MS, Mano O, Mandel HB, Matulis CA, Pottackal J, Clark DA. 2016. Direct measurement of correlation responses in Drosophila elementary motion detectors reveals fast timescale tuning. *Neuron* **92**:227–239.

Schilling T, Borst A. 2015. Local motion detectors are required for the computation of expansion flow-fields. *Biol Open* **4**:1105–1108.

Schnell B, Raghu SV, Nern A, Borst A. 2012. Columnar cells necessary for motion responses of wide-field visual interneurons in Drosophila. *J Comp Physiol A* **198**:389–395.

Serbe E, Meier M, Leonhardt A, Borst A. 2016. Comprehensive characterization of the major presynaptic elements to the Drosophila OFF motion detector. *Neuron* **89**:829–841.

Shinomiya K, Huang G, Lu Z, Parag T, Xu CS, Aniceto R, Ansari N, Cheatham N, Lauchie S, Neace E, Ogundeyi O, Ordish C, Peel D, Shinomiya A, Smith C, Takemura S, Talebi I, Rivlin PK, Nern A, Scheffer LK, Plaza SM, Meinertzhagen IA. 2019. Comparisons between the ON- and OFF-edge motion pathways in the Drosophila brain. *Elife* **8**:e40025.

Shinomiya K, Karuppudurai T, Lin T-Y, Lu Z, Lee C-H, Meinertzhagen IA. 2014. Candidate neural substrates for OFF-edge motion detection in Drosophila. *Curr Biol* **24**:1062–70.

Strausfeld NJ. 1971. The organization of the insect visual system (Light Microscopy). *Symp Soc Exp Biol* **25**:379–390.

Strother JA, Wu S, Wong AM, Nern A, Rogers EM, Le JQ, Rubin GM, Reiser MB. 2017. The emergence of directional selectivity in the visual motion pathway of Drosophila. *Neuron* **94**:168–182.

Sweeney ST, Broadie K, Keane J, Niemann H, Kane CJO. 1995. Targeted expression of tetanus toxin light chain in Drosophila specifically eliminates synaptic transmission and causes behavioral defects. *Neuron* **14**:341–351.

Takemura S, Nern A, Chklovskii DB, Scheffer LK, Rubin GM, Meinertzhagen IA. 2017. The comprehensive connectome of a neural substrate for 'ON' motion detection in Drosophila. *Elife* **6**:e24394.

Takemura Shin-ya, Xu CS, Lu Z, Rivlin PK, Parag T, Olbris DJ, Plaza S, Zhao T, Katz WT, Umayam L, Weaver C, Hess HF, Horne JA, Nunez-Iglesias J, Aniceto R, Chang L-A, Lauchie S, Nasca A, Ogundeyi O, Sigmund C, Takemura Satoko, Tran J, Langille C, Le Lacheur K, McLin S, Shinomiya A, Chklovskii DB, Meinertzhagen IA, Scheffer LK. 2015. Synaptic circuits and their variations within different columns in the visual system of Drosophila. *Proc Natl Acad Sci* **112**:13711–13716.

TABLE1

Figure	Test	p-value	Data	Cell	Condition
1f	Paired Student's t-test	0.138239	PD amplitudes	T5	Control
1g	Paired Student's t-test	0.000007	ND amplitudes	T5	Control
1h	Paired Student's t-test	0.000001	DSIs	T5	Control
1i	Paired Student's t-test	0.235891	PD amplitudes	T4	Control
1j	Paired Student's t-test	0.000323	ND amplitudes	T4	Control
1k	Paired Student's t-test	0.000014	DSIs	T4	Control
1l	Paired Student's t-test	0.000003	Decay times	T4	Control
1m	Paired Student's t-test	0.001422	Decay times	T5	Control
2e left	Paired Student's t-test	0.000113	PD amplitudes	T5	Control
2e right	Paired Student's t-test	0.000000	ND amplitudes	T5	Control
2f left	Paired Student's t-test	0.730467	PD amplitudes	T5	Control
2f right	Paired Student's t-test	0.000173	ND amplitudes	T5	Control
2g left	Paired Student's t-test	0.000000	DSIs	T5	Control
2g right	Paired Student's t-test	0.000003	DSIs	T5	Control
2i left	Paired Student's t-test	0.031593	PD amplitudes	T4	Control
2i right	Paired Student's t-test	0.000013	ND amplitudes	T4	Control
2j left	Paired Student's t-test	0.014008	PD amplitudes	T4	Control
2j right	Paired Student's t-test	0.009983	ND amplitudes	T4	Control
2k left	Paired Student's t-test	0.000000	DSIs	T4	Control
2k right	Paired Student's t-test	0.000018	DSIs	T4	Control
3a left	Paired Student's t-test	0.520257	PD amplitudes	T5	Block
3a right	Paired Student's t-test	0.287448	ND amplitudes	T5	Block
3b	Paired Student's t-test	0.227059	DSIs	T5	Block
3c	Paired Student's t-test	0.387859	Decay times	T5	Block
3d left	Student's t-test	0.003536	ΔDecay times	T5	Ctrl vs. Block
3d right	Student's t-test	0.000215	ΔDSIs	T5	Ctrl vs. Block
3f left	Paired Student's t-test	0.000007	ND amplitudes	T5	Control
3f right	Paired Student's t-test	0.287448	ND amplitudes	T5	Block
3g left	Paired Student's t-test	0.422686	PD amplitudes	T4	Block
3g right	Paired Student's t-test	0.036053	ND amplitudes	T4	Block
3h	Paired Student's t-test	0.049270	DSIs	T4	Block
3i	Paired Student's t-test	0.000158	Decay times	T4	Block
3j left	Student's t-test	0.085772	ΔDecay times	T4	Ctrl vs. Block
3j right	Student's t-test	0.560212	ΔDSIs	T4	Ctrl vs. Block
3k left	Paired Student's t-test	0.000323	ND amplitudes	T4	Control
3k right	Paired Student's t-test	0.036053	ND amplitudes	T4	Block

5 DISCUSSION

Seeing things move is so commonplace to us that we are not aware of the many processing steps that lie between the sensory cells in the retina and our awareness of a moving object. However, detecting visual motion is a task that visual systems in animals and humans alike have to solve based on the luminance signals of the photoreceptors. Spatio-temporal correlations in luminance imply visual motion. How the visual system is extracting and representing visual movement from luminance signals is a classical example of a neural computation, and has been under investigation, in both theory and experiment, for several decades. Throughout, the technologies that enable the study of neurons and neural circuits have been key to scientific progress. With the advent of electrophysiology, the first motion selective cells have been found in the frog retina and subsequently in the visual system of many other species. With two-photon calcium imaging and methods to record from and intervene with the circuit on a cell type-specific level, the computation of visual motion can today be probed with great accuracy. Nevertheless, only with the continuous development and improvement of tools will we eventually be able to fully understand the circuit.

In the course of this thesis, I have developed a technique that allows us to image the retinotopic plane of the *Drosophila* visual system *in vivo* during visual stimulation. Previously, it was only possible to record few cells at once, and the investigation of spatial relationships would have been tedious and ineffective. Imaging the retinotopic plane makes it possible to efficiently image of over 100 cells simultaneously, enables the study of spatial relationships and effectively reconstructs the fly's view of the surrounding world through the filter of a specific processing step. I applied this method to investigate the representation of natural scenes by L2, a cell type that feeds into the OFF motion pathway.

Natural scenes can vary substantially in their statistics, and a large body of evidence suggests that visual neurons dynamically adapt to the statistics of the surround. In the context of motion vision, how neural circuits adapt to robustly detect motion regardless of variations in the surround scenes has been a long-standing question. As a collaborative project, we have investigated the robustness of motion vision to varying contrast statistics. We could demonstrate that flies robustly respond to motion regardless of surround contrast, and that cells presynaptic to T4 and T5 normalize their response depending on surround contrast. Thus, we established that the fundamental computational principle of normalization also takes place in the motion vision circuit of the fly, shedding light on the complex dynamics of a comparably simple neural circuit.

Lastly, I have investigated properties of the local motion detectors T4 and T5 directly. Numerous recent efforts focused on investigating the computation of local motion on the dendrites of T4 and T5. In the course of this thesis, I could show that the local motion

signals of T4 and T5 are influenced by motion in the surround, partly due to the interactions between local motion detectors themselves. Thus, I could identify yet another layer of complexity that needs to be taken into account when attempting to understand motion vision. Overall, with this thesis, I could advance the technological options for studying the fly visual system, as well as contribute to our understanding of neuronal computation in the model circuit of fly motion vision.

5.1 TECHNOLOGICAL ADVANCES IN IMAGING

The last few years have seen a continued, rapid development in imaging techniques. Two-photon imaging has been improved in terms of speed, axial resolution and volumetric imaging options. Advanced methods of stimulation in conjunction with imaging have refined the possibilities for circuit intervention; miniature two-photon microscopes make it possible to record from freely behaving animals; improved fluorophores for sensing calcium, voltage and neurotransmitters allow the study of different signaling molecules; light-sheet microscopy enables whole-brain cellular resolution imaging in transparent species such as the larval zebrafish. In the following, I will discuss a number of advances in imaging technologies that are relevant for the investigation of neural circuits.

5.1.1 VOLUME IMAGING APPROACHES

A major area of development has been volume imaging, with the goal to image the 3D structure of neurons or neural populations. Previously, imaging was largely restricted to one plane in the sample, which clearly does not reflect the complexity and 3D structure of neuronal populations. To overcome this constraint, several alternative approaches for fast focal point scanning in three dimensions have emerged. As the simplest option, refocusing the objective lens can be controlled with a piezoelectric motor. For switching between a small number of imaging planes, reasonable scan speeds can be reached with this approach (Göbel and Helmchen, 2007). However, apart from the speed limitation due to high inertia of the objective lens, motion artefacts can be introduced when using water immersion lenses. Thus, the approaches for fast refocusing discussed in the following circumvent this problem altogether. Instead of moving the objective or the sample, the divergence of the beam entering the objective lens is modulated using different strategies, creating an axially offset focal spot in the sample.

Electrically tunable lenses (ETLs) have been employed for this purpose in a number of studies. Specific properties of an ETL, for example the refractive index or its curvature, change depending on the applied current, which alters the focal length of the lens. In other words, the beam divergence can be modulated by applying current. Consequently, feeding the divergent or convergent beam into the objective lens creates an axially shifted focal spot. Grewe et al. (2011) showed the applicability of ETLs to two-photon imaging by scanning two imaging planes in the mouse barrel cortex. Subsequently, ETLs have been employed for recording from several imaging planes with high scan speeds in different implementation

strategies (dal Maschio et al., 2017; Sheffield and Dombeck, 2015; Smirnov et al., 2017; Stirman et al., 2016).

Another method to modulate the wave front shape is with acusto-optic modulators (AODs). AODs are based on a principle whereby properties of light can be changed by acoustic waves. Acoustic waves at a specific frequency are created in a particular material, which changes the refractory index of this material. A light beam passing through will be deflected depending on the frequency of the sound waves. AODs can be used for moving the beam axially as well as laterally. A fundamental difference of AODs from other approaches for volume imaging is that the frequency of the generated acoustic wave can be changed in discrete steps (as opposed to mechanical movement or voltage, which change continuously). Thus, with AODs it is possible to change the position of the focal spot in discrete steps and randomly access different regions of interest without losing time on scanning the space in between. 3D random-access two-photon systems have been used to image different z planes as well as to image only selected locations with cell bodies or dendritic spines (Duemani Reddy et al., 2008; Rózsa et al., 2014) . The major advantage of AODs is their ability to circumvent inertia and to enable efficient imaging of sparse regions of interest (Ji et al., 2016). On the other hand, AODs widen the laser pulses, thus requiring the use of elaborate dispersion control systems and beam stabilization (Iyer et al., 2003; Svoboda and Yasuda, 2006). In addition, preselection of the target regions to be imaged is only possible in samples with little or no movement, which can be a problem for *in vivo* imaging.

Other approaches for volume imaging include imaging with ultrasound or acoustic gradient lenses (Darabi et al., 2018; Kong et al., 2015). These techniques change the beam divergence with a fixed oscillation frequency, similar to resonant galvanometer scanners. Spatial light modulators (SLMs) can also be used for quasi-volume imaging if the fluorescence is sparse, or spatially resolved detection systems are employed (W. Yang et al., 2016). SLMs for photostimulation will be discussed below. Likewise, imaging with axially elongated beams, called Bessel beams, provides signals from three dimensions; however, they are not resolved along the optical axis. Still, when fluorescence expression is sparse, some 3D information can also be gained by this method (Lu et al., 2017).

The majority of commercially available objective lenses are optimized for imaging of planes and are corrected for chromatic aberrations, these being the main requirements for most of microscopy (Ji et al., 2016). Thus, while these lenses enable diffraction-limited imaging in a large focal plane with different wavelengths, the lens characteristics are not usually optimal for forming a focal spot outside of the focal plane. A divergent or convergent wave front, created with AODs or ETLs, is therefore subject to spherical aberrations, especially at large distances from the focal plane. For two-photon microscopy, this is detrimental not only because the resolution outside of the focal plane is reduced, but, more importantly because

the two-photon effect depends on the high number of photons in a small area. A widened focal spot will cause efficiency to drop and phototoxic effects to increase (Ji et al., 2016).

The remote focusing approach that we have chosen for our imaging system circumvents this problem by deliberately introducing these aberrations at the remote objective lens, and then cancelling them out in the sample objective lens (Botcherby et al., 2008). Thus, with remote focusing, diffraction-limited focal spot sizes are possible across the entire imaging volume (Botcherby et al., 2012). Several requirements have to be met in order for the introduced aberrations to be cancelled out. First, ideally, the same model should be used for imaging and remote objective lens. If this is not possible due to the need for water immersion, objective lenses with the same general specifications will still produce near diffraction-limited resolution. Second, the back aperture of the remote lens has to be imaged onto the back aperture of the imaging lens, which is possible with a simple telescope in 4f configuration. Third, the magnification between remote and imaging objective has to be one (Botcherby et al., 2008). Remote focusing has been successfully employed in a number of configurations to scan several imaging planes in the mouse and zebrafish (Rupprecht et al., 2016; Sofroniew et al., 2016).

In conclusion, depending on the question in place, some volume imaging approaches will be favored over others, and no approach is equally suited for all application. For imaging of sparsely distributed, large cell bodies in the mammalian cortex or retina, speed and access to different areas can be prioritized over spatial resolution, and AOD systems or ETLs may be the method of choice. For small structures like dendritic spines or the structures in the fly visual system, the preferred systems will have diffraction-limited resolution in the entire volume, such as remote focusing. Yet other applications call for different systems that are targeted to the specific task at hand. To date no perfect imaging system without tradeoffs in either speed or resolution is available. Thus, continued advancement of volume imaging systems along several lines of development will be needed to eventually surmount the complexity and size of neuronal circuits.

5.1.2 OTHER ADVANCES IN MULTI-PHOTON MICROSCOPY AND LIGHT MICROSCOPY

With increasing scanning speeds in two-photon imaging, another factor is reaching its limits, the signal-to-background ratio. The time spent on each voxel by the laser focus is decreased with faster scan speeds, and fewer laser pulses excite the fluorophores in each voxel. This intrinsic limitation of point scanning approaches is circumvented by multiplexing methods, whereby several laser foci are created simultaneously, and the pixel dwell time is increased again by sampling from several points at once (Yang and Yuste, 2017). Here, several detectors have to be used to recover the signal from different areas (Kim et al., 2007). In an alternative approach called chromatic multiplexing, detectors for different wavelengths can recover the spatial information from the areas excited at the same time (Mahou et al., 2012).

For true volume imaging, that is the scanning of not only a few planes but the entire volume, these types of approaches may prove necessary, as either speed or signal-to-background ratio will be severely compromised otherwise.

The nature of microscopy traditionally limited *in vivo* imaging to head-restrained animals. However, two-photon imaging in freely behaving animals has been demonstrated with miniature two-photon headpieces for mice and rats (Helmchen et al., 2001), recently with a headpiece weighing only 2.5 g (Zong et al., 2017), and in combination with volume imaging (Ozbay et al., 2018). Furthermore, slow moving organisms like *Drosophila* larvae can now be tracked and imaged on-line (Karagyozov et al., 2018). These approaches allow us to observe the dynamics of the brain in freely behaving animals for the first time, which constitutes a big step in systemic neurosciences. However, the interpretation of neural recordings from behaving animals is challenging since the sensory input cannot be controlled for in the same way as for head-restrained animals.

Major advantages of two-photon microscopy, as compared to one-photon confocal or epifluorescence microscopes, are increased imaging depth and reduced scattering of excitation light. In near-transparent tissue with little scattering, however, these issues are less pronounced. In the transparent larval zebrafish, it is possible to record activity with cellular resolution from the entire brain using the light sheet microscopy (Ahrens et al., 2013; Power and Huisken, 2017). In this method, a sheet of light created with a cylindrical lens is exciting a thin plane in the sample, which is then imaged with a camera. Furthermore, high-resolution light field microscopy (where not only intensity information but also directional information of the light rays is captured) has recently been employed for whole-brain imaging in *Drosophila* with a near-cellular resolution when fluorophores are sparse (Aimon et al., 2019). This fast method can also be used for calcium imaging in freely moving zebrafish larvae (Cong et al., 2017). On the other hand, in two-photon microscopy, depth penetration is still limited to about 300 μm, and without surgical removal of occluding tissue this value is greatly decreased. Three-photon imaging provides an alternative here, as the excitation light can penetrate yet deeper into the tissue. Imaging in intact animals, through the mouse skull or *Drosophila* cuticle, has been demonstrated (Horton et al., 2013; Ouzounov et al., 2017; Tao et al., 2017).

In addition to the methodological improvements on the optical side, major advancements have been made in the development of fluorophores. For calcium imaging, new sensors with improved linearity and signal-to-noise ratio have been released recently (Inoue et al., 2019; Y. Yang et al., 2018). Moreover, genetically encoded fluorescent indicators that sense voltage for the first time allow the imaging of electrical potential in neural structures inaccessible to electrophysiology (Panzera and Hoppa, 2019; Platisa and Pieribone, 2018). The sensors Asap and Voltron have been successfully applied to study the *Drosophila* visual system (Abdelfattah et al., 2018; Chamberland et al., 2017; H. H. H. Yang et al., 2016). However, for most sensors, the signal-to-noise ratio is still significantly lower than for

calcium sensors. In addition to measuring the internal signaling of neurons in terms of calcium or voltage, fluorophores for the sensing of neurotransmitters are also in development (Wang et al., 2018). The genetically encoded glutamate sensor iGluSnFR as well as the acetylcholine sensor GAch have been applied in *Drosophila* (Jing et al., 2018; Richter et al., 2018). With sensors like these, it will be possible to investigate essential aspects of neural communication that were previously inaccessible, as for example neurotransmitter reuptake and other processes in the synaptic cleft.

5.1.3 OPTICAL CIRCUIT MANIPULATION

Light-sensitive ion channels allow the manipulation of a cell's potential by exposing them to light. This way, neural circuit components can be switched on and off rapidly during an experiment, which offers a crucial advantage over synaptic blocking tools that are continuously expressed in the animal or regulated by temperature. An array of genetically encoded optogenetic tools is now available for either depolarizing or hyperpolarizing neurons, activateable with different wavelengths (Chen et al., 2018). In visual neuroscience, care has to be taken that the applied light does not interfere with animal seeing it. For *Drosophila*, for example, Channelrhodopsin (Pulver 2009, Berndt 2009), GtACR (Govorunova et al., 2015) and Chrimson (Klapoetke et al., 2014) have been successfully applied to hyperpolarize or activate neurons (Haikala et al., 2013; Mauss et al., 2017a, 2014; Ribeiro et al., 2018; Strother et al., 2017).

For these applications, light was applied to the whole brain or brain hemisphere, and cell type specificity was reached by expressing the optogenetic tools only in a specific cell type. In order to selectively stimulate specific cells or sub-cellular compartments with light, several techniques have recently emerged (Chen et al., 2018). In the same way as delivering the laser focus for two-photon microscopy, lasers can be used for optogenetic stimulation, and they can be scanned across the sample with the same techniques. To simultaneously stimulate several cells in a complex spatial pattern, arbitrary 3D light patterns can be created with spatial light modulators (SLMs). SLMs change the phase of a laser wave front that is used as the optogenetic light and create holograms (Nikolenko, 2008). This method has been developed extensively over the last few years and it allows for circuit manipulation with unparalleled precision. Several systems have recently combined 3D two-photon imaging with 3D optogenetic stimulation to create an all-optical approach for studying neural circuits (dal Maschio et al., 2017; W. Yang et al., 2018).

In conclusion, these recent developments provide us with refined and flexible tools to both observe and interfere with a neural circuit. It is feasible now to simultaneously interact with a neural circuit while observing its dynamics with to date unseen precision, opening the doors for entirely new approaches to study the brain.

5.2 IMAGING RETINOTOPIC MAPS

The fundamental structural principle of visual systems is that they form two-dimensional maps, which represent visual space. The anatomy of the visual system of the fly as well as that of the mammalian retina display this principle beautifully. In the mammalian brain, map representations also dominate the visual cortex, as has been shown in the seminal studies of Hubel and Wiesel (Hubel and Wiesel, 1968, 1962, 1959). This section will discuss the possibilities we attain from being able to record from retinotopic maps, using the example of the retina.

In the mammalian retina, numerous types of retinal ganglion cells (RGCs) each tile the retina with the extent of their dendritic trees (Fig. 11a, Bae et al., 2018; Masland, 2012; Wässle et al., 1981). Physiologically, it became possible to record from the spatial array of RGCs already in the 1990s after the invention of microelectrode arrays, on which the retina could be mounted flat (Meister et al., 1994). With this method, the complete coverage of the visual space by RGCs as well as photoreceptors could also be shown functionally (Devries and Baylor, 1997; Gauthier et al., 2009). Using two-photon imaging, calcium signals can be recorded from the spatial maps of the retina (Denk and Detwiler, 1999) which eventually enabled the simultaneous recording of large numbers of RGCs, and led to the classification of over 30 functionally and morphologically different RGC types (Fig. 11b, Baden et al., 2016).

Extensive literature on the mammalian retina concerns the functional relationship between neighboring ganglion cells. As natural scenes are highly correlated both spatially and temporally, and an efficient way to code information contained in natural scenes would involve reducing these redundancies, as stated in an influential theory on efficient coding (Barlow, 1961). Conversely, redundancies in the neural code can reduce noise and therefore improve information transmission. In the mammalian retina, these hypotheses have been studied in depth. A multitude of studies describe both redundancies and noise correlations, as well as decorrelation mechanisms and how the population of neighboring cells can
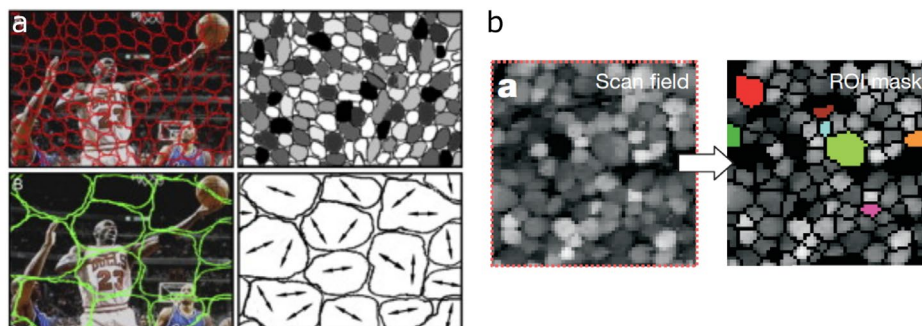

Figure 11 | Imaging the retinotopic array of ganglion cells. a) Different types of retinal ganglion cells tile the visual field with their dendritic coverage. Taken with permission from Masland (2012). b) Two-photon calcium imaging of retinal ganglion cells reveals over 30 different morphological and functional types. Taken with permission from Baden 2016).

encode information more efficiently than individual cells (e.g. Doi et al., 2012; Marre et al., 2012; Meister, 1995; Pillow et al., 2008; Puchalla et al., 2005; Simmons et al., 2013).

Another recent study concerned with the feature representation of ganglion cells found that based on spatial correlations between neighboring ganglion cells it is possible for a population to code for several features simultaneously (Kühn and Gollisch, 2019). This principle called multiplexing has also been described in OFF ganglion cells (Deny et al., 2017). In this study, it was shown that while fast OFF ganglion cells encode the position of close by bars, they encode bar velocity of distant bars, emphasizing the complex spatial. Comparisons to these studies in the light of direction encoding by T4 and T5 will be discussed in section 7.3.2.

These studies in the mammalian retina highlight the possibilities that we have through studying retinotopic planes in the visual system. It enables us to investigate spatial aspects of neural coding in *Drosophila* in a way that has not been possible previously. In this thesis, I exemplified these possibilities by studying the representation of natural scenes by L2. I showed that L2 codes linearly for local luminance in natural scenes, regardless of the surround statistics. Consistently, L2 is not modulated by surround contrast as we could show in Manuscript 2. Future experiments in medulla cell types will show whether their representation of natural scenes reflects their contrast normalization properties. In addition, they may uncover aspects of spatial coding that we have not previously considered, such as multiplexing.

5.3 TOWARD UNDERSTANDING MOTION VISION

Recent years have seen immense progress in the understanding of the cellular basis of motion vision. The cells presynaptic to the first motion selective cells, T4 and T5, have been identified, and for the main ones, their basic properties have been measured and their role in motion processing has been investigated. Still, it remains unclear how precisely motion is computed. Additional factors such as adaptive mechanisms, state-dependency, and interactions between local motion detectors complicate the picture. In the following, I will discuss our current understanding of motion detection in T4 and T5 in the light of the findings in this thesis, first the possible interactions between T4 and T5 and second the adaptive properties of the circuit.

5.3.1 MOTION COMPUTATION ON T4 AND T5 DENDRITES

Using electron microscopy, Mi1 and Tm3 were initially identified as presynaptic to T4 (Takemura et al., 2013). This was later expanded to include Mi4 and Mi9, as well as C3, CT1, TmY15, and T4 of the same subtype (Shinomiya et al., 2019; Takemura et al., 2017). These cell types account for more than 90% of the total number of synapses on T4, and they provide spatially offset input from different visual columns (Fig. 12a). In addition, the neurotransmitter phenotypes of these cell types could be identified, with Mi9 being
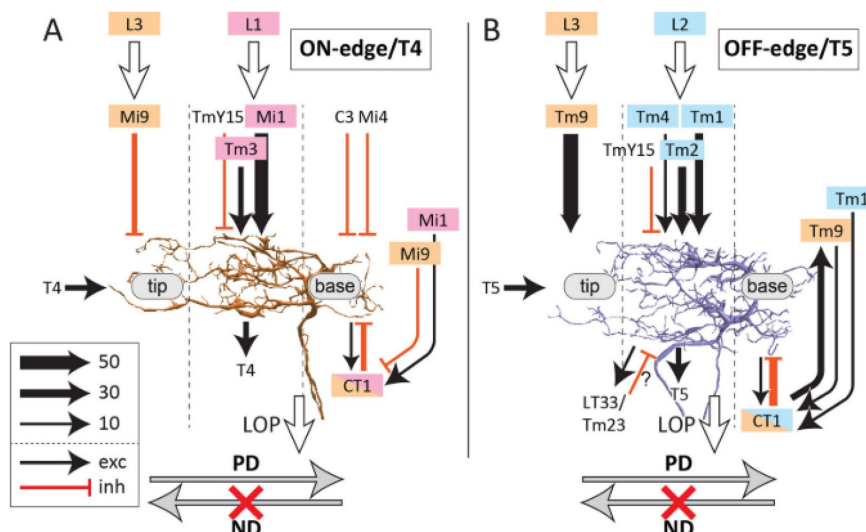
Figure 12 | Presynaptic elements to T4 and T5. a) The cell types presynaptic to T4 as identified by electron microscopy. The thickness of arrows corresponds to the number of synapses. The dashed lines roughly correspond to visual columns, indicating the spatial offset between different input cells. Colors indicate the presynaptic partners L1, L2 or L3. b) Analogous to a) for T5. Taken with permission from Shinomiya et al. (2019).

glutamatergic, Mi4, CT1 and C3 GABAergic, and Mi1 and Tm3 cholinergic (Pankova and Borst, 2016; Takemura et al., 2017). The response properties of Mi1, Tm3, Mi4, Mi9, as well as CT1 have been described (Arenz et al., 2017; Behnia et al., 2014; Meier and Borst, 2019). Except for Mi9, all cells depolarize to ON stimuli and they exhibit different temporal tuning, some showing band-pass, others low-pass characteristics. CT1, an individual amacrine cell spanning the entire medulla and lobula, has local response properties in each of its columnar neurite terminals, which each have pre- and postsynaptic sites (Meier and Borst, 2019). Alongside the identification of its input elements, the algorithm of motion computation on T4 has been investigated in several studies. Some studies, using apparent motion stimuli, report preferred-direction enhancement (PDE) as the mechanism that renders T4 direction selective (Fisher et al., 2015b). Other studies, using electrophysiology, find only null-direction suppression (NDS, Gruntman et al., 2018), and yet others, using apparent motion or white noise stimuli, or optogenetic activation experiments, find both (Haag et al., 2017, 2016; Leong et al., 2016; Strother et al., 2017). In addition, blocking the synaptic output as well as optogenetic activation of Mi1, Tm3, Mi4 and Mi9 reveals differential contributions of each of the cell types (Ammer et al., 2015; Strother et al., 2017) to motion detection.

The cell types presynaptic to T5 could be identified as Tm1, Tm2, Tm4, Tm9, CT1, TmY15 as well as LT33 and TM23 (Fig. 12b Shinomiya et al., 2019, 2014). All input cells except for CT1 are cholinergic (Shinomiya et al., 2019). Response properties of Tm1, Tm2, Tm4, Tm9, as well as CT1, have been described, showing that all the cells respond to OFF stimuli, with different temporal and spatial response characteristics (Behnia et al., 2014; Meier et al., 2014; Serbe et al., 2016). As to the algorithm for motion detection that is implemented in T5, both PDE and NDS were found in two studies, using calcium imaging in conjunction with

apparent motion stimuli or white noise (Haag et al., 2017; Leong et al., 2016). Another study using voltage indicators proposed a model whereby voltage signals are summed up linearly in T5, and only the non-linear relationship between calcium and voltage leads to direction-selective calcium responses (Wienecke et al., 2018). When blocking synaptic output of Tm1, Tm2, Tm4 or Tm9, or combinations thereof, each of the cell type shows contributions to motion detection in LPTCs (Fisher et al., 2015a; Serbe et al., 2016).

In essence, it is likely that for T4 as well as for T5, both PDE and NDS take place under certain conditions. However, to date it is not clear what these conditions are. Additionally, while each of the input elements evidently shapes motion computation to some extent, the role of each of them in terms of the algorithmic description is not yet resolved. Overall, therefore, the picture of motion detection that starts to emerge in the light of these findings is a more complex one than what the current algorithmic models portray. As an extreme case, it is conceivable that entirely different mechanisms for motion detection are in place under varying conditions, as, for example, different intensity, color, contrast or velocity regimes, or distinct states of the animal, such as during walking or flying. Different cell types may mediate and modulate motion computation under these diverse circumstances. In the following sections, I will discuss the consequences for local motion computation of two aspects that I investigated in the course of this thesis: First, in the light of possible interactions between T4 and T5 cells; and, second, in the light of contrast normalization.

5.3.2 INTERACTIONS BETWEEN LOCAL MOTION DETECTORS

As part of this thesis, I demonstrated that T4 and T5 responses to local motion differ from those to global motion. Consequently, I showed that for T5, these differences are mediated by interactions between T4/T5. It is important to note that these results do not determine whether the interaction takes place between T4/T5 cells of the same subtype (that is, e.g. T4c interacting with T4c), or between different subtypes (and if so, which ones). Dendro-dendritic synaptic connections between T4/T5 of the same subtype in the medulla and lobula, respectively, could in principle mediate the effect observed (Shinomiya et al., 2019; Takemura et al., 2017). However, it is also possible that T4/T5 interact via interneurons, or via direct synaptic contacts, in the lobula plate, where the connectome is not yet resolved. In this section, I will first discuss possible synaptic connections that could account for these observations. Subsequently, I will compare these results to other studies that blocked synaptic output of T4 and T5 while simultaneously recording from them, as well as to a study that also suggests the presence of T4/T5 interactions. Lastly, I will discuss the possibility that T4 and T5 encode stimulus properties more complex than local motion.

The synaptic connections between T4/T5 of the same subtype are directed along the PD and are cholinergic. These synapses should therefore be most relevant for PD stimuli, to which the response of individual T4 or T5 cells is high. However, the difference in direction selectivity between local and global stimulation is due to elevated ND responses. Thus, it is unlikely that these synapses mediate the observed effect, namely elevated ND responses to

local motion compared to global motion. Alternatively, a candidate cell type that could potentially mediate the feedback between T4 and T5 is TmY15. This GABAergic cell was identified in the most recent electron microscopy study (Shinomiya et al., 2019). It receives input from a variety of medulla cells, as well as from T4 and T5 in the lobula plate layers b and c, and it is presynaptic to T4 and T5 in the medulla and lobula, respectively. While TmY15's synaptic connections suggest that it provides direction selective feedback signals, it is not clear how this inhibitory cell can mediate the effects we observed. Thus, further experiments will have to elucidate the role of this cell type. Another option that could explain the observed effects involves interactions between T4 and T5 in the lobula plate. In the lobula plate, LPi cells receive input from T4 and T5 of a specific subtype, and project to the layer with opposite PD, where they inhibit LPTCs (Mauss et al., 2015). Their dendritic arbors span several columns (Klapoetke et al., 2017), and their receptive fields are slightly larger than those of T4 and T5 (Richter et al., 2018). Thus, they are likely to pool T4 and T5 signals from an area larger than the local stimulus we used. If LPis would inhibit T4 and T5 cells of the opposite PD, they could account for elevated ND responses during local stimulation, when these inhibitory signals would be smaller than during global stimulation. However, it is unclear whether an interaction on the level of the lobula plate would at all manifest in the calcium signals that we observed in the dendrites of T4 and T5.

Several other studies have blocked synaptic output of T4 and T5 while imaging from one of them. In Haag et al. (2016), T4 and T5 directional tunings were measured in control flies as well as flies with T4 and T5 blocked using TNT. Stimulating the cells with full field square wave gratings, no obvious differences between the responses of control and T4/T5 output block flies were seen, even though, strikingly, the small ND response in control flies seemed to disappear entirely in block flies. Considering the results presented in this thesis, it is likely that responses to full field grating stimuli would not reveal the interactions between T5 cells. In another study by Strother et al. (2017), T4 responses to flashes and local edge stimuli were measured in control flies as well as in flies with T4 and T5 output blocked. This study used a line containing all subtypes of T4. Thus, when recording from the dendritic area, they could not distinguish responses of the different subtypes of T4, and in that, PD or ND responses. Interestingly, however, the temporal dynamics of the edge response changed slightly when T4 and T5 output was blocked, consistent with our results. Overall, while these studies reported no differences in responses of T4 and T5 during the simultaneous block of their synaptic output, it is likely that this is due to the choice of stimulus and experimental parameters.

Another recent study by Barnhart et al. (2018) investigated the integration of T4 and T5 signals on HS cells. In their study, they used a stimulus where two adjacent local edges moved simultaneously. Strikingly, they found that, when compared to only a single local edge moving, T4 responses are heavily suppressed in a certain spatial combination of the two edges. Specifically, the response of a T4 cell to a local PD edge was suppressed by a

simultaneously moving local edge on its null side (for T4c, this corresponds to an edge moving at a location slightly higher in elevation). These experiments were not performed while blocking T4 and T5 output, and therefore it cannot be excluded that presynaptic mechanisms can account for this effect. Nevertheless, they are another indication for possible interactions between local motion detectors.

In the mouse retina, it was recently shown that even stimuli at a large distance from a RGC's receptive field can elicit responses of that cell (Deny et al., 2017). This result emphasizes that, while cells may appear to process only local information based on their linear receptive field, it is still possible that they also respond to stimuli in the surround. Moreover, the study showed that, while fast OFF RGCs code for bar position when the bar is close to their RF, they code for bar velocity when the bar is distant. Thus, in this case, an individual cell codes for several visual features at the same time. Moreover, as another study highlights, while a cell type may encode a specific visual feature, such as motion, it does not usually respond exclusively to its preferred stimulus (Kühn and Gollisch, 2019). T4 and T5, for instance, also respond substantially to luminance flicker. Taking together T5 responses from the contrast normalization and the surround motion experiments, an implicit result from this thesis is that T5 responses to global PD gratings can be smaller than its responses to local ND gratings under certain contrast conditions. T5 PD responses to local gratings (no surround contrast) are twice as large as its responses to global gratings (full surround contrast), as I showed in Manuscript 2. In addition, T5 responses to local ND motion, as in Manuscript 3, showed that the amplitude of ND responses amounts to about half the amplitude of PD responses. Taken together, remarkably, this shows that the output signal of an individual T5 cell cannot unambiguously code for direction of local motion. It is conceivable that, only as a population, T4 and T5 signals will actually encode visual motion, and future experiments in this direction will have to systematically investigate this phenomenon. With our newly acquired possibility to image the retinotopic array of T4 and T5 cells, experiments using natural movies as in Kuehn et al. (2019) may reveal novel encoding properties of T4 and T5.

5.3.3 CONTRAST NORMALIZATION IN THE MOTION VISION CIRCUIT

As part of this thesis, we could establish that neurons in the *Drosophila* motion vision circuit perform contrast normalization, and that this renders the optomotor response as well as LPTC velocity estimation robust to the wide contrast fluctuations observed in natural scenes. In the following, I will discuss selected aspects of this phenomenon.

We showed that the normalization signal that cells in the medulla receive is at least in part formed by pooling the signals of the same cell type. The self-evident question arising from these results is then: Which cell type constitutes this spatial pool, and, is there a specific pool cell for each medulla cell type performing contrast normalization, or is there a common pool for several or all of them? A candidate pool cell has to sample signals from several columns of medulla cells, and in turn provide the spatially combined output back to them.

Several wide-field neurons in the medulla have been described that have dendritic trees spanning several columns up to covering the entire medulla (Fischbach and Dittrich, 1989), and genetic tools to access several of these cell types are available (Nern et al., 2015). It will thus be the goal of future studies to further characterize the properties of spatial pooling and to identify the cells underlying the mechanism of contrast normalization.

The signals of T4 and T5 depend on surround contrast in a way that is qualitatively similar to the medulla cells. However, importantly, their sensitivity to surround contrast differs: While T4 cell responses saturate already at relatively low contrasts, T5 responses are less sensitive, such that even at the highest contrast of our stimulation arena, T5 signals are still suppressed by surround contrast. Remarkably, the cells presynaptic to T4 that are subject to contrast normalization, namely, Mi1 and Tm3, exhibit contrast sensitivity similar to that of T4, while the cells presynaptic to T5 that are subject to contrast normalization, Tm1, Tm2 and Tm4, exhibit contrast tunings similar to T5. This suggests that T4 and T5 inherit these properties from their presynaptic partners.

As a consequence of the differential sensitivities of T4 and T5 to local contrast, and together with other known asymmetries in T4 and T5 signaling (Leonhardt et al., 2016), their contributions to the motion signal of LPTCs are not necessarily balanced. Instead, under certain contrast or velocity regimes, LPTCs might receive predominantly only T4 or only T5 input. While the significance of this finding for the motion coding of LPTCs is uncertain, it might reflect fundamental statistics of natural scenes, as is the case with the asymmetrical velocity tuning of ON and OFF pathways.

In the blowfly, Reichardt et al. (1983) observed that the responses of LPTCs as well as optomotor responses initially increase in strength with increasing area of a grating stimulus, but eventually saturate at a certain level. The level at which the responses saturate depends on the grating velocity. In this study, a gain control mechanism involving a pool cell was proposed by which this phenomenon could be explained. Borst et al. (1995) proposed an alternative mechanism to explain these data, involving nonlinear dendritic integration of local motion signals. In these studies, it was assumed that local motion detectors act as independent units that are not modulated by properties of the surround; in other words, that T4/T5 cells respond the same, regardless of stimulus area. Taking into account the findings of this thesis, both in terms of contrast gain control in the medulla and in terms of interactions between local motion detectors, it becomes clear that the allegedly local signals received by LPTCs already intricately reflect properties of the spatial surround. A grating stimulus increasing in area will increase surround contrast and thus suppress local T4/T5 responses. Summing up T4 and T5 responses will produce a sub-linear response function with respect to stimulus area, qualitatively similar to the LPTC responses. Thus, it is possible that saturating LPTC responses can be explained the responses of their inputs alone. The dendritic integration of T4 and T5 signals on LPTCs has recently been addressed experimentally by Barnhart et al. (2018). The study showed that both interactions in the

presynaptic circuit (possibly between local motion detectors, as discussed above), as well as non-linear integration of T4 and T5 signals, shape the response properties of LPTCs. In summary, the output signals of LPTCs that trigger motion-based behaviors are created by a complex combination of mechanisms that will require further research.

5.3.4 CLOSING THE LOOP

Our understanding of motion encoding in T4 and T5 and LPTCs is crucially dependent on what these signals are used for in the fly, how they are processed downstream, and how they are influenced by the statistics of ego-motion. Several recent studies have addressed motion vision using a more systemic approach, and I will briefly discuss them here.

The visual motion perceived (and expected) during walking or flying is in stark contrast to motion signals that may be perceived while standing still. Specifically, during walking or flying the expected visual motion signals will be faster than during inactivity. The neurotransmitter octopamine is involved in modulating circuits in a state-dependent manner (Suver et al., 2012). Arenz et al. (2017) found that the temporal characteristics of T4 and T5 cells and their presynaptic partners change, when the octopaminergic system is activated by means of an octopamine agonist. Strother et al. (2018) found that some of the input cells to T4 are modulated by walking. These results shows that the motion vision circuit shifts its sensitivities depending on the expected visual input during flight or walking.

Flies perform saccades - fast, abrupt movements - when flying and walking, both in terms of their head and body movements. Saccades are interspersed with fixation periods of little or no movement. During a saccade, the perceived visual motion is very fast for a short amount of time. In a group of studies, the effect of saccades onto the motion signals of LPTCs has been investigated (Kim et al., 2017, 2015; Schnell et al., 2014). An efferent copy, that is, a copy of the motor signal eliciting the saccade, is inhibiting LPTCs, and thus suppressing motion signaling during saccades. This inhibition is precisely tuned to the amplitude of the saccade thus also effectively suppresses the optomotor response. During walking, LPTC cell signals have been found to represent forward walking and turning speeds as well as visual motion, suggesting that their signal in fact codes for the movement of the body through space (Fujiwara et al., 2017).

In summary, these studies point to the importance of studying a neural circuit while it is engaged in the tasks that it was constructed to perform. Only by closing the loop between sensory input and behavior will it be possible to fully uncover the dynamics of neural circuits in the behaving animal.

5.4 CONCLUSION AND OUTLOOK

Recent years have seen an immense progress in the field of fly motion vision. Eventually, these findings will contribute to a better understanding of the principles that govern neural computation in general. Until we can say that we have fully unraveled this neural circuit,

however, a long way still lies ahead. This becomes apparent considering that the recent results, in particular the connectome, reveal a substantial complexity of the circuitry. In this thesis, I could pinpoint and investigate two phenomena that illustrate how motion computation is shaped by factors such as surround contrast and surround motion. However, many additional elements that influence motion detection are still to be investigated. After all, it is still largely unclear how each cell presynaptic to T4 and T5 is contributing to motion computation. Progress was possible to a large extent because of the tools that were developed to study the neural circuits of the fly. Most of the recent work was only possible owing to calcium imaging, two-photon microscopy and genetic tools. In this thesis, I developed a novel technique that enables us to study the retinotopic array of cells that constitute visual maps. For example, showing natural movies, we might uncover unexpected spatial processing properties. Overall, with this method, we are able to investigate the spatial relationships of the neurons in the visual system and reconstitute the fly's view of the surroundings through the filter of specific cell types, which will hopefully add to our understanding of motion vision and neural computation in the future.

6 REFERENCES

Abdelfattah A, Kawashima T, Singh A, Novak O, Liu H, Shuai Y, Huang Y-C, Grimm J, Patel R, Friedrich J, Mensh B, Paninski L, Macklin J, Podgorski K, Lin B-J, Chen T-W, Turner G, Liu Z, Koyama M, Svoboda K, Ahrens M, Lavis L, Schreiter E. 2018. Bright and photostable chemigenetic indicators for extended in vivo voltage imaging. *bioRxiv*.

Abdeljalil J, Hamid M, Abdel-Mouttalib O, Stéphane R, Raymond R, Johan A, José S, Pierre C, Serge P. 2005. The optomotor response: A robust first-line visual screening method for mice. *Vision Res* **45**:1439–1446.

Ahrens MB, Orger MB, Robson DN, Li JM, Keller PJ. 2013. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat Methods* **10**:413–420.

Aimon S, Katsuki T, Jia T, Grosenick L, Broxton M, Deisseroth K, Sejnowski TJ, Greenspan RJ. 2019. Fast near-whole-brain imaging in adult Drosophila during responses to stimuli and behavior. *PLoS Biol* **17**:e2006732.

Ammer G, Leonhardt A, Bahl A, Dickson BJ, Borst A. 2015. Functional specialization of neural input elements to the Drosophila ON motion detector. *Curr Biol* **25**.

Arenz A, Drews MS, Richter FG, Ammer G, Borst A. 2017. The temporal tuning of the Drosophila motion detectors is determined by the dynamics of their input elements. *Curr Biol* **27**:929–944.

Baden T, Berens P, Franke K, Roman-Roson M, Bethge M, Euler, Román Rosón M, Bethge M, Euler T. 2016. The functional diversity of retinal ganglion cells in the mouse. *Nature* **529**:345–350.

Bae JA, Mu S, Kim JS, Turner NL, Tartavull I, Kemnitz N, Jordan CS, Norton AD, Silversmith WM, Prentki R, Sorek M, David C, Jones DL, Bland D, Sterling ALR, Park J, Briggman KL, Seung HS. 2018. Digital museum of retinal ganglion cells with dense anatomy and physiology. *Cell* **173**:1293–1306.

Bahl A, Ammer G, Schilling T, Borst A. 2013. Object tracking in motion-blind flies. *Nat Neurosci* **16**:730–738.

Bahl A, Serbe E, Meier M, Ammer G, Borst A, Bahl A, Serbe E, Meier M, Ammer G, Borst A. 2015. Neural mechanisms for Drosophila contrast vision. *Neuron* **88**:1240–1252.

Barlow HB. 1961. Possible principles underlying the transformations of sensory messages In: Rosenblith WA, editor. Sensory Communication. Cambridge: MIT Press. pp. 216–234.

Barlow HB, Levick WR. 1965. The mechanism of directionally selective units in rabbit's retina. *J Physiol* **178**:477–504.

Behnia R, Clark DA, Carter AG, Clandinin TR, Desplan C. 2014. Processing properties of ON and OFF pathways for Drosophila motion detection. *Nature* **512**:427–430.

Bellen HJ, Tong C, Tsuda H. 2010. 100 years of Drosophila research and its impact on vertebrate neuroscience: a history lesson for the future. *Nat Rev Neurosci* **11**:514–522.

Borst a, Haag J. 2002. Neural networks in the cockpit of the fly. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **188**:419–37.

Borst A, Egelhaaf M, Haag J. 1995. Mechanisms of dendritic integration underlying gain control in fly motion-sensitive interneurons. *J Comput Neurosci* **2**:5–18.

Borst A, Helmstaedter M. 2015. Common circuit design in fly and mammalian motion vision. *Nat Neurosci* **18**:1067–1076.

Botcherby EJ, Juškaitis R, Booth MJ, Wilson T. 2008. An optical technique for remote focusing in microscopy. *Opt Commun* **281**:880–887.

Botcherby EJ, Juškaitis R, Booth MJ, Wilson T, Juskaitis R, Booth MJ, Wilson T. 2007. Aberration-free optical refocusing in high numerical aperture microscopy. *Opt Lett* **32**:2007–2009.

Botcherby EJ, Smith CW, Kohl MM, Débarre D, Booth MJ, Juškaitis R, Paulsen O, Wilson T. 2012. Aberration-free three-dimensional multiphoton imaging of neuronal activity at kHz rates. *Proc Natl Acad Sci USA* **109**:2919–24.

Boyden ES, Zhang F, Bamberg E, Nagel G, Deisseroth K. 2005. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat Neurosci* **8**:1263–1268.

Braitenberg V. 1967. Patterns of projection in the visual system of the fly. I. Retina-lamina projections. *Exp brain*

Res **3**:271–98.

Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **415**:401–415.

Buchner E, Buchner S, Bülthoff I. 1984. Deoxyglucose mapping of nervous activity induced in Drosophila brain by visual movement. *J Comp Physiol A* **155**:471–483.

Carandini M, Heeger D. 2012. Normalization as a canonical neural computation. *Nat Rev Neurosci* **13**:51–62.

Chalfie M, Tu Y, Ghia Euskirchen, Ward WW, Pracher DC. 1994. Green fluorescent protein as a marker for gene expression. *Science* **263**:802–805.

Chamberland S, Yang HH, Pan MM, Evans SW, Guan S, Chavarha M, Yang Y, Salesse C, Wu H, Wu JC, Clandinin TR, Toth K, Lin MZ, St-Pierre F. 2017. Fast two-photon imaging of subcellular voltage dynamics in neuronal tissue with genetically encoded indicators. *Elife* **6**:e25690.

Chen IW, Papagiakoumou E, Emiliani V. 2018. Towards circuit optogenetics. *Curr Opin Neurobiol* **50**:179–189.

Chen T-W, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, Looger LL, Svoboda K, Kim DS. 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**:295–300.

Cheng A, Gonçalves JT, Golshani P, Arisaka K, Portera-Cailliau C. 2011. Simultaneous two-photon calcium imaging at different depths with spatiotemporal multiplexing. *Nat Methods* **8**:139–142.

Clark DA, Bursztyn L, Horowitz MA, Schnitzer MJ, Clandinin TR. 2011. Defining the computational structure of the motion detector in Drosophila. *Neuron* **70**:1165–1177.

Cong L, Wang Z, Chai Y, Hang W, Shang C, Yang W, Bai L, Du J, Wang K, Wen Q. 2017. Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (Danio rerio). *Elife* **6**:e28158.

Cremer C, Cremer T. 1978. Considerations on a laser-scanning-microscope with high resolution and depth of field. *Microsc Acta* **81**:31–44.

dal Maschio M, Donovan JC, Helmbrecht TO, Baier H. 2017. Linking neurons to network function and behavior by two-photon holographic optogenetics and volumetric imaging. *Neuron* **94**:774–789.

Darabi A, Zareei A, Alam MR, Leamy MJ. 2018. Broadband bending of fexural waves: acoustic shapes and patterns. *Sci Rep* **8**:11219.

Deisseroth K, Hegemann P. 2017. The form and function of channelrhodopsin. *Science* **357**:eaan5544.

Denk W, Delaney KR, Gelperin A, Kleinfeld D, Strowbridge BW, Tank DW, Yuste R. 1994. Anatomical and functional imaging of neurons using 2-photon laser scanning microscopy. *J Neurosci Methods* **54**:151–162.

Denk W, Detwiler PB. 1999. Optical recording of light-evoked calcium signals in the functionally intact retina. *Proc Natl Acad Sci* **96**:7035–7040.

Denk W, Strickler JH, Webb WW. 1990. Two-photon laser scanning fluorescence microscopy. *Science* **248**:73–76.

Deny S, Ferrari U, Macé E, Yger P, Caplette R, Picaud S, Tkačik G, Marre O. 2017. Multiplexed computations in retinal ganglion cells of a single type. *Nat Commun* **8**:1964.

Devries SH, Baylor DA. 1997. Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *J Neurophysiol* **78**:2048–2060.

Doi E, Gauthier JL, Field GD, Shlens J, Sher A, Greschner M, Machado TA, Jepson LH, Mathieson K, Gunning DE, Litke AM, Paninski L, Chichilnisky EJ, Simoncelli EP. 2012. Efficient coding of spatial information in the primate retina. *J Neurosci* **32**:16256–16264.

Duemani Reddy G, Kelleher K, Fink R, Saggau P. 2008. Three-dimensional random access multiphoton microscopy for functional imaging of neuronal activity. *Nat Neurosci* **11**:713–20.

Durst ME, Zhu G, Xu C. 2006. Simultaneous spatial and temporal focusing for axial scanning. *Opt Express* **14**:12243–12254.

Dvorak DR, Bishop LG, Eckert HE. 1975a. Intracellular recording and staining of directionally selective motion detecting neurons in fly optic lobe. *Vision Res* **15**:451–453.

Dvorak DR, Bishop LG, Eckert HE. 1975b. On the identification of movement detectors in the fly optic lobe. *J Comp Physiol A* **100**:5–23.

References

Fan GY, Fujisaki H, Miyawaki A, Tsay RK, Tsien RY, Ellisman MH. 1999. Video-rate scanning two-photon excitation fluorescence microscopy and ratio imaging with cameleons. *Biophys J* **76**:2412–2420.

Fermi G, Reichardt W. 1963. Optomotorische Reaktionen der Fliege Musca Domestica. *Kybernetik* **2**:15–28.

Fischbach K-F, Dittrich APM. 1989. The optic lobe of Drosophila melanogaster. I . A Golgi analysis of wild-type structure. *Cell Tissue* **258**:441–475.

Fisher YE, Leong JCS, Sporar K, Ketkar MD, Gohl DM, Clandinin TR, Silies M. 2015a. A class of visual neurons with wide-field properties is required for local motion detection. *Curr Biol* **25**.

Fisher YE, Silies M, Clandinin TR, Fisher YE, Silies M, Clandinin TR. 2015b. Orientation selectivity sharpens motion detection in Drosophila. *Neuron* **88**:390–402.

Freifeld L, Clark DA, Schnitzer MJ, Horowitz MA, Clandinin TR. 2013. GABAergic lateral interactions tune the early stages of visual processing in Drosophila. *Neuron* **78**:1075–1089.

Fujiwara T, Cruz TL, Bohnslav JP, Chiappe ME. 2017. A faithful internal representation of walking movements in the Drosophila visual system. *Nat Neurosci* **20**:72–81.

Gauthier JL, Field GD, Sher A, Greschner M, Shlens J, Litke AM, Chichilnisky EJ. 2009. Receptive fields in primate retina are coordinated to sample visual space more uniformly. *PLoS Biol* **7**:747–755.

Göbel W, Helmchen F. 2007. New angles on neuronal dendrites in vivo. *J Neurophysiol* **98**:3770–3779.

Goetz KG. 1964. Optomotorische Untersuchung des visuellen Systems einiger Augenmutanten der Fruchtfliege Drosophila. *Kybernetik* **2**:77–92.

Golic KG. 1991. Site-specific recombination between homologous chromosomes in Drosophila. *Science* **252**:958–961.

Golic KG, Lindquist S. 1989. The FLP recombinase of yeast catalyzes site-specific recombination in the drosophila genome. *Cell* **59**:499–509.

Göppert-Mayer M. 1931. Über Elementarakte mit zwei Quantensprüngen. *Ann Phys* **9**:273–294.

Gosnell TR, Taylor AJ. 1991. Selected papers on ultrafast laser technology. United States: SPIE Optical Engineering Press.

Govorunova EG, Sineshchekov OA, Janz R, Liu X, Spudich JL. 2015. Natural light-gated anion channels: A family of microbial rhodopsins for advanced optogenetics. *Science* **349**:647–650.

Grewe BF, Voigt FF, van 't Hoff M, Helmchen F. 2011. Fast two-layer two-photon imaging of neuronal cell populations using an electrically tunable lens. *Biomed Opt Express* **2**:2035–2046.

Grienberger C, Konnerth A. 2012. Imaging calcium in neurons. *Neuron* **73**:862–85.

Gruntman E, Romani S, Reiser MB. 2018. Simple integration of fast excitation and offset, delayed inhibition computes directional selectivity in Drosophila. *Nat Neurosci* **21**:250–257.

Guild JB, Xu C, Webb WW. 1997. Measurement of group delay dispersion of high numerical aperture objective lenses using two-photon excited fluorescence. *Appl Opt* **36**:397–401.

Haag J, Arenz A, Serbe E, Gabbiani F, Borst A. 2016. Complementary mechanisms create direction selectivity in the fly. *Elife* **5**:e17421.

Haag J, Mishra A, Borst A. 2017. A common directional tuning mechanism of Drosophila motion-sensing neurons in the ON and in the OFF pathway. *Elife* **6**:e29044.

Haikala V, Joesch M, Borst A, Mauss AS. 2013. Optogenetic control of fly optomotor responses. *J Neurosci* **33**:13927–13934.

Hamill OP, Marty A, Neher E, Sakmann B, Sigworth FJ. 1981. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Eur J Physiol* **391**:85–100.

Hardie RC. 1991. Whole-cell recordings of the light induced current in dissociated Drosophila photoreceptors: evidence for feedback by calcium permeating the light-sensitive channels. *Proc R Soc London* **245**:203–210.

Hardie RC. 1989. A histamine-activated chloride channel involved in neurotransmission at a photoreceptor synapse. *Nature* **339**:704–706.

Hardie RC, Raghu P. 2001. Visual transduction in Drosophila. *Nature* **15**:186–193.

References

Hassenstein B. 1951. Ommatidienraster und afferente Bewegungsintegration. Versuche an dem Rüsselkäfer Chlorophanus viridis. *Z Vgl Physiol* **33**:301–326.

Hassenstein B, Reichardt W. 1956. Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. *Zeitschrift für Naturforsch* **11**:513–524.

Hausen K. 1984. The lobula-complex of the fly: Structure, function and significance in visual behavior. In: Ali MA, editor. Photoreception and Vision in Invertebrates. New York: Plenum Press. pp. 532–559.

Hausen K. 1982. Motion sensitive interneurons in the optomotor system of the fly - I. The horizontal cells: Structure and signals. *Biol Cybern* **45**:143–156.

Hausen K. 1976. Functional characterization and anatomical identification of motion sensitive neurons in the lobula plate of the blowfly Calliphora erythrocephala. *Zeitschrift fur Naturforsch* **31**:629–633.

Heisenberg M, Buchner E. 1977. The rôle of retinula cell types in visual behavior of Drosophila melanogaster. *J Comp Physiol A* **117**:127–162.

Heisenberg M, Wolf R. 1984. The Compound Eye In: Heisenberg M, Wolf R, editors. Vision in Drosophila. Berlin, Heidelberg: Springer-Verlag. pp. 10–24.

Helmchen F, Denk W. 2005. Deep tissue two-photon microscopy. *Nat Methods* **2**:932–940.

Helmchen F, Fee MS, Tank DW, Denk W. 2001. A miniature head-mounted two-photon microscope. *Neuron* **31**:903–912.

Hengstenberg R. 1982. Common visual response properties of giant vertical cells in the lobula plate of the blowfly Calliphora. *J Comp Physiol A* **149**:179–193.

Hofbauer A, Campos-Ortega JA. 1990. Proliferation pattern and early differentiation of the optic lobes in Drosophila melanogaster. *Roux's Arch Dev Biol* **198**:264–274.

Horton NG, Wang K, Kobat D, Clark CG, Wise FW, Schaffer CB, Xu C. 2013. In vivo three-photon microscopy of subcortical structures within an intact mouse brain. *Nat Photonics* **7**:205–209.

Hubel D, Wiesel T. 1968. Receptive fields and functional architecture of monkey striate cortex. *J Physiol* **195**:215–243.

Hubel DH, Wiesel TN. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* **160**:106–154.

Hubel DH, Wiesel TN. 1959. Receptive fields of single neurones in the cat's striate cortex. *J Physiol* **148**:574–591.

Inoue M, Takeuchi A, Manita S, Horigane S, Sakamoto M, Kawakami R, Yamaguchi K, Otomo K, Yokoyama H, Kim R, Yokoyama T, Takemoto-Kimura S, Abe M, Okamura M, Kondo Y, Quirin S, Ramakrishnan C, Imamura T, Sakimura K, Nemoto T, Kano M, Fujii H, Deisseroth K, Kitamura K, Bito H. 2019. Rational engineering of XCaMPs, a multicolor GECI suite for in vivo imaging of complex brain circuit dynamics. *Cell* **177**.

Iyer V, Losavio BE, Saggau P. 2003. Compensation of spatial and temporal dispersion for acousto-optic multiphoton laser-scanning microscopy. *J Biomed Opt* **8**:460–471.

J. Nakai, M. Ohkura, K. Imoto. 2001. A high signal-to-noise Ca2+ probe composed of a single green fluorescent protein. *Nat Biotechnol* **3**:137–141.

Jenett A, Rubin GM, Ngo TTB, Shepherd D, Murphy C, Dionne H, Pfeiffer BD, Cavallaro A, Hall D, Jeter J, Iyer N, Fetter D, Hausenfluck JH, Peng H, Trautman ET, Svirskas RR, Myers EW, Iwinski ZR, Aso Y, DePasquale GM, Enos A, Hulamm P, Lam SCB, Li HH, Laverty TR, Long F, Qu L, Murphy SD, Rokicki K, Safford T, Shaw K, Simpson JH, Sowell A, Tae S, Yu Y, Zugates CT. 2012. A GAL4-Driver Line Resource for Drosophila Neurobiology. *Cell Rep* **2**:991–1001.

Ji N, Freeman J, Smith SL. 2016. Technologies for imaging neural activity in large volumes. *Nat Neurosci* **19**:1154–1164.

Jing M, Zhang P, Wang G, Feng J, Mesik L, Zeng J, Jiang H, Wang S, Looby JC, Guagliardo NA, Langma LW, Lu J, Zuo Y, Talmage DA, Role LW, Barrett PQ, Zhang LI, Luo M, Song Y, Zhu JJ, Li Y. 2018. A genetically encoded fluorescent acetylcholine indicator for in vitro and in vivo studies. *Nat Biotechnol* **36**:726–737.

Joesch M, Plett J, Borst A, Reiff DF. 2008. Response properties of motion-sensitive visual interneurons in the lobula plate of Drosophila melanogaster. *Curr Biol* **18**:368–74.

Joesch M, Schnell B, Raghu SV, Reiff DF, Borst A. 2010. ON and OFF pathways in Drosophila motion vision. *Nature* **468**:300–304.

References

Juusola M, Hardie RC. 2001. Light adaptation in Drosophila photoreceptors: I. Response dynamics and signaling efficiency at 25°C. *J Gen Physiol* **117**:3–25.

Kaiser W, Garrett CGB. 1961. Two-photon excitation in CaF2:Eu2+. *Phys Rev Lett* **7**:229–231.

Karagyozov D, Mihovilovic Skanata M, Lesar A, Gershow M. 2018. Recording neural activity in unrestrained animals with three-dimensional tracking two-photon microscopy. *Cell Rep* **25**:1371–1383.

Katona G, Szalay G, Maák P, Kaszás A, Veress M, Hillier D, Chiovini B, Vizi ES, Roska B, Rózsa B. 2012. Fast two-photon in vivo imaging with three-dimensional random-access scanning in large tissue volumes. *Nat Methods* **9**:201–208.

Kim AJ, Fenk LM, Lyu C, Maimon G. 2017. Quantitative predictions orchestrate visual signaling in Drosophila. *Cell.*

Kim AJ, Fitzgerald JK, Maimon G. 2015. Cellular evidence for efference copy in Drosophila visuomotor processing. *Nat Neurosci* **15**:1247–1255.

Kim KH, Buehler C, Bahlmann K, Ragan T, Lee W-CA, Nedivi E, Heffer EL, Fantini S, So PTC. 2007. Multifocal multiphoton microscopy based on multianode photomultiplier tubes. *Opt Express* **15**:11658–78.

Kirkby PA, Srinivas Nadella KMN, Silver RA. 2010. A compact acousto-optic lens for 2D and 3D femtosecond based 2-photon microscopy. *Opt Express* **18**:13720.

Kitamoto T. 2001. Conditional modification of behavior in Drosophila by targeted expression of a temperature-sensitive shibire allele in defined neurons. *J Neurobiol* **47**:81–91.

Klapoetke NC, Murata Y, Kim SS, Pulver SR, Birdsey-Benson A, Cho YK, Morimoto TK, Chuong AS, Carpenter EJ, Tian Z, Wang J, Xie Y, Yan Z, Zhang Y, Chow BY, Surek B, Melkonian M, Jayaraman V, Constantine-Paton M, Wong GK-S, Boyden ES. 2014. Independent optical excitation of distinct neural populations. *Nat Methods* **11**:338–46.

Klapoetke NC, Nern A, Peek MY, Rogers EM, Breads P, Rubin GM, Reiser MB, Card GM. 2017. Ultra-selective looming detection from radial motion opponency. *Nature* **551**:237–241.

Kong L, Tang J, Little JP, Yu Y, Lämmermann T, Lin CP, Germain RN, Cui M. 2015. Continuous volumetric imaging via an optical phase-locked ultrasound lens. *Nat Methods* **12**:759–762.

Kühn NK, Gollisch T. 2019. Activity correlations between direction-selective retinal ganglion cells synergistically enhance motion decoding from complex visual scenes. *Neuron* **101**:963–976.

Lai SL, Lee T. 2006. Genetic mosaic with dual binary transcriptional systems in Drosophila. *Nat Neurosci* **9**:703–709.

Land MF, Collett TS. 1974. Chasing behaviour of houseflies (Fannia canicularis) - A description and analysis. *J Comp Physiol* **89**:331–357.

Laughlin SB, Hardie RC. 1978. Common strategies for light adaptation in the peripheral visual systems of fly and dragonfly. *J Comp Physiol A* **128**:319–340.

Lee T, Luo L. 1999. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**:451–461.

Leong JCS, Esch JJ, Poole B, Ganguli S, Clandinin TR. 2016. Direction selectivity in Drosophila emerges from preferred-direction enhancement and null-direction suppression. *J Neurosci* **36**:8078–8092.

Leonhardt A, Ammer G, Meier M, Serbe E, Bahl A, Borst A. 2016. Asymmetry of Drosophila ON and OFF motion detectors enhances real-world velocity estimation. *Nat Neurosci* **19**:706–715.

Leonhardt A, Meier M, Serbe E, Eichner H, Borst A. 2017. Neural mechanisms underlying sensitivity to reverse-phi motion in the fly. *PLoS One* **12**:e0189019.

Lichtman JW, Conchello J-A. 2005. Fluorescence microscopy. *Nat Methods* **2**:85–132.

Lu R, Sun W, Liang Y, Kerlin A, Bierfeld J, Seelig JD, Wilson DE, Scholl B, Mohar B, Tanimoto M, Koyama M, Fitzpatrick D, Orger MB, Ji N. 2017. Video-rate volumetric functional imaging of the brain at synaptic resolution. *Nat Neurosci* **20**:620–628.

Luan H, Peabody NC, Vinson CR, White BH. 2006. Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. *Neuron* **52**:425–436.

Mahou P, Zimmerley M, Loulier K, Matho KS, Labroille G, Morin X, Supatto W, Livet J, Débarre D, Beaurepaire E. 2012. Multicolor two-photon tissue imaging by wavelength mixing. *Nat Methods* **9**:815–818.

References

Maisak MS, Haag J, Ammer G, Serbe E, Meier M, Leonhardt A, Schilling T, Bahl A, Rubin GM, Nern A, Dickson BJ, Reiff DF, Hopp E, Borst A. 2013. A directional tuning map of Drosophila elementary motion detectors. *Nature* **500**:212–216.

Marre O, Amodei D, Deshmukh N, Sadeghi K, Soo F, Holy TE, Berry MJ. 2012. Mapping a complete neural population in the retina. *J Neurosci* **32**:14859–14873.

Masland RH. 2012. The neuronal organization of the retina. *Neuron* **76**:266–280.

Masters BR. 2008. History of the optical microscope in cell biology and medicineEncyclopedia of Life Sciences. Chichester: John Wiley & Sons, Ltd.

Maturana H, Lettvin J. 1960. Anatomy and physiology of vision in the frog (Rana pipiens). *J Gen Physiol* **43**:129–175.

Mauss AS, Busch C, Borst A. 2017a. Optogenetic neuronal silencing in Drosophila during visual processing. *Sci Rep* **7**:13823.

Mauss AS, Meier M, Serbe E, Borst A. 2014. Optogenetic and pharmacologic dissection of feedforward inhibition in Drosophila motion vision. *J Neurosci* **34**:2254–2263.

Mauss AS, Pankova K, Arenz A, Nern A, Rubin GM, Borst A. 2015. Neural circuit to integrate opposing motions in the visual field. *Cell* **162**:351–362.

Mauss AS, Vlasits A, Borst A, Feller M. 2017b. Visual circuits for direction selectivity. *Annu Rev Neurosci* **40**:211–230.

Meier M, Borst A. 2019. Extreme compartmentalization in a Drosophila amacrine cell. *Curr Biol* **29**:1545–1550.

Meier M, Serbe E, Maisak MS, Haag J, Dickson BJ, Borst A. 2014. Neural circuit components of the Drosophila OFF motion vision pathway. *Curr Biol* 385–392.

Meinertzhagen IA, O'Neil SD. 1991. Synaptic organization of columnar elements in the lamina of the wild type in Drosophila melanogaster. *J Comp Neurol* **305**:232–263.

Meister M. 1995. Multineuronal codes in retinal signaling. *Proc Natl Acad Sci* **93**:609–614.

Meister M, Pine J, Baylor DA. 1994. Multi-neuronal signals from the retina: acquisition and analysis. *J Neurosci* **51**:95–106.

Minsky M. 1955. Confocal Microscope. US3013467.

Morgan TH. 1910. Sex limited inheritance. *Science* **32**:120–122.

Nagel G, Ollig D, Fuhrmann M, Kateriya S, Musti AM, Bamberg E, Hegemann P. 2002. Channelrhodopsin-1: A light-gated proton channel in green algae. *Science* **296**:2395–2398.

Nern A, Pfeiffer BD, Rubin GM. 2015. Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. *Proc Natl Acad Sci USA* **112**:E2967–E2976.

Nguyen QT, Callamaras N, Hsieh C, Parker I. 2001. Construction of a two-photon microscope for video-rate Ca2+imaging. *Cell Calcium* **30**:383–393.

Nikolenko V. 2008. SLM microscopy: scanless two-photon imaging and photostimulation using spatial light modulators. *Front Neural Circuits* **2**.

Ohno N, Katoh M, Saitoh Y, Saitoh S. 2016. Recent advancement in the challenges to connectomics. *Microscopy* **65**:97–107.

Olberg RM, Seaman RC, Coats MI, Henry AF. 2007. Eye movements and target fixation during dragonfly prey-interception flights. *J Comp Physiol A* **193**:685–693.

Ouzounov DG, Wang T, Wang M, Feng DD, Horton NG, Cruz-Hernández JC, Cheng Y-T, Reimer J, Tolias AS, Nishimura N, Xu C. 2017. In vivo three-photon imaging of activity of GCaMP6-labeled neurons deep in intact mouse brain. *Nat Methods* 20–24.

Ozbay BN, Futia GL, Ma M, Bright VM, Gopinath JT, Hughes EG, Restrepo D, Gibson EA. 2018. Three dimensional two-photon brain imaging in freely moving mice using a miniature fiber coupled microscope with active axial-scanning. *Sci Rep* **8**:8108.

Pankova K, Borst A. 2016. RNA-seq transcriptome analysis of direction-selective T4/T5 neurons in Drosophila. *PLoS One* **11**:e0163986.

References

Panzera LC, Hoppa MB. 2019. Genetically encoded voltage indicators are illuminating subcellular physiology of the axon. *Front Cell Neurosci* **13**.

Peli E. 1990. Contrast in complex images. *J Opt Soc Am A* **7**:2032–40.

Pfeiffer BD, Jenett A, Hammonds AS, Ngo T-TB, Misra S, Murphy C, Scully A, Carlson JW, Wan KH, Laverty TR, Mungall C, Svirskas R, Kadonaga JT, Doe CQ, Eisen MB, Celniker SE, Rubin GM. 2008. Tools for neuroanatomy and neurogenetics in Drosophila. *Proc Natl Acad Sci USA* **105**:9715–9720.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP. 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**:995–999.

Platisa J, Pieribone VA. 2018. Genetically encoded fluorescent voltage indicators: are we there yet? *Curr Opin Neurobiol* **50**:146–153.

Potter CJ, Tasic B, Russler E V., Liang L, Luo L. 2010. The Q system: A repressible binary system for transgene expression, lineage tracing, and mosaic analysis. *Cell* **141**:536–548.

Power RM, Huisken J. 2017. A guide to light-sheet fluorescence microscopy for multiscale imaging. *Nat Methods* **14**:360–373.

Puchalla JL, Schneidman E, Harris RA, Berry MJ. 2005. Redundancy in the population code of the retina. *Neuron* **46**:493–504.

Ramon y Cajal S, Sanches D. 1915. Contribución al conocimiento de los centros nerviosos de los insectos.

Ready DF, Hanson TE, Benzer S. 1976. Development of the Drosophila retina, a neurocrystalline lattice. *Dev Biol* **53**:217–240.

Reichardt W. 2013. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system In: Rosenblith WA, editor. Sensory Communication. Cambridge: MIT Press Scholarship Online.

Reichardt W. 1987. Evaluation of optical motion information by movement detectors. *J Comp Physiol A* **161**:533–547.

Reichardt W, Poggio T, Hausen K. 1983. Figure-ground discrimination by relative movement in the visual system of the fly. *Biol Cybern* **162**:153–162.

Reiff DF, Plett J, Mank M, Griesbeck O, Borst A. 2010. Visualizing retinotopic half-wave rectified input to the motion detection circuitry of Drosophila. *Nat Neurosci* **13**:973–978.

Ribeiro IMA, Drews M, Bahl A, Machacek C, Borst A, Dickson BJ. 2018. Visual projection neurons mediating directed courtship in Drosophila. *Cell* **174**:607–621.

Richter FG, Fendl S, Haag J, Drews MS, Borst A. 2018. Glutamate signaling in the fly visual system. *iScience* **7**:85–95.

Rieke F, Rudd ME. 2009. The challenges natural images pose for visual adaptation. *Neuron* **64**:605–616.

Rister J, Pauls D, Schnell B, Ting C-Y, Lee C-H, Sinakevitch I, Morante J, Strausfeld NJ, Ito K, Heisenberg M. 2007. Dissection of the peripheral motion channel in the visual system of Drosophila melanogaster. *Neuron* **56**:155–70.

Rivera-Alba M, Vitaladevuni SN, Mischenko Y, Lu Z, Takemura S, Scheffer L, Meinertzhagen IA, Chklovskii DB, De Polavieja GG, Gonzalo G. 2011. Wiring economy and volume exclusion determine neuronal placement in the Drosophila brain. *Curr Biol* **21**:2000–2005.

Rock I, Smith D. 1986. The optomotor response and induced motion of the self. *Perception* **15**:497–502.

Rózsa B, Szalay G, Katona G. 2014. Acousto-optical scanning – based high-speed 3D two-photon imaging in vivo In: Korngreen A, editor. Advanced Patch-Clamp Analysis for Neuroscientists. New York: Humana Press. pp. 213–245.

Ruderman DL, Bialek W. 1994. Statistics of natural images: Scaling in the woods. *Phys Rev Lett* **73**:814–817.

Rupprecht P, Prendergast A, Wyart C, Friedrich RW. 2016. Remote z-scanning with a macroscopic voice coil motor for fast 3D multiphoton laser scanning microscopy. *Biomed Opt Express* **7**:1656–1671.

Salazar-Gatzimas E, Chen J, Creamer MS, Mano O, Mandel HB, Matulis CA, Pottackal J, Clark DA. 2016. Direct measurement of correlation responses in Drosophila elementary motion detectors reveals fast timescale tuning. *Neuron* **92**:227–239.

Schilling T, Borst A. 2015. Local motion detectors are required for the computation of expansion flow-fields. *Biol*

References

Open **4**:1105–1108.

Schnell B, Joesch M, Forstner F, Raghu SV, Otsuna H, Ito K, Borst A, Reiff DF. 2010. Processing of horizontal optic flow in three visual interneurons of the Drosophila brain. *J Neurophysiol* **103**:1646–57.

Schnell B, Raghu SV, Nern A, Borst A. 2012. Columnar cells necessary for motion responses of wide-field visual interneurons in Drosophila. *J Comp Physiol A* **198**:389–395.

Schnell B, Weir PT, Roth E, Fairhall AL, Dickinson MH. 2014. Cellular mechanisms for integral feedback in visually guided behavior. *Proc Natl Acad Sci USA* **111**:5700–5.

Serbe E, Meier M, Leonhardt A, Borst A. 2016. Comprehensive characterization of the major presynaptic elements to the Drosophila OFF motion detector. *Neuron* **89**:829–841.

Sheffield MEJ, Dombeck DA. 2015. Calcium transient prevalence across the dendritic arbour predicts place field properties. *Nature* **517**:200–204.

Shinomiya K, Huang G, Lu Z, Parag T, Xu CS, Aniceto R, Ansari N, Cheatham N, Lauchie S, Neace E, Ogundeyi O, Ordish C, Peel D, Shinomiya A, Smith C, Takemura S, Talebi I, Rivlin PK, Nern A, Scheffer LK, Plaza SM, Meinertzhagen IA. 2019. Comparisons between the ON- and OFF-edge motion pathways in the Drosophila brain. *Elife* **8**:e40025.

Shinomiya K, Karuppudurai T, Lin T-Y, Lu Z, Lee C-H, Meinertzhagen IA. 2014. Candidate neural substrates for OFF-edge motion detection in Drosophila. *Curr Biol* **24**:1062–70.

Silies M, Gohl D, Fisher Y, Freifeld L, Clark DA, Clandinin TR. 2013. Modular use of peripheral input channels tunes motion-detecting circuitry. *Neuron* **79**:111–127.

Simmons KD, Prentice JS, Tkačik G, Homann J, Yee HK, Palmer SE, Nelson PC, Balasubramanian V. 2013. Transformation of stimulus correlations by the retina. *PLoS Comput Biol* **9**:e1003344.

Simoncelli EP, Olshausen BA. 2001. Natural image statistics and neural representation. *Annu Rev Neurosci* **24**:1193–1216.

Smirnov MS, Evans PR, Garrett TR, Yan L, Yasuda R. 2017. Automated remote focusing, drift correction, and photostimulation to evaluate structural plasticity in dendritic spines. *PLoS One* **12**:e0170586.

So PTC, Dong CY, Masters BR, Berland KM. 2000. Two-photon excitation fluorescence microscopy. *Annu Rev Biomed engeneering* **2**:399–429.

Sofroniew NJ, Flickinger D, King J, Svoboda K. 2016. A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *Elife* **5**:e14472.

St Johnston D. 2002. The art and design of genetic screens: Drosophila melanogaster. *Nat Rev Genet* **3**:176–88.

Stirman JN, Smith IT, Kudenov MW, Smith SL. 2016. Wide field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain. *Nat Biotechnol* **34**:857–862.

Stöckl AL, O'Carroll DC, Warrant EJ. 2016. Neural summation in the hawkmoth visual system extends the limits of vision in dim light. *Curr Biol* **26**:821–826.

Strausfeld NJ. 1971. The organization of the insect visual system (Light Microscopy). *Symp Soc Exp Biol* **25**:379–390.

Strother JA, Wu S-T, Rogers EM, Eliason JLM, Wong AM, Nern A, Reiser MB. 2018. Behavioral state modulates the ON visual motion pathway of Drosophila. *Proc Natl Acad Sci USA* **115**:E102–E111.

Strother JA, Wu S, Wong AM, Nern A, Rogers EM, Le JQ, Rubin GM, Reiser MB. 2017. The emergence of directional selectivity in the visual motion pathway of Drosophila. *Neuron* **94**:168–182.

Suver MP, Mamiya A, Dickinson MH. 2012. Octopamine neurons mediate flight-induced modulation of visual processing in drosophila. *Curr Biol* **22**:2294–2302.

Svoboda K, Yasuda R. 2006. Principles of two-photon excitation microscopy and its applications to neuroscience. *Neuron* **50**:823–839.

Swanson LW, Lichtman JW. 2016. From Cajal to connectome and beyond. *Annu Rev Neurosci* **39**:197–216.

Sweeney ST, Broadie K, Keane J, Niemann H, Kane CJO. 1995. Targeted expression of tetanus toxin light chain in Drosophila specifically eliminates synaptic transmission and causes behavioral defects. *Neuron* **14**:341–351.

Takemura S-Y. 2014. Connectome of the fly visual circuitry. *Microscopy* 1–8.

References

Takemura S-Y, Lu Z, Meinertzhagen IA. 2008. Synaptic circuits of the Drosophila optic lobe: the input terminals to the medulla. *J Comp Neurol* **509**:493–513.

Takemura S, Nern A, Chklovskii DB, Scheffer LK, Rubin GM, Meinertzhagen IA. 2017. The comprehensive connectome of a neural substrate for 'ON' motion detection in Drosophila. *Elife* **6**:e24394.

Takemura Shin-ya, Bharioke A, Lu Z, Nern A, Vitaladevuni S, Rivlin PK, Katz WT, Olbris DJ, Plaza SM, Winston P, Zhao T, Horne JA, Fetter RD, Takemura Satoko, Blazek K, Chang L-A, Ogundeyi O, Saunders MA, Shapiro V, Sigmund C, Rubin GM, Scheffer LK, Meinertzhagen IA, Chklovskii DB. 2013. A visual motion detection circuit suggested by Drosophila connectomics. *Nature* **500**:175–81.

Takemura Shin-ya, Xu CS, Lu Z, Rivlin PK, Parag T, Olbris DJ, Plaza S, Zhao T, Katz WT, Umayam L, Weaver C, Hess HF, Horne JA, Nunez-Iglesias J, Aniceto R, Chang L-A, Lauchie S, Nasca A, Ogundeyi O, Sigmund C, Takemura Satoko, Tran J, Langille C, Le Lacheur K, McLin S, Shinomiya A, Chklovskii DB, Meinertzhagen IA, Scheffer LK. 2015. Synaptic circuits and their variations within different columns in the visual system of Drosophila. *Proc Natl Acad Sci* **112**:13711–13716.

Tao X, Lin H-H, Lam T, Rodriguez R, Wang JW, Kubby J. 2017. Transcutical imaging with cellular and subcellular resolution. *Biomed Opt Express* **8**:1277–1289.

Tuthill JC, Chiappe ME, Reiser MB. 2011. Neural correlates of illusory motion perception in Drosophila. *Proc Natl Acad Sci USA* **108**:9685–90.

Tuthill JC, Nern A, Holtz SL, Rubin GM, Reiser MB. 2013. Contributions of the 12 neuron classes in the fly lamina to motion vision. *Neuron* **79**:128–40.

Tuthill JC, Nern A, Rubin GM, Reiser MB. 2014. Wide-field feedback neurons dynamically tune early visual processing. *Neuron* **82**:887–95.

Venken KJT, Simpson JH, Bellen HJ. 2011. Genetic manipulation of genes and cells in the nervous system of the fruit fly. *Neuron* **72**:202–230.

Wang H, Jing M, Li Y. 2018. Lighting up the brain: genetically encoded fluorescent sensors for imaging neurotransmitters and neuromodulators. *Curr Opin Neurobiol* **50**:171–178.

Wardill TJ, Fabian ST, Pettigrew AC, Stavenga DG, Nordström K, Gonzalez-Bellido PT. 2017. A novel interception strategy in a miniature robber fly with extreme visual acuity. *Curr Biol* **27**:854–859.

Wardill TJ, List O, Li X, Dongre S, McCullock M, Ting C-Y, O'Kane CJ, Tang S, Lee C-H, Hardie RC, Juusola M. 2012. Multiple spectral inputs improve motion discrimination in the Drosophila visual system. *Science* **11**:925–932.

Wässle H, Peichl L, Boycott BB. 1981. Dendritic territories of cat retinal ganglion cells. *Nature* **292**:344–345.

Weber B, Helmchen F. 2014. Optical imaging of neocortical dynamics. New York: Humana Press.

White JG, Southgate E, Thomson JN, Brenner S. 1986. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc B Biol Sci* **314**:1–340.

Wienecke CFR, Leong JCS, Clandinin TR. 2018. Linear summation underlies direction selectivity in Drosophila. *Neuron* **99**:680–688.

Wilson RI, Turner GC, Laurent G. 2004. Transformation of olfactory representations in the Drosophila antennal lobe. *Science* **303**:366–370.

Yamaguchi S, Wolf R, Desplan C, Heisenberg M. 2008. Motion vision is independent of color in Drosophila. *Proc Natl Acad Sci* **105**:4910–4915.

Yang HHH, St-Pierre F, Sun X, Ding X, Lin MZZ, Clandinin TR. 2016. Subcellular imaging of voltage and calcium signals reveals neural processing in vivo. *Cell* **166**:245–257.

Yang W, Carrillo-Reid L, Bando Y, Peterka DS, Yuste R. 2018. Simultaneous two-photon imaging and two-photon optogenetics of cortical circuits in three dimensions. *Elife* **7**:e32671.

Yang W, Kang Miller J, Carrillo-Reid L, Pnevmatikakis E, Paninski L, Yuste R, Peterka DS. 2016. Simultaneous multi-plane imaging of neural circuits. *Neuron* **89**:269–284.

Yang W, Yuste R. 2017. In vivo imaging of neural activity. *Nat Methods* **14**:349–359.

Yang Y, Liu N, He Y, Liu Y, Ge L, Zou L, Song S, Xiong W, Liu X. 2018. Improved calcium sensor GCaMP-X overcomes the calcium channel perturbations induced by the calmodulin in GCaMP. *Nat Commun* **9**:1504.

Zheng Z, Lauritzen JS, Perlman E, Robinson CG, Nichols M, Milkie D, Torrens O, Price J, Fisher CB, Sharifi N,

References

Calle-Schuler SA, Kmecova L, Ali IJ, Karsh B, Trautman ET, Bogovic JA, Hanslovsky P, Jefferis GSXE, Kazhdan M, Khairy K, Saalfeld S, Fetter RD, Bock DD. 2018. A complete electron microscopy volume of the brain of adult Drosophila melanogaster. *Cell* **174**:730–743.

Zong W, Wu R, Li M, Hu Y, Li Y, Li J, Rong H, Wu H, Xu Y, Lu Y, Jia H, Fan M, Zhou Z, Zhang Y, Wang A, Chen L, Cheng H. 2017. Fast high-resolution miniature two-photon microscopy for brain imaging in freely behaving mice. *Nat Methods* **14**:713–719.

7   APPENDIX

7.1   ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Axel. With his guidance, I learned about countless aspects of science. He gave me the freedom and room to discover and learn at my own pace and endowed me with the trust to build a microscope. His scientific curiosity and his continued, genuine interest in science create a unique atmosphere in the lab that make it a pleasure and privilege to work in.

Many other people have been directly or indirectly involved in my PhD, which would not have been possible without their continuous help and support. Them I would like to thank in the following.

My TAC members, Andreas and Ruben, who gave me several pieces of very valuable advice. Winfried Denk for the initial design of the remote-focusing module. Robert, who helped me a lot at a time when I still knew very little about microscopes. Bulle for sharing his two-photon expertise with me throughout, and who is always there, ready to help when something doesn't work. Thomas and Joe who made me give up on galvos and use a piezo instead. Armin, who was supervising me through my first internship, welcomed me to the lab, and whose excitement about science has been contagious. Flo and Michi, my longest P7 contestants, for countless discussions and a lot of fun. Aljoscha for lots of helpful advice, discussions and for sharing his skills. Sandra for her good questions and her exemplary steadiness. Nadya for a lot of encouragement, discussions and because she dares stuff. Georg, MM, Jesus, Tabea, ET, Niko, Jonatan for input and support in many aspects. Together, all members of the lab for an amazing environment. The proofreaders of this thesis, Nadya, Flo, Aljoscha, Georg, Martin, Lee and Vilim. The GSN for providing the framework for my Master's as well as my PhD.

Finally, I would like to thank all the amazing people that have come with me through this time: the gardening friends, dinner friends, dancing friends, best friends, and friends for all circumstances: Ksenia, Vero, Miguel, Vilim, Alex, Judi, Nadya, Elena, Martin, Sara and Lee.

Lastly and most importantly, I would like to thank my parents for their unconditional support and trust in me, with anything I pursue.

7.2 AFFIDAVIT

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation „Investigation of the *Drosophila* motion vision circuit using remote-focusing two-photon microscopy" selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation „Investigation of the *Drosophila* motion vision circuit using remote-focusing two-photon microscopy" is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

München, den   Unterschrift

Munich, date   Signature

4. October 2019   Anna Schuetzenberger

7.3   AUTHOR CONTRIBUTIONS

MANUSCRIPT 1: Seeing through the eye of a fly: Reconstructing *Drosophila*'s visual map with remote focusing two-photon microscopy

Authors	**Anna Schuetzenberger** and Alexander Borst
Contributions	A.B. and **A.S.** jointly designed the study. **A.S.** built the remote-focusing module, performed all experiments and data analysis, and wrote the manuscript. A.B. edited the manuscript.

MANUSCRIPT 2: Dynamic signal compression for robust motion vision in flies

Authors	Michael S. Drews, Aljoscha Leonhardt, Nadezhda Pirogova, Florian G. Richter, **Anna Schuetzenberger**, Lukas Braun, Etienne Serbe, and Alexander Borst
Contributions	M.S.D., A.L., and A.B. jointly conceived the study. M.S.D. and A.L. designed all experiments. A.L. and L.B. conducted behavioral experiments. E.S. recorded electrophysiological responses. M.S.D., N.P., F.G.R., and **A.S.** performed calcium imaging. A.L. designed and analyzed the convolutional model. M.S.D. and A.L. analyzed data, performed modelling, and wrote the manuscript. All authors participated in editing the manuscript.

MANUSCRIPT 3: Surround motion shapes local motion computation on the dendrites of T4 and T5 cells in *Drosophila*

Authors	**Anna Schuetzenberger** and Alexander Borst
Contributions	A.B. and **A.S.** jointly designed the study. **A.S.** performed all experiments and data analysis, and wrote the manuscript. A.B. edited the manuscript.

…………………………..   ……………………………．

Anna Schuetzenberger   Prof. Dr. Alexander Borst

                                                           Supervisor