DANIEL HARNACK

EFFECTIVE INFLUENCES IN NEURONAL NETWORKS: ATTENTIONAL MODULATIONS OF EFFECTIVE INFLUENCES UNDERLYING FLEXIBLE PROCESSING AND HOW TO MEASURE THEM

Dissertation
in partial fulfilment of the degree
Doctor rerum naturalium (Dr. rer. nat.)

EFFECTIVE INFLUENCES IN NEURONAL NETWORKS: ATTENTIONAL MODULATIONS OF EFFECTIVE INFLUENCES UNDERLYING FLEXIBLE PROCESSING AND HOW TO MEASURE THEM

submitted by
DANIEL HARNACK, M.SC.

Univerität Bremen
Fachbereich für Physik und Elektrotechnik
Institut für Theoretische Physik

Gutachter:
Dr. Udo A. Ernst
Prof. Dr. Andreas K. Kreiter

Dedicated to friends and family

What is mind?
No matter.
What is matter?
Nevermind.

George Berkeley

The Magician ©James R. Eads,
with permission from the artist.

ABSTRACT

Selective routing of information between brain areas is a key prerequisite for flexible adaptive behaviour. It allows to focus on relevant information and to ignore potentially distracting influences. Selective attention is a psychological process which controls this preferential processing of relevant information. The neuronal network structures and dynamics, and the attentional mechanisms by which this routing is enabled are not fully clarified. Based on previous experimental findings and theories, a network model is proposed which reproduces a range of results from the attention literature. It depends on shifting of phase relations between oscillating neuronal populations to modulate the effective influence of synapses. This network model might serve as a generic routing motif throughout the brain. The attentional modifications of activity in this network are investigated experimentally and found to employ two distinct channels to influence processing: facilitation of relevant information and independent suppression of distracting information. These findings are in agreement with the model and previously unreported on the level of neuronal populations.

Furthermore, effective influence in dynamical systems is investigated more closely. Due to a lack of a theoretical underpinning for measurements of influence in non-linear dynamical systems such as neuronal networks, often unsuited measures are used for experimental data that can lead to erroneous conclusions. Based on a central theorem in dynamical systems, a novel theory of effective influence is developed. Measures derived from this theory are demonstrated to capture the time dependent effective influence and the asymmetry of influences in model systems and experimental data. This new theory holds the potential to uncover previously concealed interactions in generic non-linear systems studied in a range of disciplines, such as neuroscience, ecology, economy and climatology.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

LFP	local field potential
MUA	multi unit activity
RISER	rate-imbalance induced selective routing
CTC	communication through coherence
ING	interneuron network γ
PING	pyramidal-interneuron network γ
GC	Granger causality
CCM	convergent cross-mapping
TC	topological causality
TE	transfer entropy
EEG	electroencephalography
LGN	lateral geniculate nucleus
V1	primary visual cortex
V2	visual area 2
V4	visual area 4

1 MOTIVATION

history of brain research

Research of the human brain has an eventful history. From being hypothesized by Aristotle as a cooling device for the blood heated by emotions [AriBCb; Gro95], in contemporary science it evolved into the primary seat of intelligence and information processing in the body. It was established that it is made up of neurons, specialized complex cell types, that form highly connected networks [Caj94; DB55], and that information can be transmitted between the neurons by electrical discharges [HH52]. The brain receives input through different sensory pathways and can affect the actions of the body by muscle activation. Several subnetworks within the brain are crucial for formation and retrieval of memories, performing calculations, planning, and reasoning. In short: it is a biological manifestation of a generally intelligent system. The vast majority of scientific approaches towards the brain is guided by the materialist view that its capacities emerge from the interaction within neuronal networks, and that every behaviour or cognitive state correlates to a state of these networks (see e.g. [Jan08]).

artificial Intelligence

Not surprisingly, analogies to the functioning of the human brain permeate the underpinnings of the field of artificial intelligence [McC+06]. Most recently, fertilized by results from neuroscience research [Has+17], impressive and media-effective feats have been accomplished, such as the passing of the Turing test [Tur50] in a number of domains [LST15; Owe+16], which was designed as a yardstick to asses whether human and machine behaviour can be discriminated. But not only can human performance be mimicked, it can also be surpassed: The game of Go, seen by many players as an art form due to its impressive combinatorics and hence necessary reliance of players on keen intuition, has been mastered at a super-human performance level without the need of any human teaching [Sil+17].

flexibility

In light of these developments, it seems that understanding of the inner workings of the brain has gone a long way. So, following the "build it, then you understand it" credo, one might ask: what is left to do? While there is certainly a multitude of answers to this question, I want to focus on one issue in this thesis which I deem of special importance: that of flexibility. Humans effortlessly adapt behaviour to changing context, react quickly to unexpected events, and reallocate resources to process information which is of importance for the current situation. In fact, this flexibility is in most parts what still sets humans apart from intelligent machines, which are typically very adept at one task but unable to perform similarly well under different circumstances and task demands: an artificial neuronal network trained to detect faces can match human performance [Tai+14], but will typically be bad at detecting anything else, whereas a normal human has no problem to take in a complex visual scenery and flexibly change what they are looking for, one moment a face, the next maybe a car or a tiger. How do human brains achieve this flexibility?

Formally, flexibility can be defined as the ability of a system to generate different output, such as an action, in response to the same input, depending on the task demand. In a very simple example, which is close to the experiments analysed later, subjects could be shown a video where streams of numbers are displayed at several different locations, and asked to press a button if a 4 appears at one location which was specified beforehand. Now, the same video could be shown, but a response to the appearance of a 4 at a different location could be required. It is clear that the input to the neuronal network, here the visual scene, is the same in both tasks, whereas the output differs.

visual hierarchy

Understanding how flexibility under these circumstances is achieved in neuronal networks is constrained by the brain's architecture. The network processing the visual input can be approximated as having a feed-forward, fan-in structure, meaning several neurons in one layer project to the same neuron in the next layer. This leads to a visual hierarchy where neurons higher up in the network have access to more information than neurons at the bottom. As a consequence, while neurons in early layers

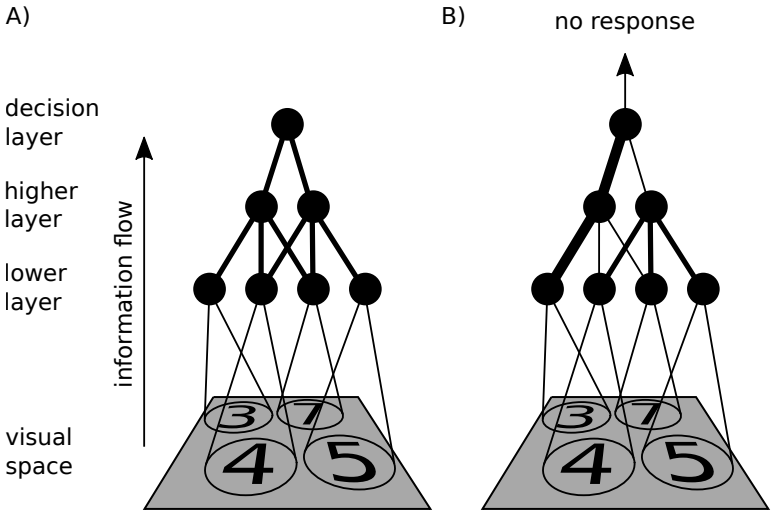

Figure 1.1: Hierarchical network organization of visual processing. **A)** Different numbers are shown at four locations in the visual scenery. The area in visual space which evokes responses in each neuron in the lower layer is symbolized by the black circle. Several neurons project to the same neuron in a higher area. The lines connecting neurons symbolize directed information flow to the ones higher up in the hierarchy. The decision neuron at the top can potentially represent information from all four locations in visual space. **B)** If the task is to detect a 4 in the upper left location, the neuronal network must route information to the decision stage such that competing information from other locations is attenuated. This is achieved by increasing the effective influence of some connections, symbolized by thicker lines, and decreasing the effective influence of others, symbolized by thinner lines.

only represent information of small areas in visual space, this area grows when ascending the hierarchy. At the top of the hierarchy, a decision stage can be postulated which has potential access to information from the whole scene and controls the output of the network. This hierarchical network is illustrated in figure 1.1 A).

information routing bottleneck

An important implication of this structure is that it creates a "bottleneck". The diagram in figure 1.1 A) shows that the neuron at the top of the hierarchy can potentially represent information from all possible locations. However, for the task of detecting a 4 at one location, only the information from this location is relevant and should be routed to the decision stage. Information from the other locations, which could interfere at the decision stage, should be gated out. This can be achieved by changing the impact, or the effective influence, of neurons onto the decision stage: The effective influence of neurons which carry information about the task-relevant location should be increased, whereas the influence of neurons processing task-irrelevant locations should be decreased. This situation is depicted in figure 1.1 B).

attention guides information flow

In psychological terms, the process which controls the flow of information at this bottleneck, and hence is a prerequisite for flexible behaviour, is selective attention. Attention is typically loosely defined as a "spotlight" or "zoom lens" emphasizing defining features of stimuli that are important for the task while diminishing potentially distracting influences [Von67; Jam90; BK15b]. In the example task, the location of interest would be put in the "spotlight" such that things happening here are clearly visible, whereas the other locations can be dimmed down because the events at these locations are not important at the moment.

The effects of attention can be observed in the activity of single neurons and neuronal populations in the visual system of the brain. Several studies, where the activity of neurons during an attention demanding task was directly recorded, showed indeed that neurons higher up in the visual hierarchy, which can respond to several locations, behave as if only stimuli at attended locations are present although others are shown at the same time [MD85; Gro+12; Bos+12].

flexibility, information routing and attention

To summarize: in hierarchical feed-forward neuronal networks, flexibility can be achieved by changing effective influences among neurons, which allows for selective information routing. This changing is controlled by a psychological process called attention.

The first part of this thesis now revolves around two specific questions:

1. Which network structures and dynamics naturally support selective information routing?

2. How and where does attention intervene to bring about selective routing in the network?

The first question is addressed by building on previous theories and proposing a network structure that is shown by simulations to perform selective routing with minimal interventions. By global parameter optimization this structure is verified to be an optimal routing configuration under realistic neurophysiological constraints. For the second question, neuronal activity of the primate brain is analysed from a specifically designed experiment, where a direct prediction from the model is tested. Attentional effects are in accordance with previous results and with the proposed attentional mechanism embedded in the model. The results are further used to refine the model of selective routing networks.

effective influence in dynamical systems

In a second part, the quantity that needs to be modulated for selective routing to take place, the "effective influence", is scrutinized more closely. Up to this point, a somewhat intuitive definition of effective influence will suffice. Here now, a theoretically solid quantification is sought. This endeavour leads away from the details of biological neuronal networks and to a broader perspective of non-linear dynamical systems, which encompass neuronal networks. Currently most experimental studies rely on measures of effective influence developed for linear stochastic systems, which can lead to erroneous conclusions when applied to non-linear dynamical systems. For the lack of a comprehensive theory for non-linear systems, a novel approach is introduced. First, it is developed purely theoretically based on a fundamental mathematical theorem of non-linear dynamics. Successful applicability to complex model systems and experimental data is demonstrated in a first test. Furthermore, the merits in comparison to other established measures and theoretical links are discussed.

summary

Taken together, in the course of this thesis a new network model of selective routing underlying flexible information processing is proposed, the understanding of attentional mechanisms in the primate brain is advanced, and a novel theory of effective influence for non-linear dynamical systems in general is presented.

Part I

ATTENTIONAL MECHANISMS OF INFORMATION ROUTING

OUTLINE PART I

This first part of the thesis covers selective routing in biologically realistic neuronal networks and the necessary attentional interventions. It is organized in three chapters.

In chapter 3, a model neuronal network representing two successive stages in the visual hierarchy is developed, adhering to known properties of biological neurons and networks. It is shown that a specific recurrent connectivity structure among neurons of the same layer, along with attention dependent input to the first layer, is instrumental in selectively modulating the effective influence of connections to the following layer. In addition, the model faithfully reproduces several electrophysiological attention related results from the experimental literature for the first time in a single framework [HEP15].

Further, in chapter 4 a simplification of the network model is introduced, which allows to optimize the model parameters to route information most effectively, and to investigate whether there may be networks which route information in a fundamentally different way to the one presented in chapter 3. It turns out that the network previously discussed is indeed optimal to route information, given some reasonable biological constraints.

From the type of routing networks modelled in chapters 3 and 4, a prediction is generated on the exact relation of activities of neurons in the first layer under attentional conditions. In chapter 5, analyses of experimental data are presented that was recorded to test this prediction. The experimental findings are in agreement with this proposal, underlining the explanatory power of the model approach. Furthermore, the data suggests that attentional changes of neurons' activities are brought about by two independent processes, which are shown to be easily incorporated in the model network.

These three chapters, are written in a paper-style format, each having an own introduction and discussion section, where the introduction provides more specialized information needed for the

understanding of the chapter. At the end, a general conclusion will place the research contained in this part in a wider scope and summarize the scientific achievements.

The three main chapters are preceded by a short overview of the physics of neuronal networks, introducing some tools of the trade of the theoretical neuroscientist.

2 PHYSICS OF NEURONAL NETWORKS

2.1 NEURONS AND SYNAPSES

The following paragraphs will briefly introduce the morphology and functional properties of neurons to an extent which is sufficient for understanding the research presented in later chapters. More comprehensive accounts can be found in any standard neuroscience textbook, e.g. [Kano0].

neuron morphology

The neuron is a specific cell type found in the brain and the whole nervous system including the spinal cord and the enteric nervous system. Morphologically, it consists of a central cell body (soma) housing the cell nucleus and other typical cellular components, usually highly branched appendices called the dendrites, and one longer and less branched appendix termed the axon. When the neuron is embedded in a network, connection sites (synapses) between the axon of one neuron and a dendrite of another are formed. Figure 2.1 A) shows a schematic drawing of neurons and synapses.

electrical properties

The functional state of the neuron depends on its membrane potential, the potential difference between the inside and the outside of the cell body separated by a lipid bilayer membrane. Note that the terminology is a bit confusing from a physical standpoint, where it should rather be called the membrane voltage, but the term "membrane potential" is so deeply engrained in neuroscience that I will keep with the tradition. It is determined by Na^+,K^+ and Cl^- ion concentrations that differ across the cell membrane. The concentration of Na^+ ions and Cl^- ions is higher in the surrounding liquid, whereas the one of K^+ is higher on the inside of the cell.

These ions are in principle capable of slowly diffusing through the membrane and eventually abolishing the concentration gradient. However, in the resting state, the membrane potential of a cortical neuron is held at ≈ -65 mV due to an active process that

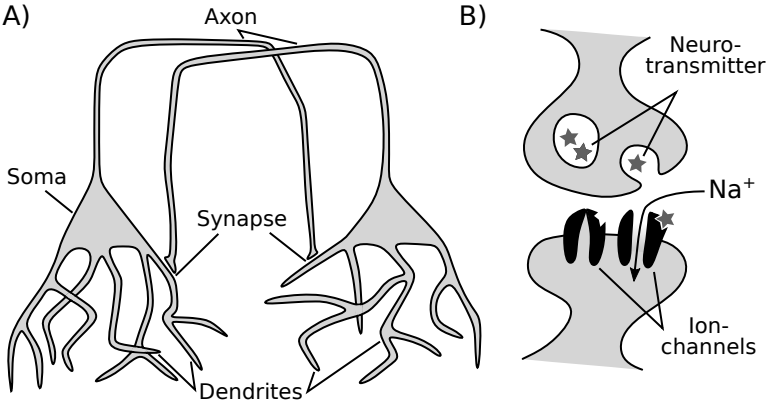

Figure 2.1: Morphology of neurons and synapses. **A)** The cell body (soma) houses most of the typical cellular components. Axons and dendrites are appendices which allow contact points between neurons, called synapses. **B)** At the synapse, neurotransmitter can be released from the axon terminal. Opening of transmitter-gated channels allows the flow of ions (here Na^+) across the membrane, changing the potential of the post-synaptic neuron.

preserves the concentration gradients: sodium-potassium pumps transfer Na^+ ions from the inside of the cell to the surround and at the same time K^+ ions in the reverse direction.

The membrane potential of a neuron is subject to change by selective activation of ion channel proteins in the membrane, mostly through the binding of a chemical substance, a neurotransmitter, to the channel protein. In an activated state, the channel protein forms a pore in the membrane and by virtue of its electrostatic and spatial structure becomes selectively permeable for one or a few ion types (see figure 2.1 B)). An increase in membrane potential (depolarization) is mediated by sodium channels, whereas a decrease (hyperpolarization) relies on chloride and potassium channels. Most channel proteins are located at the dendritic or post-synaptic site of a synapse, whereas the axonal or pre-synaptic site contains vessels filled with neurotransmitters.

spikes

If the membrane potential is depolarized sufficiently to surpass a threshold, a spike is triggered. This sharp and short pos-

itive peak in the potential is initiated at the axon hillock, the region where the axon branches off from the soma. It arises from a sequence of opening and closing of ion channel proteins at different time scales, first leading to a rapid increase of the local membrane potential followed by a strong hyperpolarization which brings the membrane voltage back to the resting potential. In contrast to the transmitter gated channels described before, the channels responsible for the spike are voltage gated, i.e. open and close as a function of the membrane potential.

synaptic weights and Dale's law

The spike travels along the axon and arrives at the axon terminals where the release of neurotransmitter is triggered. The released transmitter in turn activates ion channels located in the membrane the post-synaptic neuron. Transmitters are selective for a specific kind of ion channel, and most neurons only express one type of transmitter, hence it exerts either a depolarizing (excitatory) or hyperpolarizing (inhibitory) influence on other neurons. This principle is often referred to as Dale's law. The weight of a synapse, i.e. the amount the membrane potential changes in the post-synaptic neuron when the pre-synaptic neuron spikes, can vary strongly and is subject to dynamic changes.

2.2 NETWORKS

By these basic principles of interaction, neuronal networks are formed throughout the brain that convey information with spikes. The topology and interaction strengths within a network are crucial in shaping its functional properties, as well as the time constants that govern the dynamics of single neurons and synapses.

mathematical formulation of neuronal networks

In terms of dynamical systems, a neuronal network of size N can be formalized as

$$\tau \frac{dV_i}{dt} = f(V_i) + f_{rec}\left(\{A_1, ..., A_N\}\right) + I_i$$
$$A_i(t) = h(V_i) \quad .$$

The membrane potential V_i of neuron i changes with a time constant τ (by units a capacitance) as a function of the external input current to each neuron, contained in I_i, a function $f()$ of the membrane potential modelling passive diffusion and voltage

gated processes, and on a function $f_{rec}()$ of activity of the network modelling synaptic interaction. The variable A_i reflects the activity of neuron i which is generated by passing V_i through a non-linearity $h()$.

The complexity of the functions $f()$, $f_{rec}()$ and $h()$ is guided by the level of abstraction at which a network is studied. Choices vary from very detailed, biologically realistic models which take much of the cellular and synaptic machinery into account [HH52], to more coarse grained descriptions that focus on the average behaviour of larger cell groups [WC72].

2.2.1 *Diffusion and voltage gated processes*

Hodgkin-Huxley type model

For a detailed description, let $f()$ be given as

$$f(V_i) = \sum_{l=1}^{L} g_l(V_i)(V_i - E_l) \quad , \tag{2.1}$$

where g_l is the conductance of the membrane for ion type l and E_l the reversal potential for that ion type. The conductances are constant if they model passive leak conductances, can also be functions of V_i if voltage gated channels are modelled and even itself be described by differential equations. A prominent example for these conductance based models is the classical Hodgkin-Huxley formulation of spike generation in the squid giant axon [HH52].

In this class of models, the variable A_i reflects the spiking activity of neuron i. A spike of neuron i is detected when V_i crosses a chosen threshold V_{thr} from below, i.e. A_i can be written as

$$A_i(t) = \delta(V_i(t) - V_{thr})\Theta\left(\frac{dV_i}{dt}(t)\right) \tag{2.2}$$

where $\delta()$ denotes the Dirac delta function and Θ the Heaviside function. Since a spike is a stereotypical positive excursion of the membrane potential way above the typical range of V_i, followed by setting back to the resting potential, any value V_i which is exclusive to spikes can be chosen for V_{thr}.

integrate and fire model

The description of $f()$ can be simplified by a polynomial approximation, i.e. setting

$$f(V_i) = \sum_{k=0}^{K} p_k V_i^k \quad , \tag{2.3}$$

which comes at the expense of losing the voltage gated spike generation. Instead, spikes are then not only detected, but also elicited by a threshold mechanism. this means that equation 2.2 still applies, but V_{thr} has to be set to the typical value of V_i at the onset of a spike. Also, to mimic the behaviour of biological neurons, V_i has to be manually reset to the resting potential V_{rest} after a spike occurred. This model class is called integrate and fire neurons. The most widely used variant is the linear integrate and fire model with $K = 1$. In chapter 3, a quadratic version ($K = 2$) is used. A detailed and mathematically stringent reduction of conductance based neurons to integrate and fire models can be found in [AK90].

2.2.2 *Synaptic interaction*

conductance based synapse

In a biologically realistic setting, the recurrent interaction term can also be modelled in a similar way to equation 2.1:

$$f_{rec}(\{A_1, ..., A_N\}) = \sum_{l=1}^{L} g_l(x_l)(E_l - V_i) \tag{2.4}$$

$$x_l = \sum_{j=1}^{N} w_{ij}^l A_j(t - d_{ij}) \quad ,$$

with l now as the synapse type. Every type has a corresponding non-negative weight matrix W^l with entries w_{ij}^l, and the matrix D composed of d_{ij} contains the delays associated with each synapse due to the conduction speed of spikes and the distance between neurons i and j. If E_l is greater than the maximum of the range of $V(t)$ (excluding spikes), the corresponding synapse type is excitatory, and if E_l is smaller it is inhibitory. The conductance g_l is typically modelled as a convolution of the weighted summed

spike train x_l of the pre-synaptic neurons of a synapse type l with a causal kernel $\kappa_l(t)$:

$$g_l(x_l) = (x_l * \kappa_l)(t) \quad .$$

A popular choice is a negative exponential kernel, which is also used in chapter 3.

weight based synapse

If a detailed account of synaptic dynamics is not necessary, equation 2.4 can be simplified to

$$f_{rec}\left(\{A_1, ..., A_N\}\right) = \sum_{j=1}^{N} w_{ij} A_j(t - d_{ij}) \quad . \tag{2.5}$$

This equation is obtained by assuming $\kappa_l(t) = \delta(t)\ \forall l$, replacing $(E_l - V_i)$ by $\mathrm{sgn}(V_i - E_l)$, and absorbing it into the weights w_{ij}^l. Then the set of weight matrices W^l can be collapsed into a single weight matrix W with entries w_{ij}. Thus the impact of a synapse is completely described by its weight.

integrate and fire neurons with weight based synapses

Using both a linear approximation of $f()$ (equation 2.3) with $p_0 = 0, p_1 = -1$ and the simplified $f_{rec}()$ from equation 2.5 leads to a network model given by

$$\begin{aligned} \tau \frac{dV_i}{dt} &= -V_i + \sum_{j=1}^{N} w_{ij} A_j(t - d_{ij}) + I_i \\ A(t) &= h(V) \quad . \end{aligned} \tag{2.6}$$

Here, dV_i/dt turned into a simple linear differential equation which yields some numerical and analytical advantages.

2.2.3 *Population models*

If furthermore the activity of neuronal populations, i.e. a large number of similar neurons, is rather of interest than the behaviour of single neurons, the same equations 2.6 can be used for this purpose, albeit with different interpretations for the variables. V_i can be loosely related to the average membrane potential of neurons in population i, and $h()$ becomes a function that maps V_i to the now average firing rate $A_i(t)$. The activity $A_i(t)$ thus turns

into a continuous variable. Typically, $h()$ is chosen as a bounded, monotonously increasing function of positive range, such as a sigmoid or a piecewise linear function. That these models are actually valid descriptions of the activity of neuronal populations was pioneered in [WC72]. A model of this type is used in chapter 4.

This concludes the short introduction to the modelling of neural networks. The last few paragraphs took a quite condensed mathematical tour through several decades of research in theoretical neuroscience and modelling. Should more background knowledge be required: a more extensive introduction to the description of neuronal networks can be found in [DA01].

2.3 NEURONAL OSCILLATIONS

functional role of oscillations

Oscillations are one among many interesting behaviours neuronal networks can exhibit. Emerging from synchronization of spikes or firing rates among a group of neurons, oscillations are ubiquitous in biological neural networks [BD04]. They naturally occur in different frequency bands and are hypothesized to play important roles for brain function. For example, the prominence of oscillations in the $7-12$ Hz range (so called α range) is correlated with wakefulness. The interplay of oscillations in the θ ($5-7$ Hz) and γ ($40-100$ Hz) range in the hippocampus is modulated during memory recall [Tor+09]. γ oscillation have also been hypothesized to play an important role in selective information routing throughout the cortex, which will be discussed and investigated in depth in chapters 3 and 4.

emergence of γ oscillations

In principle, γ oscillations can emerge in a neuronal network by a strongly interconnected inhibitory population of neurons (interneuron network γ (ING) mechanism), or by an excitation - inhibition loop, where excitatory neurons activate inhibitory neurons which in turn inhibit the excitatory population (pyramidal-interneuron network γ (PING) mechanism) [BW12]. In both cases, the network frequency depends on the conduction delay and synaptic time constants of the involved neurons.

2.4 EARLY VISUAL SYSTEM

In the studies described in chapters 3 and 5, activity was modelled respectively recorded in areas of the early visual system, which is why a short introduction into the relevant anatomy and physiology will follow.

In humans and primates, whose conscious percept of the world is largely dominated by vision, the visual system is among the largest and most complex structures in the brain. The visual system, apart from the eyes, consists of various subcortical and cortical visual areas, and considerable effort was and still is directed towards uncovering their exact function and interactions. As definition of a visual area, a set of neurons, grouped together by spatial proximity and similar response properties shall suffice at this point.

2.4.1 *Anatomical structure*

the eye

The first stage of the visual system is the eye. Here, a two-dimensional image is projected onto the retina, which contains an array of specialized receptors reacting to electromagnetic waves in the visible spectrum. By these receptors, physical properties of light are transformed into electrochemical signals. Following some early processing, signals are ultimately passed on to the ganglion cells, neurons whose axons are bundled in the optic nerve and later in the optic tract. These axons terminate in the lateral geniculate nucleus (LGN).

LGN, V1, V2 and V4

The LGN is a structure that resides in the sensory thalamus and its main feed-forward projection target is the primary visual cortex (V1). From V1, directly and via visual area 2 (V2), the signal ultimately reaches neurons in visual area 4 (V4). These early visual areas are the biological counterparts of the networks modelled in this thesis. It shall be noted though that the hierarchical feed-forward structure is a simplification, since feed-back connections from higher to lower cortical areas exist [FV91]. For a more comprehensive account of the early visual system see [Kano0] chapters 26 and 27.

2.4.2 *Receptive fields and tuning*

receptive field

As noted earlier, neurons can be classified into areas by their response properties. The response property of a neuron is typically characterized by its receptive field. The receptive field is, in an abstract definition, the n-dimensional subspace or lower dimensional manifold of an n-dimensional physical stimulus space in which a significant deviation of the neuron's firing rate from baseline is observed. The baseline firing rate is the rate which is measured in the absence of external stimulation. For some neurons, the receptive field can be subdivided into an excitatory part, that elicits a higher firing rate than the baseline activity, and an inhibitory part, that suppresses the activity of the neuron. The excitatory part of a receptive field is also sometimes referred to as the classical receptive field. Dimensions in which the visual stimulus space is typically described encompass two spatial dimensions of the image that is processed (spatial receptive field), spatial and temporal frequency (e.g of a grating with alternating luminance), colour, contrast, and others. Throughout this thesis, mostly the spatial receptive field will be considered and thus the term "receptive field" on its own will refer to the spatial receptive field.

Retinal ganglion cells typically exhibit a circular centre - surround receptive field organisation, meaning that presentation of a stimulus in the centre elicits a strong response, whereas a stimulus in the immediate surrounding of the centre suppresses the response (on-centre cell) or vice versa (off-centre cell). This can be understood by a simplified picture in which receptors covering the spatial centre of the receptive field have an excitatory (inhibitory) synapse onto the on-centre (off-centre) ganglion cell and receptors in the surrounding have an inhibitory (excitatory) synapse.

convergence and receptive field growth

From the previous example it becomes clear that the receptive field of a neuron can be shaped by its afferent connections. A common motif in the visual hierarchy is that several neurons in a lower area form feed-forward synapses with the same neuron in a higher area, entailing that receptive field sizes tend to grow along the visual hierarchy (also compare figure 1.1): The size of

a receptive field of a retinal ganglion cell close to the fovea is several arc minutes of visual angle ([Kan00] p. 517), around 0.8 degrees for a V1 neuron [CBM02] and approximately 2 degrees for a V4 neuron [Mot09]. Neurons with receptive fields in the visual periphery, i.e the visual space far away from the fovea, show an even more pronounced growth in receptive field size, such that several V1 non-overlapping receptive fields fit into a V4 receptive field (see e.g. [MD85]). A general observation is that receptive field sizes approximately double at any given eccentricity when going from V1 to V2 and also when going from V2 to V4 [GGS81; GSG88].

retinotopy

Another aspect of this convergent connectivity is that the neurons in one area that synapse with the same neuron in a higher area are typically in spatial proximity. Given that this principle holds more or less strictly for all connections from the retina up to V4 leads to retinotopy in early visual areas [Mot09], which relates the location of spatial receptive fields to the visual space: neurons close by in V1, V2 and V4 likely have a similar spatial receptive field.

As the size of receptive fields grows when ascending the visual hierarchy, so does the complexity. Where retinal ganglion cells show mainly on-centre or off-centre receptive fields, a class of V1 neurons responds best to oriented line segments, e.g. generated by a luminance Gabor filter. This receptive field structure is presumably due to the neurons receiving afferent input from ganglion cells via LGN that cover this oriented region with their excitatory centre. This concept also holds for the transition from V1 to V4, where some neurons respond best to shapes that are made up of combinations of oriented line segments [Mot09; PC99]. This simple model of feed-forward shaping of receptive fields is not uncontested (e.g. [HV06]), but further details are of no importance for all intents and purposes in this thesis.

tuning

The concept of tuning is closely related to the receptive field. A neuron is considered tuned with respect to dimension d of the receptive field if the function mapping the receptive field to the neuron's activity is *not* flat in this dimension. For example, a stimulus can elicit variably strong responses at different location within the spatial receptive field. The corresponding neuron is

thus spatially tuned. On the other hand, if the stimulus evokes the same response at every location inside the receptive field, it has no spatial tuning. A tuning many visual neurons share is that for contrast, where a stimulus typically evokes a higher response at a higher contrast irrespective of other tuning properties [Alb+02].

It shall be noted that the terms "receptive field" and "tuning curve" are not consistently defined in the literature and are sometimes used interchangeably, so the definitions given here are relatively general and encompassing, but not agreed upon.

2.5 ELECTROPHYSIOLOGY

Since this part of the thesis navigates the cross section between theoretical neuroscience and electrophysiology, also containing data analysis from experiments, a short introduction will follow on how the data is obtained.

Invasive electrophysiological recordings involve insertion of an electrode into the brain tissue. For this work, only extracellular recordings are important. Here, the electrode is placed in the area of interest, presumably in between intact neuronal cell bodies, as opposed to patch clamping techniques where the membrane of neurons is punctured and intracellular fields are recorded. The electrical field the electrode measures contains contributions from spikes, synaptic currents and other electrical processes of neurons close to the electrode position.

spikes and MUA

Since spikes are very rapid excursions in the membrane potential on a millisecond timescale, they can be extracted from the recorded signal by high pass filtering above $\approx$ 500 Hz. Approximately 100 neurons in the neighbourhood of the electrode tip contribute to this filtered signal [Buz04]. By thresholding, a spike train can be generated which is called a multi unit activity (MUA), since up to 10 neurons generate spikes of sufficient amplitude to be clearly distinguishable from background noise. The spikes of pyramidal cells, which are named for the shape of their central cell body, are overrepresented in this sample because transmembrane currents mostly flow along the axonal-dendritic axis, they are aligned which each other, and have large cell bodies [Buz04].

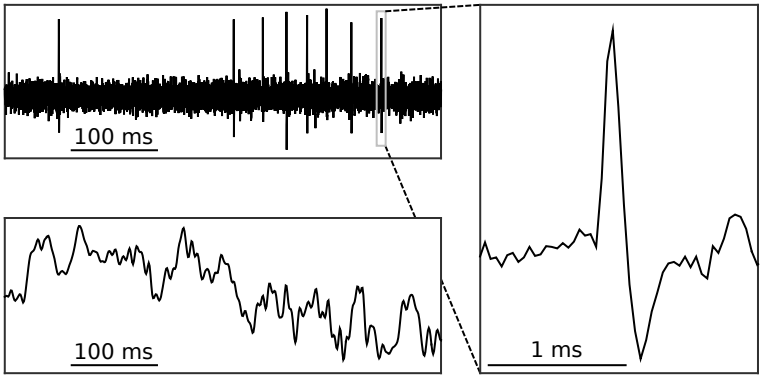

Figure 2.2: MUA and LFP signals extracted from experimental data from a recording in V1. The top shows the high pass filtered raw data above 500 Hz. Spikes can be readily seen against the background noise. The right plot shows a zoom onto one single spike in the signal. The lower plot shows the low pass filtered raw signal below 500 Hz. Data recorded by Eric Drebitz.

Pyramidal cells are excitatory. On the other hand, inhibitory neurons have smaller cell bodies and less clear cut geometry. In addition, pyramidal cells are more abundant in the cortex, at a ratio of approximately 4:1, than inhibitory cells [Bea+92]. Due to these reasons, the MUA can be interpreted as an averaged activity of a small local population of excitatory neurons.

local field potential (LFP)

The local field potential (LFP) is extracted as the low frequency component of the raw signal below ≈ 500 Hz. The LFP is thought to mostly represent an average over synaptic currents within a local population, since the high-frequency spikes are filtered out. However, it cannot be clearly identified with the input from other populations, since the output of a population can partially become its input again due to recurrence in local networks. Also, when neurons spike approximately synchronously, these "population spikes" considerably contribute to the LFP. Thus, the LFP is not as defined as the MUA, and there is still research into how exactly it is composed under different network states and over which spatial region it averages [Ein+13]. Leaving this aside, in

the common interpretation it is a measure of the average activity of a local population bigger than the one represented in the MUA.

Figure 2.2 shows examples of MUA and LFP signals extracted from experimental data.

Extensive reviews of the biophysics of extracellular fields and the contributions to the LFP and MUA are given in [BAK12; Buz04].

3

THE RISER MODEL OF ATTENTIONAL SELECTIVE ROUTING

3.1 INTRODUCTION

The following chapter is adapted with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

It is a wide-spread belief that the brain has limited resources available for information processing [MI05; BK15b]. Selective attention provides a means of efficient and flexible operation under this restriction by allowing preferential allocation of processing resources to a particular area or feature in the visual scenery while ignoring distractors.

In several experimental studies, the effect selective visual attention has on neuronal activity was probed in a paradigm where two stimuli are presented within the receptive field of a recorded neuron in V4, but far enough apart to be represented by two separate populations in V1. By the arrangement of receptive fields, it is assumed that the two activated populations in V1 project (possibly indirectly) to the population in V4 (see figure 3.1).

biased competition

This experiment was performed in two variations. In the first, the two stimuli were chosen to be a preferred stimulus for the V4 neuron, i.e. eliciting a high firing rate, and a non-preferred one, eliciting a low firing rate. It was found that presenting both at the same time results in an averaged rate between the rates the stimuli would elicit if presented alone. Attending one stimulus biased the response towards the rate the attended stimulus would evoke if presented alone [MD85]. This effect was dubbed biased competition.

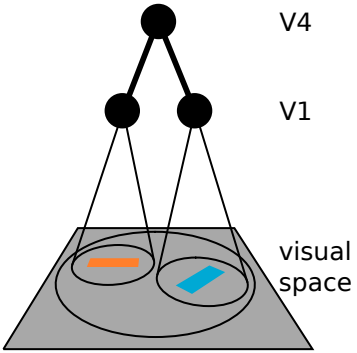

Figure 3.1: Experimental setup to probe effects of selective attention. Two stimuli, here coloured oriented bars, are presented in the receptive field of a V4 neuron, symbolized by the big circle. The stimuli were separated enough to be processed by distinct populations in V1, whose receptive fields are symbolized by the small circles.

information routing

In a second variant, two equally preferred stimuli for a V4 neuron were presented. By independently varying the luminance of the two stimuli, it was found that the luminance modulation of the attended stimulus was represented in the activity of the V4 population, the modulation of the non-attended stimulus however to a much lesser degree [Gro+15]. This effect will be termed information routing.

Assuming that the triangle network setup sketched in figure 3.1 is accurate, the results could be explained by "silencing" the population in V1 processing the distractor. However, such drastic firing rate effects in V1 have not been observed. Thus, both studies suggest that the effective influence of the V1 population processing the attended stimulus is increased, whereas the influence of the V1 population processing the non-attended one is decreased.

Following the experiments on biased competition, it was suggested that the effect could be explained by a five-fold increase of synaptic weights from the attended V1 population to V4 [RCD99]. However, known mechanisms that change synaptic weights, such as long term potentiation, are by far not fast enough to serve as a possible explanation for changes on the time scale of sev-

eral hundred milliseconds on which selective attention operates [WAH00].

communication through coherence (CTC)

One hypothesis about how synaptic efficacies could be modulated dynamically and quickly is based on the idea of binding by synchrony [Sin99; Mal81]. After finding that presentation of stimuli elicited a strong oscillatory response of both firing rate and LFP in the visual cortex in the γ frequency range [GS89], it was hypothesized that phase-locked oscillatory activity of several neurons forms transient ensembles that are engaged in the same computational task. The concept of CTC [Fri05; Fri15] builds on this theory: a neuron with oscillating membrane potential can be driven to fire by relatively weak input when it is close to the threshold, while if it is in a phase of low excitability, i.e. directly after it spiked or when the membrane potential is far from threshold, even stronger input might fail to elicit an action potential. Thus, shifting relative phases of oscillating sending and receiving populations, thereby changing the effective influence of the synapses, could be utilized to gate information.

experimental support of CTC

Experimental support is provided by the findings that attention increases the synchrony of γ oscillations within V_4 [Tay+05; Fri+01; Fri+08], that communication between different populations oscillating in the γ band is most effective for a specific phase relation [Wom+07], and that the γ coherence between neuronal ensembles in V_1 and V_4 is elevated under attention [Gro+12; Bos+12]. Also, in [Gre+09] it was shown that CTC might be used for effective communication between the frontal eye field and V_4. Simultaneous recordings of single neurons in V_1 and V_2 with overlapping receptive fields furthermore showed that spikes in V_2 are most likely to occur when afferent V_1 spikes are synchronized [ZK15]. In [Ni+16], γ oscillations were artificially enhanced in V_4 by optogenetic stimulation, leading to a stronger response of neurons to visual input.

models implementing CTC

Several models were conceived that exploit oscillatory dynamics to explain information routing or biased competition [Mas09; AK10; BK08; BEK08; TS10; WS12; ZFG08; MKW12; MFS06; WT17]. However, there is no model that has been shown to reproduce both key experimental findings of biased competition and information routing in a unifying framework while implementing a

phase shifting mechanism in a biologically plausible manner. In this study, it is hypothesized that feed forward networks of recurrently coupled, mixed excitatory and inhibitory neurons with distance dependent lateral interaction are capable of achieving just that.

By constructing a two-layered, fan-in feed-forward system including lateral connectivity, which mimics the convergent architecture of the visual system, it is shown in this chapter that intrinsically generated oscillations self-organize their relative phase relations to optimize information transmission for attended stimuli. This enables the model to reproduce both biased competition and information routing. Querying the robustness of the findings confirms that both effects are observed over a broad range of relevant model parameters. Furthermore, different working regimes of the model are identified, enabling both mixed and bistable representations of the competing stimuli, which establishes a link to the occurrence of multistable perception phenomena, such as binocular rivalry.

3.2 MATERIALS AND METHODS

3.2.1 *Neuron and synapse model*

Both excitatory and inhibitory neurons were modelled as quadratic integrate and fire units with a membrane potential described by

$$C_m \dot{V} = p_2 V^2 + p_1 V + p_0 + g_e(E_e - V) + g_i(E_i - V) \quad ,$$

where the membrane capacitance C_m is calculated as

$$C_m = 1\frac{\mu F}{cm^2} A \quad .$$

A denotes the surface area of the neuron. The surface of excitatory neurons (A_e) is bigger than the surface of inhibitory ones (A_i). Numerical values for A_e and A_i are in accordance with [Pos+08] (see table 3.1). The dynamic equations for excitatory and inhibitory neurons only differ in the value of C_m. The param-

eters $p_{0,1,2}$ were found by mathematical reduction of a Hodgkin-Huxley type model following [AK90] similar to the one used in [Bar+02] to model inhibitory neurons generating γ rhythms in the cortex. The variable g_e (g_i) is the excitatory (inhibitory) conductance with respective reversal potential E_e (E_i), governed by

$$g_e(t) = w_e \sum_{s=1}^{n_e} \Theta(t - t_{s,e} - d) \exp\left(\frac{-(t - t_{s,e} - d)}{\tau_e}\right)$$

$$g_i(t) = w_i \sum_{s=1}^{n_i} \Theta(t - t_{s,i} - d) \left[\chi_1 \exp\left(\frac{-(t - t_{s,i} - d)}{\tau_i^1}\right) + \chi_2 \exp\left(\frac{-(t - t_{s,i} - d)}{\tau_i^2}\right)\right] .$$

Here, $t_{s,e}$ denote the times of presynaptic excitatory spikes, and $t_{s,i}$ the times of inhibitory ones. Θ is the Heaviside function. The synaptic delay is symbolized by d. Furthermore, in accordance with [Bar+02], the synaptic weight w_i of inhibitory connections is generally taken to be stronger than the weight w_e of excitatory synapses and the response to an inhibitory spike has a fast and a slow component, where relative contributions are controlled by $\chi_{1,2}$ with $\chi_1 + \chi_2 = 1$. The values of w_e and w_i were chosen such that the network is in an oscillatory activity regime under visual stimulation. If the membrane potential crosses V_{thr}, a spike is generated and the potential set back to V_{rest}. The model parameters and their default values can be found in table 3.1. If not stated otherwise, the default values were used.

3.2.2 *Local network structure*

Local populations were formed as recurrent networks of 800 excitatory and 200 inhibitory neurons (ratio of 4:1 [Bea+92]) with sparse connectivity probabilities p_{loc}^{ie} and p_{loc}^{ii}, i.e. recurrent connections only exist within the inhibitory population and from the inhibitory to the excitatory population. It is assumed that neurons making up a local population represent a patch of visual cortex that is activated by a circular stimulus of one degree diameter. With retinotopy and a cortical magnification factor of

1 deg/mm [Alb75; TPR78], this translates to a stimulated cortical area of 1 mm diameter. All recurrent connections within the local population are inhibitory, which were found to have a conduction speed of approximately 0.1 m/s [SP96]. Thus one would expect conduction delays to range from 0 to 10 ms. For simplicity all delays d are set to the mean of 5 ms. Figure 3.2 A) shows a schematic drawing of the basic local circuitry.

Input to the population is delivered via afferent synaptic connections that impinge onto both excitatory and inhibitory neurons [Zem+13; HM07; Lee+14]. This setup can be seen as an extended implementation of the dynamical network motif of feed forward inhibition, which is frequently encountered in local cortical, cotico-cortical and thalamo-cortical circuits [Wom+14]. The most stripped down version of this motif only has an inhibitory connection onto the excitatory subpopulation and common input driving both excitatory and inhibitory subpopulations. The functional interpretation is that potential oscillations in the input stream are selectively extracted: the inhibitory subpopulation resonates with the input oscillation, and, with a certain conduction delay, rhythmically inhibits the excitatory subpopulation. Since the excitatory subpopulation receives the same input as the inhibitory one at the same time, but is inhibited with a delay, the peak that causes the inhibitory population spike is passed through the excitatory subpopulation. Consequently, such circuitry is prone to selectively gate oscillatory input in a certain frequency range, depending on neuronal time constants and the delay of local inhibition. The present implementation is extended in the sense that the inhibitory subpopulation generates a local γ oscillation via recurrent inhibition, which is imposed onto the excitatory population, even in the absence of oscillations in the input. The design of this inhibitory subnetwork stems from a model of γ generating circuits in the neocortex [Bar+02].

Figure 3.2 B) shows example mean firing rate traces and rate distributions of the excitatory and inhibitory subpopulation when driven by 135 independent Poissonian spike trains, each at a rate of 13 Hz, revealing stable γ oscillations. Whereas most inhibitory neurons fire slightly below the population frequency of the network, the excitatory neurons fire typically at markedly

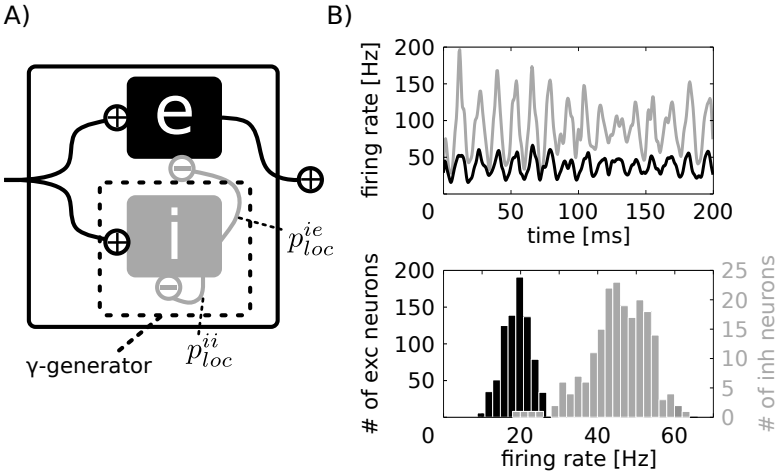

Figure 3.2: Local network setup. **A)** A local population consists of externally driven, recurrently coupled 800 excitatory and 200 inhibitory neurons. Recurrent synapses in the inhibitory subpopulation, instantiated with coupling probability p^{ii}_{loc}, generate a γ rhythm, which is imposed onto the excitatory subpopulation via connections with probability p^{ie}_{loc}. **B)** A local population generates a stable γ population-rhythm, seen in the exemplary time courses of activity of the inhibitory (grey) and excitatory (black) subpopulation (top). Inhibitory neurons fire at markedly higher rates than excitatory ones (mean firing rate histogram, bottom). Adapted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

lower rates, as observed in electrophysiological studies [Csi+98; Csi+99; Ker+10; Hof+11; Vin+13; CWG02; MSR07].

3.2.3 *Global network structure*

To emulate the converging feed-forward characteristic of the visual system, four local populations A,B,C, and D are arranged in two layers. In the first layer, populations A and B are driven

by stimuli S_A and S_B. The excitatory subpopulations of A (A_{exc}) and B (B_{exc}) project to C and D in the second layer. Pairs AC and BD are connected by equal probabilities $p_{AC} = p_{BD}$, whereas the strength of "crosstalk" between pairs AD and BC is determined by $p_{AD} = p_{BC} = \mu \cdot p_{AC}$, with $\mu \in [0, 1]$. The mixing variable μ controls the relative preference of the stimuli to the populations in the second layer. The fact that in the first layer, the populations are driven directly only by one stimulus, but by both in the second layer, reflects the growth of receptive field sizes along the visual stream [KE13].

Lateral interaction is implemented within the layers as structured according to the tuning properties of populations. Feature specific lateral coupling has been proposed to account for various computational effects, for example iso-feature suppression in V1 [Li05] or contrast invariance of orientation tuning curves [BLS95]. For a variety of features, specificity of lateral connections in visual cortex has been confirmed, showing that columns with similar tuning properties are more likely to be interconnected [Bos+97; CF04; Mal+93].

Here, the same is presumed to hold true when the feature is visual space. Hence, lateral coupling between populations with overlapping receptive fields should be different from lateral coupling between populations with non-overlapping receptive fields, following the reasoning in [DR05], where a similar model setup to study biased competition with spiking neurons in the absence of oscillations was used. In that manner, I assume that between A and B with non-overlapping classical receptive fields, lateral interaction is mediated by cross-coupling from excitatory to inhibitory subpopulations. Lateral interaction in the second layer, where receptive fields overlap, is implemented by coupling the inhibitory subpopulation of C to the inhibitory and excitatory subpopulations of D and vice versa.

This particular choice is made for two reasons. Firstly, the lateral coupling between populations C and D is similar to the coupling *within* a local population (compare figure 3.2 A)), albeit at lower strength, which is consistent with assuming that populations with overlapping receptive fields become closer to act as one single rather than two separate computational units. On the

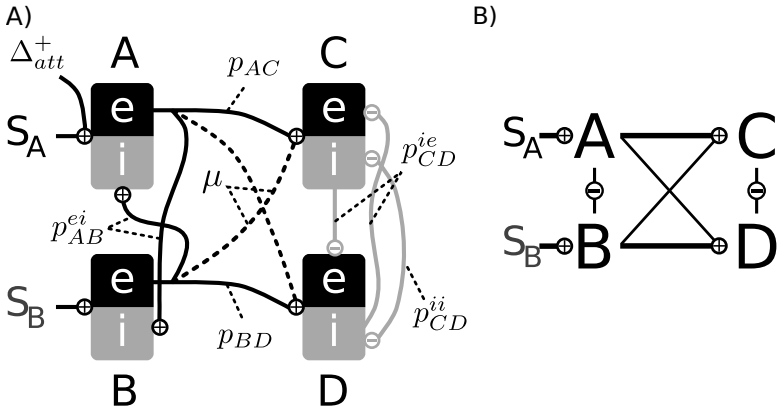

Figure 3.3: Global network setup. **A)** Four interconnected local populations form the two-layered global setup. First layer populations A and B are driven by Poissonian spike trains representing two different stimuli S_A and S_B. Both A and B project to both second layer populations C and D, whereas A and C, respectively B and D, are connected with the same coupling probability $p_{AC} = p_{BD}$. The strength of cross-talk between A and D, respectively B and C, is regulated by μ, i.e. $p_{AD} = p_{BC} = \mu \cdot p_{AC}$ with $\mu \in [0, 1]$. The value of μ determines the relative preference of S_A and S_B to C and D: if $0 < \mu < 1$, S_B (S_A) will be a non-preferred stimulus for C (D). Intra-layer lateral interactions differ between the two layers and depend on connection probabilities p^{ei}_{AB}, p^{ie}_{CD} and p^{ii}_{CD}. Populations C and D are assumed to act as similar computational units due to the partly overlapping receptive fields, thus the lateral connectivity is the same as within local populations (see figure 3.2 A)) at lower strength ($p^{ie}_{CD} < p^{ie}_{loc}$ and $p^{ii}_{CD} < p^{ii}_{loc}$). Attention is introduced as an additional input Δ^+_{att} to one first layer population (here A). All coupling probabilities and model parameters can be found in table 3.1. **B)** Simplified layout diagram of the full setup used in following figures. Adapted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

other hand, populations A and B process different stimuli, and can be interpreted as two separate computational units. Thus, the lateral coupling between A and B is *different* to the coupling within a local population. Secondly, if the layers in the model are retinotopically organized, the distance dependent lateral interaction in visual space translates to distance dependent interaction in cortical space. The connectivity scheme proposed here is then in line with the finding that in early visual cortex, lateral connections reaching beyond the local column are usually mediated by excitatory neurons [RL83], whereas inhibitory projections are confined to the same or neighbouring columns [HSF09].

The delay d of the lateral synapses between C and D is also set to 5 ms, consistent with the assumption that both populations are in close vicinity and act similar to one population. In the first layer, it is assumed that A and B are populations directly adjacent in cortical space, thus processing different stimuli in their classical receptive field while lying in each others suppressive surround. Following the same reasoning as above, the mean distance between neurons in both populations is approximately 1 mm. With a higher conductance speed of pyramidal axons which is set to 0.2 m/s [MGI93; NK95; HG91], this distance also translates to a delay of 5 ms.

The difference in the implementation of the lateral connectivity results in a differential effect on the phase relation between the populations in the first and second layer. In the second layer, C_{exc} and D_{exc} will tend to oscillate in phase, since the lateral coupling is conceived as similar to the coupling within each local population. In the first layer however, A_{exc} and B_{exc} will tend to oscillate in anti-phase. This can intuitively be understood by considering A_{exc} and B_{exc} as mutually inhibiting oscillators, where A_{inh} and B_{inh} merely serve to convert the excitatory output of A_{exc} and B_{exc} to inhibitory signals and add an additional time delay. The travelling time of the inhibitory signals between A_{exc} and B_{exc} is about 14.2 ms (5 ms from A_{exc} to B_{inh} + 5 ms from B_{inh} to B_{exc} and vice versa. In addition to this, there is also the rise time of inhibitory (1.2 ms) and excitatory (3 ms) conductances). Roughly speaking, if the travelling time of the signal is close to the period of the oscillation T, A_{exc} and B_{exc} will tend to push each other

out of phase, since this is the configuration in which the impact of the mutual inhibition is lowest. More precisely, the anti-phase steady state solution of two inhibitory pulse coupled oscillators is stable for delays approximately in the interval from $\frac{3}{4}$T to $\frac{5}{4}$T [KN11]. This situation is given with populations A and B oscillating around 70 Hz, resulting in a period T of approximately 14 ms.

The layer specific lateral connectivities also lead to distinctive effects on the mean firing rate. If populations A and B are both active, they will suppress each other, since the lateral coupling from excitatory to inhibitory subpopulations increases the overall inhibition in the layer. Thus, A and B can be interpreted as acting as each others suppressive surround. In the second layer, populations C and D do not have a strong mutual suppressive effect. The lateral coupling mainly synchronizes C_{inh} and D_{inh} and overall inhibition remains largely unchanged.

Figure 3.3 A) shows the global setup of the model and figure 3.3 B) a simplified layout diagram used in following figures. The values of the connection parameters can be found in table 3.1.

3.2.4 *Stimuli*

The first layer populations A and B are driven by different input stimuli S_A and S_B. These stimuli are modelled as firing rates from which Poissonian spike trains are drawn. Every neuron (excitatory and inhibitory) in A and B receives 135 input synapses, that each transmit a Poissonian spike train at a mean rate of 13 Hz. Around the mean, the firing rates of S_A and S_B are independently modulated by a uniformly distributed random process with an amplitude of 2 Hz, where every 10 ms a new value is drawn. This models the effect of the luminance "flicker"-signal used in the information routing experiment [Gro+15].

3.2.5 *Attention*

In accordance with many experimental studies showing that directing attention towards a stimulus leads to a moderate increase

variable	A_e	A_i	p_0	p_1	p_2
(default) value	$2.88x10^{-4}$ cm^2	$1.2x10^{-4}$ cm^2	$3.90x10^{-9}$ A	$1.30x10^{-7}$ $\frac{\text{A}}{\text{V}}$	$1.08x10^{-6}$ $\frac{\text{A}}{\text{V}^2}$
variable	V_{thr}	V_{rest}	τ_e	τ_i^1	τ_i^2
(default) value	-56.23 mV	-67.00 mV	3 ms	1.2 ms	8 ms
variable	χ_1	χ_2	E_e	E_i	w_e
(default) value	0.9	0.1	0 mV	-75 mV	0.4 nS
variable	w_i	p_{loc}^{ie}	p_{loc}^{ii}	p_{AC}	μ
(default) value	1.2 nS	0.2	0.2	0.1125	0.5
variable	p_{AB}^{ei}	p_{CD}^{ie}	p_{CD}^{ii}		
(default) value	0.08	0.10	0.10		

Table 3.1: Default model parameters. Symbols: $A_{e,i}$: Surface area of excitatory respectively inhibitory neurons. $p_{0,1,2}$: Parameters of the quadratic integrate-and-fire model. V_{thr}: Firing threshold. V_{rest}: Reset potential. τ_e: Excitatory synaptic time constant. $\tau_i^{1,2}$: Inhibitory synaptic time constants. $\chi_{1,2}$: Relative contribution of fast and slow components to the inhibitory response. $E_{e,i}$: Excitatory respectively inhibitory reversal potential. $w_{e,i}$: Excitatory respectively inhibitory synaptic weight. p_{loc}^{ie}: Connection probability from inhibitory to excitatory neurons in a local population. p_{loc}^{ii}: Connection probability between inhibitory neurons in a local population. p_{AC}: Connection probability from excitatory neurons in local population A to all neurons in local population C. μ: Cross talk parameter ($p_{BC} = \mu \cdot p_{AC}$). p_{AB}^{ei}: Connection probability from excitatory neurons in A to inhibitory neurons in B and vice versa. p_{CD}^{ie}: Connection probability from inhibitory neurons in C to excitatory neurons in D and vice versa. p_{CD}^{ii}: Connection probability between inhibitory neurons in C and D.

in firing rate in early visual neurons driven by that stimulus (e.g. [Cha+10; MCW08; Thi+09; Her+08]), both in putative excitatory and inhibitory neurons [MSR07; Vin+13], attention is manifested as a supplementary input Δ^{+}_{att} to both excitatory and inhibitory subpopulations in the respective local population (A or B) that processes the attended stimulus. More specifically, the supplementary input raises the mean rate of the Poisson process representing the attended stimulus by 1 Hz to 14 Hz.

3.2.6 *Simulation and analysis*

Neural activity of networks was simulated in epochs of 2.4 s with a time step of 0.1 ms. The simulations were realized in the framework of the BRIAN neural network simulator (V1.4.1, `http://briansimulator.org/`) [GB09]. Every simulation for each set of parameters was performed 50 times with randomized initial conditions. The trial length and number of trials per condition are chosen to approximately match the original information routing experiment [Gro+15].

When rates of neurons in response to a stimulus are considered, it refers to the firing rates averaged over time, trials and neurons within excitatory subpopulations.

intermediate response factor (IRF)

To quantify the biased competition effect, two different scores are introduced. The intermediate response factor (IRF) quantifies the relation of the rate at which neurons fire when both stimuli are present (r_b) to the rate observed when only the preferred (r_p) or the non-preferred (r_{np}) stimulus is present as

$$\frac{r_b - r_{np}}{r_p - r_{np}} \quad .$$

A value of 0.5 thus describes the situation when the response to both stimuli is exactly between the two rates elicited by the preferred respectively non-preferred stimulus alone. Here, information about both stimuli is represented to the same degree in the firing rate, concurrent with the biased competition effect.

biased competition score (BCS)

The biased competition score (BCS) indicates the effectiveness of rate recovery under attention. The cases of "attend preferred"

and "attend non-preferred" are considered separately. In the former case, the biased competition score is defined as

$$\frac{r_b^{att\,p} - r_b}{r_p - r_b} \quad ,$$

in the latter case it is

$$\frac{r_b^{att\,np} - r_b}{r_{np} - r_b} \quad ,$$

where $r_b^{att\,p}$ ($r_b^{att\,np}$) is the rate when both stimuli are present and the preferred (non-preferred) is attended. For perfect rate recovery, a value of 1 would be observed, and 0 if attention had no effect at all on the rates.

population phase relations

To investigate the phase relation between two neuronal populations or subpopulations, a measure is constructed that is 0 if both populations are in-phase and $\pi / -\pi$ if they are in anti-phase. Due to the conduction delays in the network and a phase lag between membrane current and population spikes, it is not feasible to directly compare the population rates. Instead, first the analytic signal $c_s^a(t)$ of the mean incoming current $c_s(t)$ from all sending neurons is calculated:

$$c_s^a(t) = \mathcal{F}^{-1}(\mathcal{F}(c_s(t)) \cdot 2\Theta(2\pi f)) = c_s(t) + i \cdot \mathcal{H}(c_s(t)) \quad ,$$

where i is the imaginary unit, $\mathcal{F}$ denotes the Fourier transform and $\mathcal{H}$ the Hilbert transform. Θ is the Heaviside function and f symbolizes frequency. In the same way, the analytic signal $c_r^a(t)$ of the mean *total* incoming current $c_r(t)$ to all receiving neurons is computed. The instantaneous phase of a signal can be determined as the angle of the corresponding analytic signal. Thus, the phase difference $\Delta\phi(t)$ between $c_s(t)$ and $c_r(t)$ can be calculated as

$$\Delta\phi(t) = \arg\left(\frac{c_r^a(t)}{c_s^a(t)}\right) \quad ,$$

where argument is mapped to the principal value, resulting in $\Delta\phi \in [-\pi, +\pi)$. In the course of this study only phase differences between excitatory subpopulations are considered. As an example, a phase difference $\Delta\phi$ of 0 between A and C means that

the peaks of the mean current to C_{exc} originating from A_{exc} coincide with the peaks of the total mean current flowing into C_{exc} (i.e. the sum of the mean currents to C_{exc} originating from A_{exc}, B_{exc}, C_{inh} and D_{inh}). In the same way, a phase difference $\Delta\phi$ of $\pi/-\pi$ occurs when the current peaks from A_{exc} coincide with the troughs of the total incoming current to C_{exc}.

spectral coherence score (SCS)

The spectral coherence is a measure that was used to quantify information transmission between luminance signals and the activity of neurons in V4 in the information routing experiment [Gro+15]. For two signals $x(t)$ and $y(t)$ it is defined as

$$SC_{xy}(f,\tau) = \left| \frac{\sum_k \sum_i W_x^{k*}(f,t_i) W_y^k(f,t_i+\tau)}{\sum_k \sum_i |W_x^k(f,t_i)| \sum_i |W_y^k(f,t_i+\tau)|} \right| \quad .$$

Here, W_x^k denotes the wavelet transform of signal $x(t)$ in trial k (the Morlet mother wavelet was used with width parameter 6), * denotes complex conjugation, f symbolizes frequency, τ a time delay, and i the time index. The spectral coherence is collapsed to a single spectral coherence score (SCS), independent of f and τ, by taking the mean over the full frequency spectrum and a set of τ for every frequency f. The width of this region of interest of τ (see non-hatched area in figure 3.11 A)) as a function of f is $\frac{6}{f}$. The centre of the region was found ad-hoc by taking the mean of the maxima of $SC(f,\tau) \forall f \in [25\text{ Hz}, 65\text{ Hz}]$, and was the same for all frequencies f.

Chance levels are calculated by the same procedure, pairing luminance signals with firing rates from different trials.

Error bars are obtained by bootstrapping. 50 simulated trials for every parameter combination from which the spectral coherence score is calculated are assumed to be a representative sample of the distribution. Bootstrap case resampling [Efr79] was applied 100 times for every parameter combination and the spectral coherence score calculated for each. Thus for every point in parameter space a distribution is obtained of 100 values of the spectral coherence score. The shaded areas in the plots show the 95% confidence interval.

phase coherence (PC)

The phase coherence (PC) at a specific frequency f and time t between two population rates x and y is also determined, which is given as

$$\mathrm{PC}_{xy}(f,t) = \frac{1}{N_k}\left|\sum_{k=1}^{N_k} e^{i[\arg(W_x^k(f,t))-\arg(W_y^k(f,t))]}\right| \quad .$$

$W_x^k(f,t)$ denotes the wavelet transform of the rate x at the frequency f and time t in trial k, N_k the number of trials, and $\arg()$ the argument function mapped to the principal value. Possible values of the phase coherence lie between 0 (random phase relation) and 1 (complete phase locking). Note that this value does not give information on the actual value of the phase lag between x and y, only on how stable this phase lag is over time and trials. This method is also used in [Gro+15].

Data analysis and visualization was performed using IPython [PG07], Numpy and Scipy [Oli07], and Matplotlib [Hun07].

3.3 RESULTS

When the first layer populations are stimulated, by construction they engage in γ oscillations. If only one stimulus is presented, say S_A, the second layer populations receive input according to their feed forward connection probabilities and show respective firing rates, i.e. a high rate for a preferred and a low rate for a non-preferred stimulus. The first layer population processing the attended stimulus, in this case A, entrains both C and D, setting a favourable phase relation for information transfer (figure 3.4).

Upon simultaneously presenting a second stimulus S_B, the inhibitory interaction in the first layer reduces mean rates of A_{exc} and B_{exc} and drives the populations out of phase. Simultaneously, the interactions between C and D bring both in phase, entailing that if one stimulus is in a favourable in-phase relation, the other will be forced into a non-favourable anti-phase relation with respect to both populations in the second layer. Which first layer population is in a favourable, and which is in a non-favourable phase relation to C and D changes over time, depending on temporal fluctuations in the firing rates of A_{exc} and B_{exc}. It follows

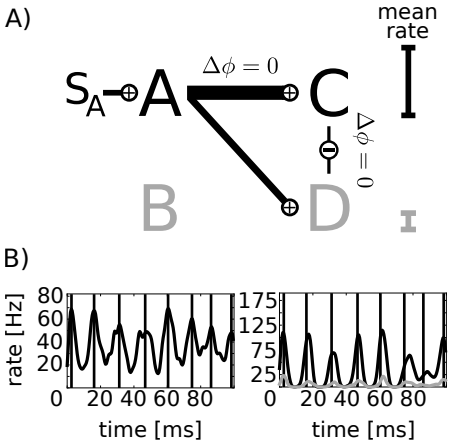

Figure 3.4: The mechanism of information routing and biased competition. **A)** In the presence of one stimulus, both C and D are entrained by the corresponding first layer population. The phase is suitable for information transfer and the mean rates over time of C_{exc} and D_{exc}, symbolized by the bars on the right, scale monotonously with the feed forward connection probabilities. The width of the lines connecting the first and second layer corresponds to total input to the second layer over the respective connection. **B)** Example traces of mean rates of the excitatory subpopulations, where the line colour corresponds to the colour of the populations in the model diagram. The vertical black lines mark the peaks of oscillations in A_{exc}. For better visualization, the firing rate traces in the first layer are shifted to account for transmission delay and the phase shift between incoming currents peaks and rate peaks in the second layer, such that rate peaks in the first layer are aligned with the corresponding peaks caused in the second layer. Please note that the phase differences are calculated between currents, as described in section 3.2.6. Adapted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

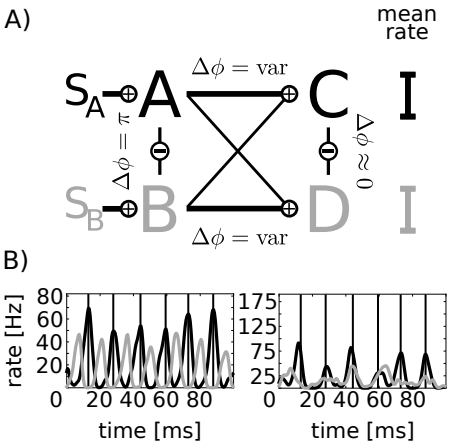

Figure 3.5: The mechanism of information routing and biased competition. **A)** Showing two stimuli brings A_{exc} and B_{exc} out of phase, whereas C_{exc} and D_{exc} tend to stay in phase, due to the respective lateral interactions. The stimuli compete for representation in the second layer. The outcome depends on temporal fluctuations in the peak rates of A_{exc} and B_{exc}. Thus the phases between the first and second layer populations vary over time, favouring either S_A or S_B. Mean firing rates lie at approximately the mean of the two rates of C_{exc} and D_{exc} observed in figure 3.4. **B)** example traces as in figure 3.4. Adapted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

that both rates of C_{exc} and D_{exc} show approximately the same, normalized rate (figure 3.5).

Attending to one stimulus, e.g. S_A, breaks the symmetry by raising the firing rate of A_{exc} and thereby reducing the rate of B_{exc}. The higher output of A_{exc} to the second layer results in a bias for A_{exc} to be in a favourable in-phase relation to transmit information to C, for which, in case of $\mu < 1$, it is the preferred stimulus. At the same time, due to the in-phase relation between C and D mediated by the lateral interaction, it also enters a favourable phase relation with D. Consequently, the weaker output from B_{exc} will be in a non-favourable phase relation to both C and D, hence

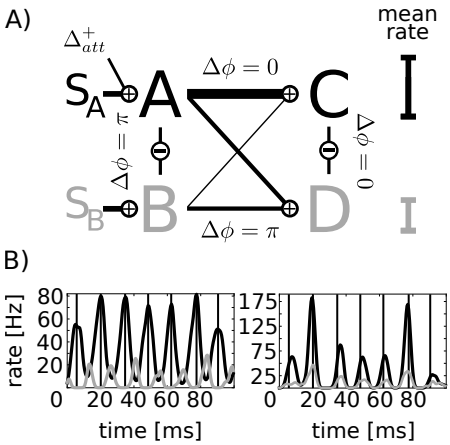

Figure 3.6: The mechanism of information routing and biased competition. **A)** If two stimuli are presented, symmetry is broken by adding an additional input Δ^+_{add} representing the influence of attention. The higher rate of A_{exc} driven by the attended stimulus S_A more likely entrains populations C and D and thus renders input from B ineffective. The mean firing rates of 3.4, where only the attended stimulus is present, are partially recovered. **B)** example traces as in figure 3.4. Adapted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

less likely to elicit spikes in the second layer. Thus it is possible that if a non-preferred stimulus is attended - here S_A with respect to D - the mean rate of the second layer population drops although the preferred stimulus is present at the same time (figure 3.6).

Since synchrony emerges in a self-organized manner rather than being imposed by a common external source, coherent oscillations need a few γ cycles to emerge. This behaviour is most obvious at stimulus onset (figure 3.7), where transient activation is followed by the organization of γ activity. In particular, a stable phase relationship between an attended sending population

in the first layer and the second layer populations takes about 9 cycles ($\approx$ 130 ms) to be fully established, thus imposing a temporal constraint on this CTC scheme. The transient of $>$ 100 ms is well in line with studies demonstrating that attention can only be shifted 3 to 5 times a second [WAH00].

Especially the strength of lateral inhibition in the first layer, which realizes competition between A and B, determines the mode of operation of the network. Figure 3.8 A) shows the biased competition effect and respective phase difference distributions between first and second layer populations for different values of p^{ei}_{AB} . For weak competition in the "both stimuli, no attention" case, weakly modulated monomodal phase distributions with broad peak widths between sending and receiving populations are observed, implying a rather continuous and "mixed" representation of both stimuli over time in both populations C and D. The stronger the cross coupling gets, the more bimodal, peaked and similar the phase distributions become (figure 3.8 B)). At the same time, the distribution of firing rates for the "both stimuli, no attention" case are also getting more and more bimodal, indicating a mode of bistable representation, where one stimulus is exclusively represented in the second layer for an extended period of time (figure 3.8 C)). Under attention, the attended first layer population is almost exclusively in a favourable phase relation and the non-attended one in anti-phase, with virtually no overlap between the phase difference distributions. This holds for all values of p^{ei}_{AB} (figure 3.8 C)). As a consequence, the rate recovery under attention is almost perfect in the bistable regime, i.e. with strong competition in the first layer (figure 3.8 A) bottom).

The emergence of bistability is further illustrated in figure 3.9, which shows example traces of the phase difference between populations A and C and the phase difference between B and C when both stimuli are present, but none is attended. With increase in the lateral inhibition in the first layer, the dominance periods lengthen and the phase differences tend to be closer to the "best" ($|\Delta\phi| = 0$) and "worst" ($|\Delta\phi| = 1$) case, and less in between (compare figure 3.8 B)).

A parameter that is furthermore important in a realistic setting is μ, describing relative stimulus preference, which in reality

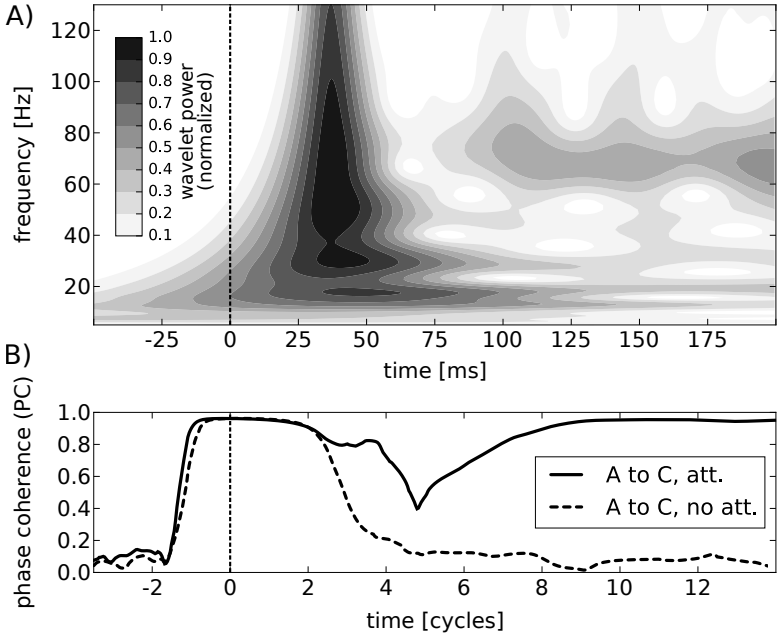

Figure 3.7: Temporal evolution of phase coherence. **A)** Wavelet power spectrum of the mean firing rate of population C_{exc} (attention on S_A), averaged over 100 trials. About 30 ms after stimulus onset (vertical dashed line), the population responds to the input with a strong, transient activation. Approximately 100 ms after stimulus onset, the population settles into a stable γ rhythm with a frequency of about 70 Hz. **B)** Amplitude of the phase coherence between A_{exc} and C_{exc} at a frequency of 70 Hz when S_A is attended (solid black line) and when no attention is involved (dashed black line). The abscissa is given in units of 70 Hz cycles. At stimulus onset, there is a period of almost complete "synchrony" since the feed-forward input arrives in every trial at the same point in time, and leads to the same initial activation. Subsequently, gamma oscillations begin to emerge in both populations which are initially in a random phase relationship (dip around 5 cycles post stimulus onset), and then become more and more coherent in the case of attention on S_A (solid line, after 9 cycles). Without attention, the phases remain on average in a random relationship leading to a low phase coherence amplitude. Adapted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

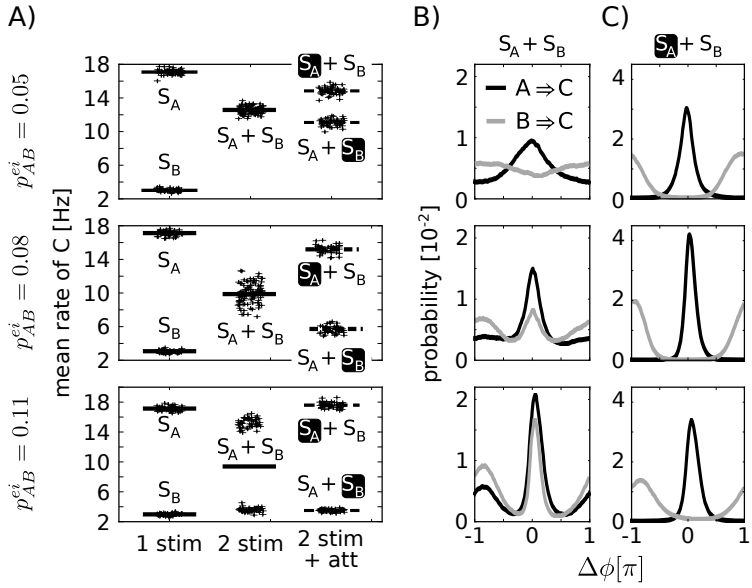

Figure 3.8: Rate- and phase-distributions depend on lateral inhibition. **A)** Mean rates (lines) of population C_{exc} in response to one stimulus, two stimuli present, and two stimuli while attending to one. attentional focus is marked by the black box. Crosses around the mean rates show single trial values, jittered horizontally. The same plot is shown for three different strengths of lateral inhibition between A and B. With increasing p^{ei}_{AB}, the response to presenting S_A and S_B at the same time becomes bimodal and the attentional rate recovery of the rates the stimuli elicit on their own becomes nearly perfect. **B)** The plots show the phase difference distribution between A and C (A ⇒ C) and B and C (B ⇒ C) when both stimuli are present for the same three values of p^{ei}_{AB} as in A). Monomodal phase difference distributions are observed between first and second layer populations for $p^{ei}_{AB} = 0.05$. Population A, representing the preferred stimulus S_A, is more likely to be in-phase and population B, representing the non-preferred stimulus, more likely in anti-phase. Increasing lateral inhibition causes the distributions to become bimodal and more similar, indicating a bistable regime where one stimulus is represented exclusively for an extended amount of time. **C)** The same situation as B), but with attention on S_A. The attended first layer population is mostly in-phase with the second layer, the non-attended mostly in anti-phase, with virtually no overlap of the phase difference distributions. This effect is independent of the strength of lateral inhibition in the first layer. Adapted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

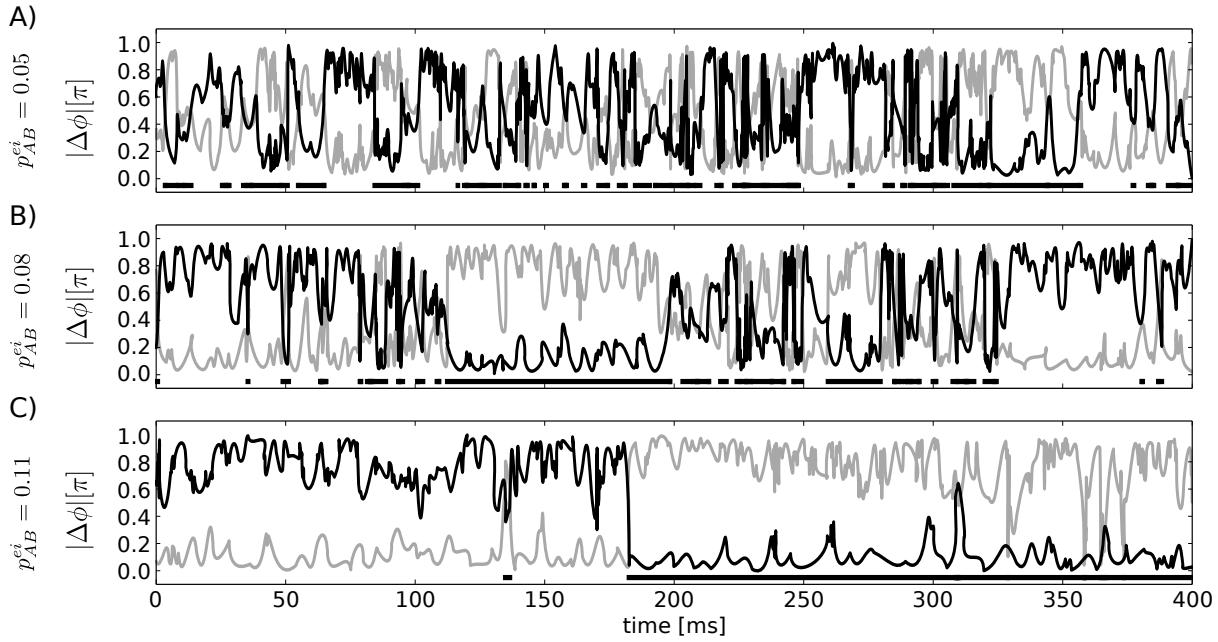

Figure 3.9: Example traces for absolute phase differences $|\Delta\phi|$ between A and C (black line) and B and C (grey line) for weak (**A)**), intermediate (**B)**) and strong (**C)**) lateral inhibition in the first layer. Both stimuli are present but none is attended. The horizontal black bars indicate the time periods in which $|\Delta\phi_{AC}| < |\Delta\phi_{BC}|$, i.e. where population A is in a more favourable position to transmit information to C than B. With rising lateral inhibition, these dominance periods lengthen and the absolute phase differences tend to lie more closely to the edges of the spectrum, i.e. 1 respectively 0 (compare figure 3.8 B)). Reprinted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

depends on the placing of the stimuli in visual space, their shape, and the recording position in the cortex. Figure 3.10 shows the biased competition effect on firing rates over the full range of possible values for μ. Here it is found that the general effect is present for all values (A)), and that the intermediate response rate - when both stimuli are present but none is attended - lies approximately at the mean of the rates when only the preferred or non-preferred stimulus is present (B)), as reported in the literature [RCD99; MD85]. The biased competition score, showing how good the recovery of the single stimulus responses under attention is, lies around 0.6 to 0.8 for intermediate values of μ, with a bias for the higher, preferred rate. This non-symmetric split up was also noted in [RCD99]. In the present model, this feature comes about due to the way attention is applied. Increasing the overall input to the first layer consequently increases the overall output to the second, increasing the activity averaged over the whole second layer.

Also information transmission is investigated for the case that both stimuli are present and S_A is attended in the same parameter regimes, using the method from the original information routing experiment [Gro+15]: the spectral coherence is computed between the two competing, temporally modulated input stimuli and the neuronal activities in the second layer. Figure 3.11 A) shows an example of the spectral coherence for $\mu = 1$. There is considerably stronger information flow from the attended stimulus S_A to the second layer populations than from the unattended stimulus S_B.

Again the variation of the parameter μ was considered. For sufficiently large values of μ (≥ 0.3), i.e where the stimulus S_A elicits a significant response in D (see figure 3.10 A)), the attended stimulus clearly dominates the activity of the second layer populations, specifically for $\mu = 1$, which corresponds to the stimulus placement in the original experiment (figure 3.11 B)).

In addition to these primary results, also the general stability of both effects was investigated for variations of the connectivity parameters. Figure 3.12 illustrates the biased competition of mean rates and information routing for variations of p^{ei}_{AB}, p^{ie}_{CD} and p^{ii}_{CD}. Qualitatively, the principal effects were observed across

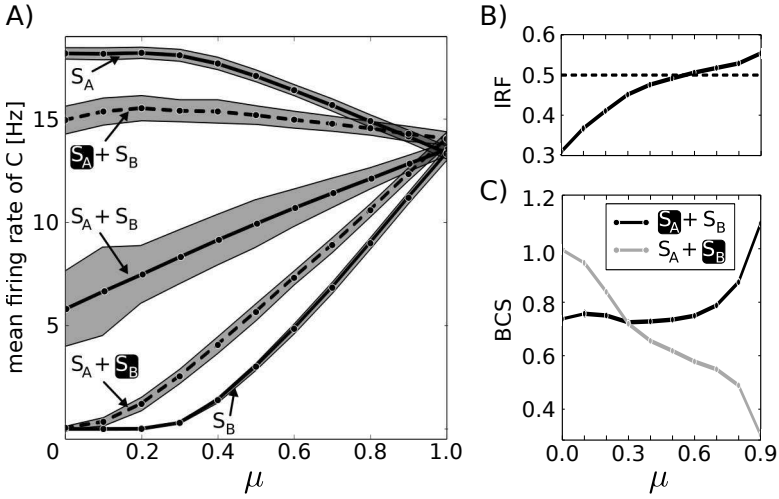

Figure 3.10: Biased competition effect for different values of the mixing parameter μ, which regulates the relative stimulus preference. **A)** In the second layer, mean firing rates show consistent normalization if both stimuli are present (S_A + S_B) and partial attentional recovery (dashed lines, black box marks attentional focus) to the rates the stimuli elicit when presented alone (S_A / S_B) over the full range of possible values for μ. If $\mu = 0$, S_B elicits no response in C, whereas if $\mu = 1$, responses elicited by S_A and S_B are equally strong. Shaded areas mark the 95% confidence interval. **B)** For values of $\mu \geq 0.3$, i.e. where the non-preferred stimulus elicits a non-zero response (see A)), the intermediate response factor (IRF) is ≈ 0.5, which means that the rate when both stimuli are presented lies approximately at the mean of the responses when only one stimulus is presented. **C)** In the same range, the biased competition score (BCS) in population C, quantifying the degree to which single stimulus responses are recovered under attention, is around 0.8 for preferred-rate recovery (black line) and around 0.6 for non-preferred-rate recovery (grey line), revealing a slight bias of the attentional rate recovery for the preferred stimulus. A value of 1 would correspond to perfect rate recovery, whereas 0 would mean that attention has no effect on the firing rates. Adapted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

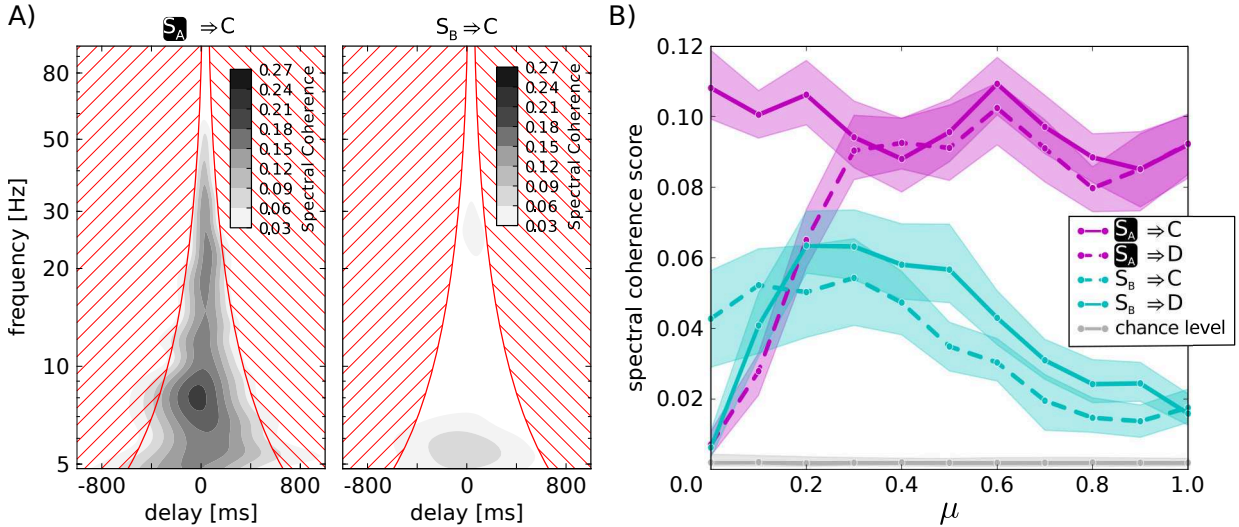

Figure 3.11: Information routing dependence on μ. **A)** Spectral coherence example for both stimuli present and S_A attended ($\mu = 1$). Transmission from the attended stimulus to C (left) is much higher than from the non-attended one (right). The spectral coherence score is calculated by averaging in the non-hatched region. **B)** The spectral coherence scores of all four possible transmission combinations from stimuli to second layer population are plotted against μ. Both stimuli are presented and S_A is attended. If $\mu = 0$, i.e. S_A (S_B) does not elicit a response in population D (C), there is no significant information transmission via these pathways. For $\mu \geq 0.3$, i.e. when the non-preferred stimulus induces a response significantly bigger than zero (see figure 3.10 A)), the spectral coherence score is consistently higher for information transfer from the attended stimulus to both second layer populations than from the non-attended stimulus. Shaded areas mark the 95% confidence interval. Adapted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

the whole parameter space, albeit at varying strengths, showing that the basic mechanisms are robust.

Furthermore, the susceptibility of information routing to neuronal noise in this framework is probed. The amount of noise in the system is controlled by introducing a noise factor that scales all connection probabilities down and all synaptic weights up accordingly. This keeps the mean input to every neuron constant, i.e. preserves important quantities like absolute rates, while increasing the noise in the system by making single spikes more influential. Figure 3.13 shows that the selectivity of information transmission degrades gradually with increasing noise factor. The main reason is that a certain degree of "smoothness" of the recurrent input is needed within the inhibitory subpopulations to generate a stable γ rhythm. As p_{loc}^{ii} decreases, this is no longer given. This point is illustrated in the insets of figure 3.13, that show example traces of the mean rate of A_{exc} and B_{exc} for a noise factor of 1 and 3.89. The coefficient of variation of inter-spike-intervals of single spike train contributing to these mean population rates is widely distributed in a range between approximately 0.35 and 1 as shown in figure 3.14, in line with studies on the variability of neuronal responses in cortex [SK93; BW76]. This holds also for conditions in which the noise factor is close to 1 and information transmission selectivity is optimal. The decrease of stability of the γ rhythm generated by the inhibitory subpopulations is reflected in the notable increase in the coefficient of variation of inhibitory neurons with the noise factor.

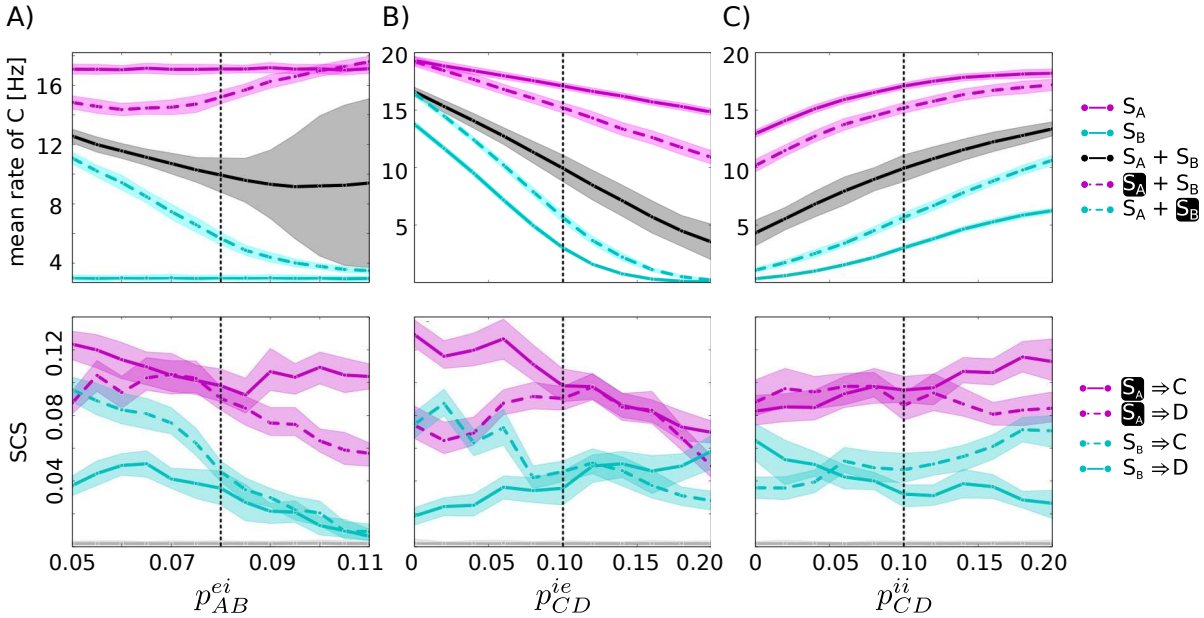

Figure 3.12: Robustness of effects to parameter variations. The first row shows plots analogous to figure 3.10 A) and the second row plots analogous to figure 3.11 B) as a function of p^{ei}_{AB} **(A))**, p^{ie}_{CD} **(B))**, p^{ii}_{CD} **(C))** for $\mu = 0.5$. The dashed black lines mark the default value of the respective parameters. Shaded areas mark the 95% confidence interval. Reprinted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

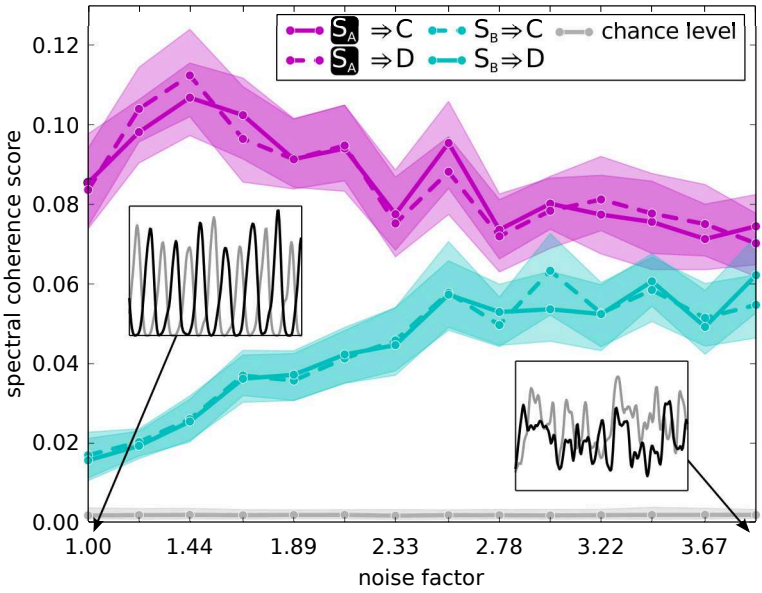

Figure 3.13: Robustness of selective information routing to noise. The noise in the system is increased by scaling all connection probabilities down and all synaptic weights up by the same noise factor. The plot shows the spectral coherence score for all four information transmission pathways for the case that both stimuli are present and S_A is attended. The mixing parameter μ is set to 1. Selectivity of information routing from the attended stimulus to the second layer degrades with increasing noise factor. The main reason is that as the recurrent connectivity within inhibitory subpopulations decreases, determined by p^{ii}_{loc}, the γ rhythms these populations generate get less and less stable. This fact is illustrated in the inset figures, that show example time courses of the mean activity of excitatory subpopulations in A (black) and B (grey) for noise factors of 1 and 3.89. Reprinted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

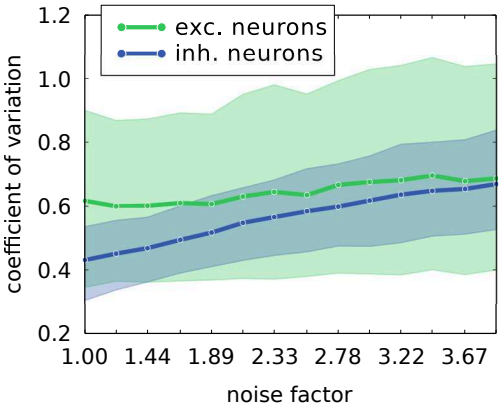

Figure 3.14: Spike-train variability in the model. The coefficient of variation of inter-spike-intervals of single neuron spike trains for all excitatory (green) and all inhibitory neurons (blue) in the network is displayed for various values of the noise factor. The solid lines denote mean values, the shaded areas 75% confidence intervals. The increase of the coefficient of variation of inhibitory neurons correlates with the decreasing stability of generated γ rhythms. Taken all neurons together, coefficients of variation range between ca. 0.35 and 1 over the whole range of the noise factor. Reprinted figure with permission from D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception". The Journal of Neurophysiology. 114.3, pp.1593-1605 (2015). Copyright 2015 by the American Physiological Society.

3.4 DISCUSSION

In summary, a realistic, two-layered neural network was designed, mimicking the converging feed-forward structure of the visual processing stream and incorporating intra-layer lateral connectivity. The mechanism for information routing relies on anti-phase synchronization in the first layer and in-phase synchronization in the second layer, thereby introducing a phase competition between the driving, first layer populations. Attention is conceived in the most simple way, namely as an additional selective input to the first layer.

routing mechanism

The idea of utilizing a phase competition to selectively route information is a direct implementation of the CTC hypothesis [Fri05; Fri09; Fri15]. The main contribution of this study is that both hallmark effects of visual attention – biased competition and information routing – can be successfully and robustly reproduced in a broad parameter regime, while single neurons show realistic response variabilities. As an explanation a twofold mechanism was found: firstly, by adding input to the first layer population processing the attended stimulus, its firing rate increases, leading to a stronger driving current targeting the second layer populations. At the same time, via the lateral connections, the increased activity in the attended population induces a decrease of activity in the unattended one, rendering its output to the second layer less effective. Secondly, this bias in the input of the higher area in favour of the attended population induces a phase relation between sending and receiving populations that facilitates information transfer from the attended population and attenuates transmission from the unattended one. This model will be referred to as the rate-imbalance induced selective routing (RISER) model in the following, due to the rate modulation in the first layer causing selective routing.

biased competition and information routing

In this study, biased competition and information routing were treated as two distinct effects. However, one could interpret biased competition as information routing where the time period for which a feature attribute is static is equal to the trial length. In the information routing experiments, the time period of static luminance was 10 ms. Considering this it is plausible that the mechanisms subserving information routing can be utilized to reproduce biased competition: essentially, the rate at which the attended first layer population fires is mirrored by the second layer populations, whereas the input by the unattended one is attenuated. Information routing is then the generic situation with biased competition as a boundary case.

comparison to related models

Other models put forward to explain information routing in biologically realistic networks of excitatory and inhibitory populations share certain commonalities: two distinct populations in one layer are connected to one population in a second layer, and competition between two stimuli is established ultimately by the

fact that both first layer populations drive the same (or partly the same) inhibitory subpopulation in the second layer, be it directly or indirectly [ZFG08; BEK08; WS12; DR05]. Of this, the RISER model proposed here is no exception. However, the lateral interactions are a novelty and allow the model to "self-organize" phase relations for stimulus selection and the use of a simple attentional direction signal, whereas in other approaches, either a desired phase relation was "manually" constructed [Mas09; TS10] or oscillations were externally imposed [ZFG08; WS12]. Building on the RISER model, [Pal+17] showed that a similar routing scheme also works in networks which do not oscillate strongly, but show transient bursts of γ activity.

3.4.1 *Physiological plausibility*

locus of competition

The model is structured as a self-consistent common motif in the visual pathway, where input from several cortical populations in one area impinges onto the same population in a higher area. Thus, going upstream from the first layer to lower visual areas will result in even smaller receptive fields that have less mutual interactions, thereby exhibiting a negligible attentional selection effect. Going downstream to higher areas, populations targeted by the input will even more act as one single rather than two different computational units, thus the attentional selection effect would stay constant or even enhance further. From this insight follows that depending on size and placing of stimuli, the actual stage at which competitive lateral interaction occurs may vary. In the information routing experiments [Gro+15; Bos+12], activity from V1 and V4 was recorded. Identifying the second layer with V4, the first model layer, following the above reasoning, does not necessarily have to correspond to V1, but could be a more downstream area like V2 or even a different cortical layer within V4.

gamma mechanisms

For explaining γ oscillations in visual cortex, two mechanisms are deemed plausible: interneuron network γ (ING) and pyramidal-interneuron network γ (PING). Whether ING or PING mechanisms underlie oscillations in the visual cortex is still debated and might depend on task requirements [TS09; IS11; BW12; GPS15], but a combination of both seems most likely, considering the local

connectivity structure [HM07]. The current study focused on a "pure" ING-mechanism (no connections between excitatory and from excitatory to inhibitory subpopulations). However, the specific implementation of the local and lateral connectivity is not of crucial importance. The proposed mechanism could be supported by any connectivity scheme that causes suppression and anti-phase synchronization in the first layer and in-phase synchronization in the second, while the local circuitry generates oscillatory activity.

frequency matching

While the exact γ mechanism seems to be irrelevant, a crucial constraint for effective routing to occur in this framework is that γ frequencies in all four populations are approximately equal. Otherwise, the phase relations could not be kept constant. The specific value of this γ frequency need not be fixed, but may vary depending on stimulus properties as has been shown experimentally [RM10]. Stimulus-dependent changes in γ frequency do not impede the routing mechanism as long as the frequencies of communicating populations co-vary. Dynamic frequency matching between V1 and V2 populations with overlapping receptive fields has indeed been demonstrated recently [Rob+13]. It can also be inferred that similar routing schemes could be used in different frequency bands, for example in the β-range, which has been proposed to be a channel for top-down information flow [BM07; BM09; Mic+16; Bas+15].

pulse-like attention signal

Another important model aspect is the attentional gating signal. Here a static additional input to the first layer is assumed. Apart from considering energy efficiency, a gating signal that is only used to "switch" the attentional state would also be more aesthetically pleasing. A recent study ventured in this direction [TS10], showing that pulses applied to an inhibitory subpopulation can shift the phase of the whole population, depending on the phase at which the pulse is applied. Simply substituting the additional input with such a pulsatile signal would likely not work in the RISER model, since the attentional selection causally relies on rate differences in the first layer. However, if the system is in the bistable regime, applying a short pulse to one population in the first layer could tip the relation of firing rates to its favour, setting a suitable phase for information transfer (figure

3.9 C)). In this state, the system would remain for a relatively long time without necessary intervention. The difference to the model in [TS10] is that the generator of the attentional signal does not have to "know" the phases of all involved populations to select from.

3.4.2 *Predictions of the RISER model*

attention vs. contrast

The model mechanisms for selective routing imply that attention can be traded against bottom-up stimulus saliency: if two stimuli of different saliency elicit different firing rates in the first layer, the weaker stimulus can not be gated unless its associated firing rate in the first layer is enhanced by attention to be higher than the rate elicited by the unattended one. Hence it may be possible to construct pairs of stimuli with strong enough luminance or contrast difference such that the weaker stimulus can not be attended any more. Whereas this has not been shown directly to my knowledge, the prediction is supported by several studies showing that attention acts similar to a contrast gain [MT02; RD03; CLR04]. In a setting similar to the one used for the information routing experiment [Gro+15], this prediction could be tested by using two flickering stimuli with the same flicker amplitude, but different contrasts.

distance depended competition

Since effective stimulus competition and routing in requires anti-phase synchronization between the two first layer populations, a further prediction is that the phase difference between these populations depends on the distance of the competing stimuli. This translates to the placing of stimuli in visual space in retinotopically organized areas if the competing feature is location in visual space (fully the case in the information routing experiment [Gro+15] and partially in biased competition experiments [MD85; RCD99]). More specifically, populations in close cortical vicinity should oscillate in phase, whereas at some "optimal" distance, the oscillations should be out of phase. However, note that distance is not necessarily difference of locations, but could be analogously defined for other receptive field dimensions such as orientation or colour. To my knowledge, the dependency of oscillation phases on cortical distance or tuning property in

visual cortex while two competing stimuli are presented has not been investigated to date.

bistability and binocular rivalry

In addition to the primary effects of biased competition and selective gating, it was found that by increasing lateral inhibition in the first layer, the system transitions into a bistable regime. Bistability has been found as an emergent model feature in [WS12], where it also depends on the strength of lateral inhibitory interactions, albeit in the second model layer. Multistability is an inherent feature of a subset of oscillating, delay coupled networks with lateral inhibition, that is not restricted to the specific realization in this model. A well studied example of bistable perception is binocular rivalry – the alternating percept that is observed when two different stimuli are presented monocularly to corresponding retinal locations (see [Blao1] for a comprehensive review). Current theories propose that the bistability of the percept is at least in part attributable to interocular inhibition between monocular neurons in V1 and/or LGN [TMB06], which concurs with the model architecture if distance in the ocularity dimension is considered (close (same eye) or far (other eye)).

By extending the distance dependent architecture to other stimuli dimensions, a prediction is that the locus at which rivalry is initiated depends on the features of the stimuli that are in competition, e.g. orientation differences will elicit bistable dynamics early in the visual system, whereas the neural correlate of shape ambiguities will be found in higher areas where neurons with respective receptive field properties are located. Consequently it can be expected that rivalry effects of competing stimuli that differ in several stimulus dimensions can be spread out through the whole visual system, consistent with the explanation for binocular rivalry promoted in [TMB06].

3.5 ACKNOWLEDGEMENTS

This chapter including figures was published in similar form as

> D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception," Journal of Neurophysiology **114.3**, 1593–1605 (2015). Copyright 2015 by the American Physiological Society.

The text underwent minor changes for consistency with other chapters and references and citations have been updated. I wrote the manuscript for submission, performed simulations and data analysis, and prepared the figures. The study was designed in collaboration with Klaus R. Pawelzik and Udo A. Ernst.

4

OPTIMAL NETWORK CONFIGURATION FOR SELECTIVE ROUTING

4.1 INTRODUCTION

the RISER model is a good routing network

A local network configuration thought to be critical for selective visual attention is the convergent fan-in structure, where several neuronal populations in one visual area provide input to the same population in another area. (e.g. [HEP15; Gro+15; Fri05; MD85]). As shown in Chapter 3 and [HEP15], the RISER model of such a network is sufficient to explain various electrophysiological findings of classical attention experiments. In the model, CTC is employed for information routing, where lateral interactions serve to automatically set suitable phase relations between oscillating populations.

what is the optimal routing configuration?

Building on these previous results, the purpose of this chapter is now answering the question: What is the optimal network configuration for efficient information routing under the constraint of a fan-in network motif? And specifically, are there network configurations that do not rely on oscillatory activity and phase shifts to achieve selective stimulus routing? In comparison to the network in chapter 3, several degrees of freedom were added, i.e. additional local and lateral excitatory interaction, feedback projections, and variable delays.

By global optimization of a selective routing objective it is found that indeed a network configuration which produces oscillations via strong local recurrent inhibition and an anti phase relation between the sending populations through lateral inhibition is optimal. The network structure of the RISER model is part of this optimal class of solutions. No qualitatively different network configurations were found. Also the addition of feedback projections and changing the delay between and within populations does not lead to fundamentally different routing behaviour.

4.2 METHODS

For efficient parameter exploration, the spiking network used in Chapter 3 was simplified to a rate based network formulation, where every neuronal subpopulation can be described by one single differential equation.

4.2.1 *Network dynamics*

The network dynamics are given by

$$\begin{aligned} \tau \dot{V}(t) &= -V(t) + Wr(t-d) + I(t) \\ r(t) &= h(V(t)) \quad . \end{aligned}$$

V represents an internal activation which is passed through the nonlinearity $h()$ to produce the output activation r. The nonlinearity $h()$ is chosen as a bounded piecewise linear function given by

$$h(x) = \Theta(x)\Theta(1-x)x \quad ,$$

with Θ denoting the Heaviside function. This produces $r \in [0,1]$ without loss of generality. The membrane time constant τ is chosen as 5 ms, as well as the delay time d. W is the coupling matrix and I external input to the network.

To asses the selective routing of the network, he differential equation was numerically integrated via Euler's method with a time step of 0.2 ms for 50 seconds.

4.2.2 *Network structure*

Three populations, each with an excitatory and an inhibitory subpopulation, are arranged in 2 layers where population A and B in layer 1 provide feed-forward input to population C in layer 2, creating a fan-in network. The difference to the 4 population setup in chapter 3 is addressed in the discussion. The detailed network structure is determined by the coupling matrix W. Non-zero values of the matrix are the free parameters of the model. These are the coupling weight from the excitatory subpopulation onto itself

(w^{ee}_{loc}), from the excitatory to the inhibitory subpopulation (w^{ei}_{loc}), the inhibitory to excitatory subpopulation (w^{ie}_{loc}) and from the inhibitory subpopulation onto itself (w^{ii}_{loc}) in every population A,B, and C. Furthermore, the lateral interaction from excitatory to excitatory subpopulation between populations in the first layer (w^{ee}_{AB}) and from excitatory to inhibitory subpopulation (w^{ei}_{AB}). The feed forward weight w_{ff} from the excitatory subpopulations of A and B to excitatory and inhibitory subpopulation of C was fixed to $w_{ff} = 0.5$. The dynamical equation for V can then be explicitly written as

$$\tau \frac{d}{dt} \begin{pmatrix} V^e_A \\ V^i_A \\ V^e_B \\ V^i_B \\ V^e_C \\ V^i_C \end{pmatrix} = - \begin{pmatrix} V^e_A \\ V^i_A \\ V^e_B \\ V^i_B \\ V^e_C \\ V^i_C \end{pmatrix} + W \begin{pmatrix} r^e_A \\ r^i_A \\ r^e_B \\ r^i_B \\ r^e_C \\ r^i_C \end{pmatrix} + \begin{pmatrix} I^e_A \\ I^i_A \\ I^e_B \\ I^i_B \\ I^e_C \\ I^i_C \end{pmatrix}$$

with

$$W = \begin{pmatrix} w^{ee}_{loc} & w^{ie}_{loc} & w^{ee}_{AB} & 0 & w_{fb} & 0 \\ w^{ei}_{loc} & w^{ii}_{loc} & w^{ei}_{AB} & 0 & w_{fb} & 0 \\ w^{ee}_{AB} & 0 & w^{ee}_{loc} & w^{ie}_{loc} & w_{fb} & 0 \\ w^{ei}_{AB} & 0 & w^{ei}_{loc} & w^{ii}_{loc} & w_{fb} & 0 \\ w_{ff} & 0 & w_{ff} & 0 & w^{ee}_{loc} & w^{ie}_{loc} \\ w_{ff} & 0 & w_{ff} & 0 & w^{ei}_{loc} & w^{ii}_{loc} \end{pmatrix},$$

where the subscript on V, r, and I signifies the population membership and the superscript whether the respective population is excitatory (e) or inhibitory (i).

4.2.3 *Input*

The external input I consists of a static and a dynamic part. The static part was set to 0.5 to excitatory and inhibitory subpopulations in A and B, and to 0.25 to both the excitatory and the

inhibitory subpopulation of C. In addition to this, populations A and B receive uncorrelated noise signals, drawn every 10 ms from a Gaussian white noise process with $\mu = 0$ and $\sigma = 0.1$, symbolized by $S_A(t)$ and $S_B(t)$. Without loss of generality, the dynamic noise input to A was chosen as the attended stimulus. Hence, both subpopulations in A received an extra static input Δ^+_{att} modelling attentional drive. The input I can then be explicitly written as

$$\begin{pmatrix} I^e_A \\ I^i_A \\ I^e_B \\ I^i_B \\ I^e_C \\ I^i_C \end{pmatrix} = \begin{pmatrix} 0.5 + \Delta^+_{att} + S_A(t) \\ 0.5 + \Delta^+_{att} + S_A(t) \\ 0.5 + S_B(t) \\ 0.5 + S_B(t) \\ 0.25 \\ 0.25 \end{pmatrix}$$

The full network configuration used for the global parameter optimization is summarized in figure 4.1.

4.2.4 *Global optimization*

The network described in the previous section shall be optimized for selective information transmission from $S_A(t)$ to the activity $r^e_C(t)$ of the excitatory subpopulation of C, which is quantified by the objective function O:

$$\begin{aligned} O = &[\chi(f(S_A), f(r^e_C)) - \chi(f(S_B), f(r^e_C))] \\ &\cdot \sigma(\chi(f(S_A), f(r^e_A))) \cdot \sigma(\chi(f(S_B), f(r^e_B))) \quad . \end{aligned}$$

Here, $f()$ denotes a band pass filtering between 5 and 30 Hz using two-way least-squares FIR filtering. $\chi()$ performs a cross correlation on the z-scored inputs and returns the maximum of this correlation in a causal shift window of 0 to 20 ms. The function $\sigma()$ is a sigmoid given as

$$\sigma(x) = (1 + \exp(-30(x - 0.7)))^{-1} \quad .$$

The first factor of O measures the difference between information content in r^e_C originating from S_A and S_B and can have values

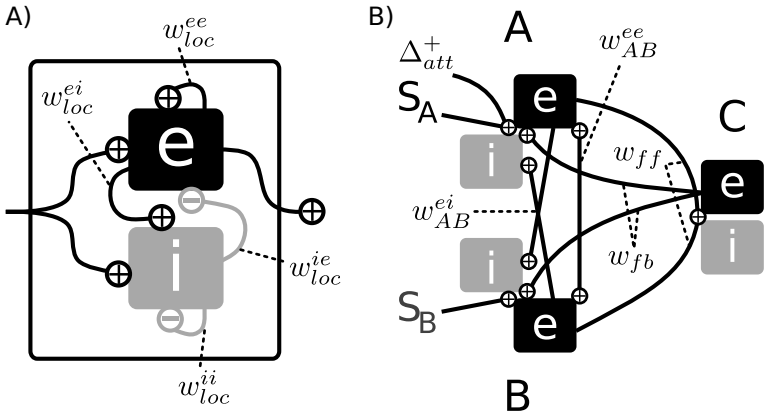

Figure 4.1: Network setup for routing optimization. **A)** Each local population consists of an excitatory and an inhibitory subpopulation. Between these subpopulations, all possible connections are allowed (compare figure 3.2 A) in contrast). **B)** Three populations are arranged in two layers to form a fan-in network structure, whereas feed forward and feedback projections exist. Within layer 1, lateral connections from excitation to inhibition and between both excitatory subpopulations exist (compare figure 3.3 A) in contrast)

between -1 and 1. 1 signifies perfect selective routing of the attended stimulus S_A, and -1 perfect routing of the unattended stimulus S_B. The last two factors quantify the information content of the stimuli in the associated first layer populations and can range between 0 and 1. These factors modulate the objective function such that networks where stimulus information is already lost in the first layer are penalized. Thus, the highest value of O to be expected for perfect selective information routing while retaining information about both stimuli in the first layer is 1.

The objective function was optimized with respect to the parameters $X = \{w^{ee}_{loc}, w^{ei}_{loc}, w^{ie}_{loc}, w^{ii}_{loc}, w^{ee}_{AB}, w^{ei}_{AB}, w_{fb}, \Delta^{+}_{att}\}$. The allowed range of these parameters was restricted by the boundary values given in table 4.1. In comparison to the network setup in chapter 3, the parameters w^{ee}_{loc} and w^{ei}_{loc} were added to allow for PING solutions, and w_{fb} to investigate the effect of feedback

projections that are known to exist. The specific choices for the boundaries are motivated in the discussion.

parameter	w_{loc}^{ee}	w_{loc}^{ei}	w_{loc}^{ie}	w_{loc}^{ii}	w_{AB}^{ee}	w_{AB}^{ei}	w_{fb}	Δ_{att}^{+}
max value	0.5	2	0	0	2	2	0.5	0.3
min value	0	0	-3	-3	0	0	0	0

Table 4.1: Parameter boundaries for global optimization.

Global optimization of $O(X)$ w.r.t X was performed using Bayesian optimization with Gaussian processes [WR06; Sha+16]. This procedure is tailored to maximize (or minimize) a noisy black box objective function by using only few function evaluations. Simply put, the method tries to exploit an assumed smoothness of the mapping from X to O to find the optimum quickly. Since this method is fairly new, the following paragraphs will provide a quick overview.

Assume that at optimization step k, the noisy scalar measurements $\{y_1, y_2, ...y_{k-1}\}$ of O at parameter values $\{X_1, X_2, ...X_{k-1}\}$ are known, where

$$y_i = y(X_i) = O(X_i) + \eta \quad ,$$

with η being a white noise process. Given this data, an estimate $\tilde{O}(X)$ for the whole D-dimensional parameter space X is constructed by a Gaussian process. A Gaussian process defines a prior over all possible functions that can give rise to the observations $\{y_1, y_2, ..., y_{k-1}\}$. It assumes that $p(O(X_1), ..., O(X_{k-1}))$ is jointly Gaussian with a mean $\mu(X)$, a standard deviation σ_f of O, an observation noise standard deviation σ_n, and a covariance matrix Σ with entries $\Sigma_{ij} = \kappa(|X_i - X_j|, \lambda_1, ..., \lambda_D)$, where κ is a smooth positive kernel function and λ_i a length scale in input dimension i. From this, a posterior prediction $\tilde{O}(X)$ can be generated for every point in X. The simple intuition behind this is that close points in X map to similar values $O(X)$, where the degree of similarity depends on the length scales λ_i. This prediction $\tilde{O}(X)$ depends on the hyperparameters $h = \lambda_1, ..., \lambda_D, \sigma_f, \sigma_n, \mu$, which can be optimized given the observations, since $p(h|X, y)$ is analytically accessible. For a Gaussian process, the prediction $\tilde{O}(X)$

naturally comes with the standard deviation $\sigma_{\tilde{O}}(X)$ which quantifies the degree of uncertainty of the prediction at X. Generally speaking, the uncertainty grows with increasing distance from the previously sampled points $\{X_1, X_2, ...X_{k-1}\}$.

Once $\tilde{O}(X)$ and $\sigma_{\tilde{O}}(X)$ are obtained, an acquisition function $AF(\tilde{O}(X), \sigma_{\tilde{O}}(X))$ is maximized that evaluates at which point the objective should be sampled next to find the maximum. Several acquisition functions are typically used that vary in their trade off between exploitation (focus on high $\tilde{O}(X)$) and exploration (focus on high $\sigma_{\tilde{O}}(X)$). The next point X_k is thus given as $X_k = \text{argmax}_X(AF(X))$. At this point, the objective is sampled, i.e. $y_k = y(X_k)$ is obtained, and the next optimization iteration begins.

In essence, this optimization procedure assumes smoothness of the objective, easily handles observation noise, and transforms the problem of optimizing $y(X)$ directly to optimizing $p(h|X, y)$ and $AF(X)$. Thus it makes sense if $p(h|X, y)$ and the acquisition function $AF(X)$ are computationally cheaper to evaluate than $y(X)$. Figure 4.2 shows a cartoon of the workflow of one optimization step. An excellent introduction to the topic is given in [WR06].

For the present application, a Matérn 5 kernel function κ was chosen, which in input dimension d reads

$$\kappa(|x_i^d - x_j^d|) = \sigma_f^2 \left(1 + \frac{\sqrt{5}|x_i^d - x_j^d|}{\lambda_d} + \frac{5|x_i^d - x_j^d|^2}{3\lambda_d^2}\right) \exp\left(\frac{\sqrt{5}|x_i^d - x_j^d|}{\lambda_d}\right) \quad .$$

Here, $||$ denotes the L2-norm.

As acquisition function, the expected improvement was used, which, with Φ as the standard normal cumulative density function and ϕ as the standard normal probability density function, can be written as

$$AF(X) = (\tilde{O}(X) - y_{max})\Phi\left(\frac{\tilde{O}(X) - y_{max}}{\sigma_{\tilde{O}}(X)}\right) + \sigma_{\tilde{O}}(X)\phi\left(\frac{\tilde{O}(X) - y_{max}}{\sigma_{\tilde{O}}(X)}\right) \quad .$$

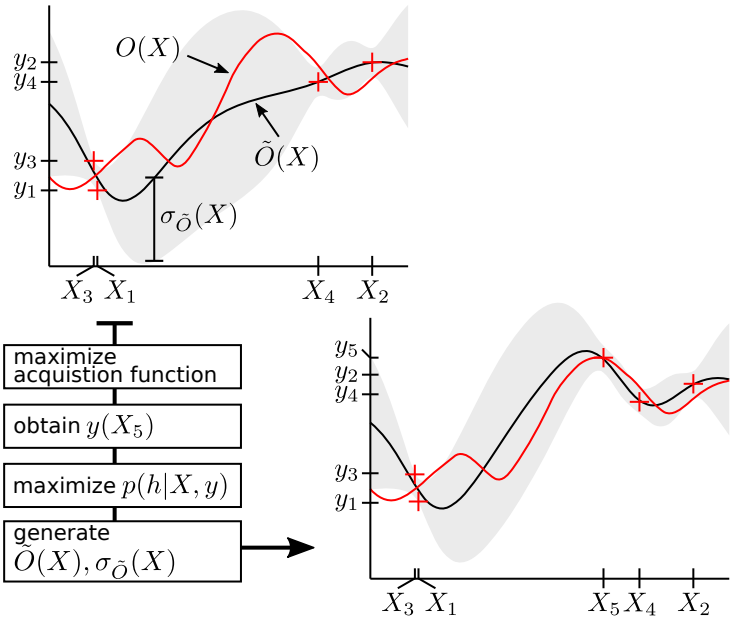

Figure 4.2: Illustration of the optimization procedure. The upper plot shows the prediction $\tilde{O}(X)$ (black line) given the observations $\{y_1, y_2, y_3, y_4\}$ (red crosses) at $\{X_1, X_2, X_3, X_4\}$ in a one dimensional example, along with the true unknown underlying objective $O(X)$ (red line). Note that the measurements y do not coincide with O due to observation noise. The uncertainty $\sigma_{\tilde{O}}(X)$ of the prediction is symbolized by the grey shaded area. The length scale λ controls how "smooth" the estimated objective $\tilde{O}$ is assumed to be, i.e. how quickly the uncertainty increases with increasing distance to the measured points $\{X_1, X_2, X_3, X_4\}$. From the prediction, the next point in X to check is determined by maximizing an acquisition function. A exploitative function would probably pick a point right of X_4, whereas an explorative one would peak in between X_1 and X_4. For this example, the latter is used. The available data is then augmented by X_5 and $y_5 = y(X_5)$, and $p(h|\{X_1, X_2, X_3, X_4, X_5\}, \{y_1, y_2, y_3, y_4, y_5\})$ is maximized to refine the prediction $\tilde{O}(X)$, which is shown in the lower right plot. This procedure is then repeated until some stopping criterion is reached.

This function quantifies the expectation that the value O at X is larger that y_{max}. The target y_{max} at iteration k was set as

$$y_{max} = \begin{cases} \max(\{y_1, ..., y_{k-1}\}) & \text{if } \max(\{y_1, ..., y_{k-1}\}) \geq 0.5 \\ 0.5 & \text{else} \end{cases},$$

incorporating prior knowledge that values of $y \approx 0.5$ were observed upon cursory inspection before optimization. The acquisition function was numerically maximized by global pattern search followed by numerical gradient ascent. In addition, at each optimization step k with probability 0.15, X_k was chosen randomly instead of by maximizing the acquisition function to prevent possible overexploitation.

Furthermore, smooth box priors on the hyperparameters were implemented, given by

$$bp(x, a, b, c) = \frac{1}{b-a}\xi(c(x-a))(1-\xi(c(x-b)))$$
$$\xi(z) = \frac{1}{1+\exp(-z)} \quad .$$

This prior is quasi constant between the lower boundary a and the higher boundary b, where c controls the drop off outside these boundaries. For μ, the parameters were $\{a = -0.5, b = 0.5, c = 15\}$, for the length scales λ_d $\{a = 0.001, b = 10, c = 15\}$, for σ_f $\{a = 0.001, b = 2, c = 15\}$ and for σ_n $\{a = 0.001, b = 0.25, c = 15\}$.

The function $p(h|X, y)$, i.e the probability for the hyperparameters given the observed data, was maximized at each optimization step after new data was acquired with standard gradient ascent algorithms using 3 different starting points. One starting point was the previous maximum of $p(h|X, y)$, and the other two were randomly chosen.

Each optimization run s was initialized with 5 randomly chosen starting values $\{X_1, ..., X_5\}$ and corresponding $\{y_1, ..., y_5\}$ and lasted $k_{max} = 300$ optimization steps. In total, $s_{max} = 200$ parallel optimization runs were performed with different random initial values $\{X_1, ..., X_5\}$. After $k_{max} = 300$ optimization steps in run s, the final prediction $\tilde{O}^s(X)$ of that run was maximized,

i.e. $X^s_{max} = \text{argmax}_X(O^s)$ was determined, by pattern search followed by numerical gradient ascent. Thus, this leads to 200 potentially different predictions of the objective with potentially different optimal routing network configurations X^s_{max}. Good routing solutions were considered to be networks with parameters X^s_{max} for which $\tilde{O}^s(X^s_{max}) \geq 0.55$.

The implementation of the full algorithm was built around the GPML Matlab toolbox [RN10].

4.2.5 *Phase analysis*

Phase differences between oscillating population were analysed analogously to chapter 3.2.6. Thus, the phase difference between signals $x(t)$ and $y(t)$ is obtained via the analytic signals $x^a(t)$ and $y^a(t)$ as

$$\Delta\phi = \langle \arg\left(\frac{x^a(t)}{y^a(t)}\right)\rangle_t \quad ,$$

The phase difference between the activity of populations A_{exc} and B_{exc} for a single simulation run was calculated directly from the activities as

$$\Delta\phi_{AB} = \langle \arg\left(\frac{r^{e,a}_A(t)}{r^{e,a}_B(t)}\right)\rangle_t \quad .$$

For the phase differences between A_{exc} and C_{exc}, the input to C_{exc} originating from A_{exc} was compared to the full recurrent input to C_{exc}, i.e.

$$\begin{aligned} J_A(t) =& w_{ff} r^e_A(t-d) \\ J_C(t) =& w_{ff} r^e_A(t-d) + w_{ff} r^e_B(t-d) + \\ & w^{ee}_{loc} r^e_C(t-d) + w^{ie}_{loc} r^i_C(t-d) \\ \Delta\phi_{AC} =& \langle \arg\left(\frac{J^a_A(t)}{J^a_C(t)}\right)\rangle_t \quad . \end{aligned}$$

Accordingly, $\Delta\phi_{BC}$ was determined.

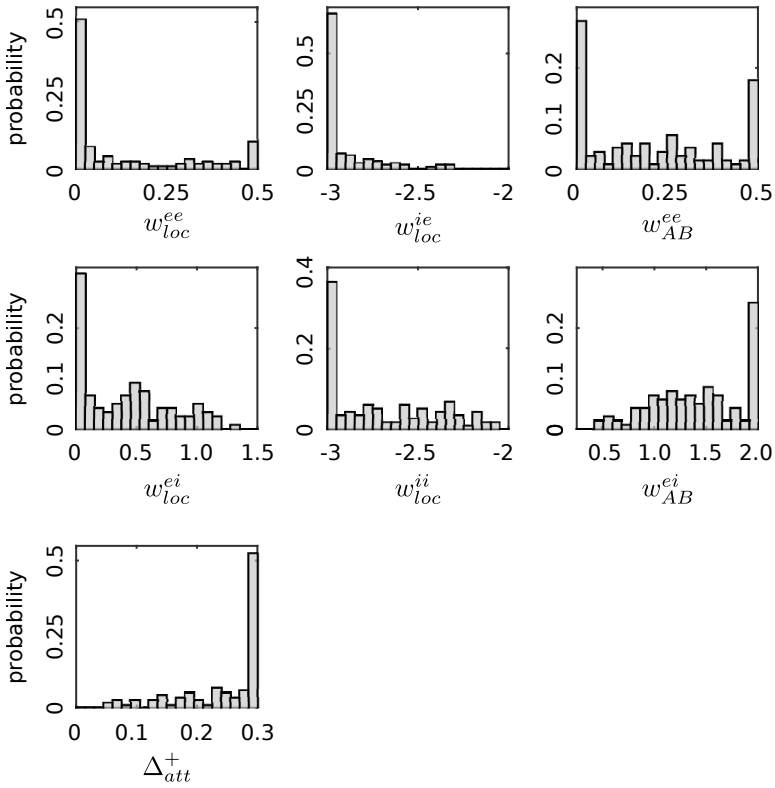

Figure 4.3: Probability distributions of the optimal parameter values for runs s with $\tilde{O}^s(X^s_{max}) \geq 0.55$. Note that the abscissae do not always span the full allowed parameter range (compare table 4.1).

4.3 RESULTS

In the default optimization setting, the feedback weight w_{fb} was set to 0. Global optimization was performed in 200 parallel runs with randomized initial conditions, with 300 objective function evaluations each as detailed in the methods section. 141 out of 200 runs were successful in finding a $\tilde{O}^s(X^s_{max}) \geq 0.55$. In the following, only these successful runs are considered. Figure 4.3 shows the distribution of the parameters $\{w^{ee}_{loc}, w^{ei}_{loc}, w^{ie}_{loc}, w^{ii}_{loc},$

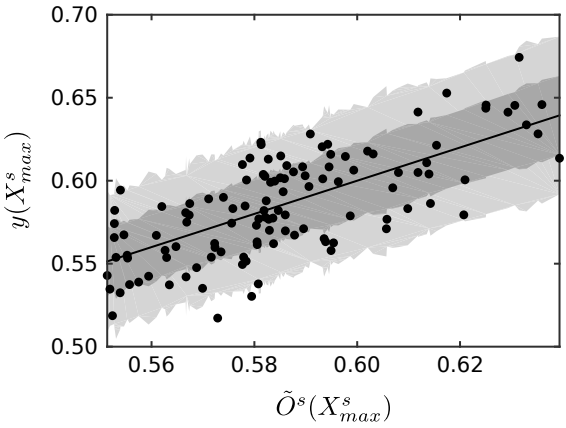

Figure 4.4: Validity of predictions for the routing objective. The predicted value $\tilde{O}^s(X^s_{max})$ for every good router is plotted against $y(X^s_{max})$, determined from one network simulation run. The dark (bright) grey area corresponds to $1\sigma_{\tilde{O}}$ ($2\sigma_{\tilde{O}}$). All predicted values are in good agreement with the measured objective.

w^{ee}_{AB}, w^{ei}_{AB}, $\Delta^{+}_{att}\}$ of these good routers. It was found that the parameters Δ^{+}_{att}, w^{ii}_{loc}, and w^{ie}_{loc} mostly assume values at or close to a parameter boundary, whereas the others have a further spread.

To verify the values of the predicted $\tilde{O}^s(X^s_{max})$, the networks were simulated at X^s_{max} and y was measured. Figure 4.4 shows that the values $y(X^s_{max})$ are in good agreement with the predictions by the Gaussian process regression.

Figure 4.5 displays the average power spectra of r^e_C and the 95% confidence interval of the good routers, revealing a strong oscillatory component in the γ frequency band in every network. Furthermore, a stable anti-phase relation between A_{exc} and B_{exc} is found and a higher mean activity of A_{exc} than B_{exc}. This leads to the peaks of the full input to C_{exc} to slightly precede the peaks of the incoming current from A_{exc} by 0.090π, and to lag behind the peaks of incoming current from B_{exc} by 0.881π. Hence, the peaks of attended incoming input to the peaks of the full incoming input are nearly aligned. The values for the phase differences are given in table 4.2.

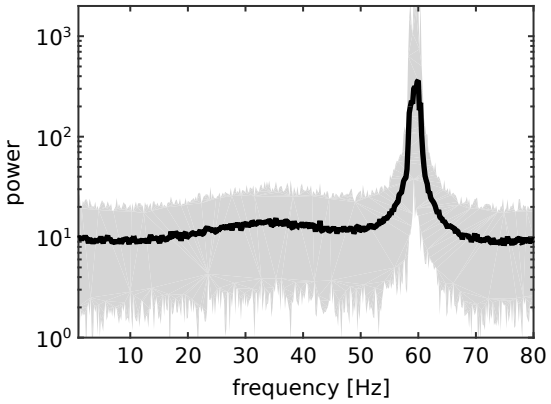

Figure 4.5: Power spectra for single runs of each good router, simulated for 50 seconds. The borders of the shaded area correspond to the 2.5 and 97.5 percentiles, and the black line the mean over good routers. In every instance, a strong oscillatory component in the γ range around 60 Hz was observed.

The fact that some parameters tend to concentrate close to a boundary value in all successful optimization runs suggests that these are most important in setting up the network for selective routing. This was tested by running 500 networks with random choices for w^{ee}_{loc} and w^{ee}_{AB} within the boundaries given in table 4.1, w^{ei}_{loc} randomly chosen in the interval $[0, 1]$, and w^{ei}_{AB} in $[1, 2]$. The other parameters were fixed at $w^{ii}_{loc} = -3$, $w^{ie}_{loc} = -3$, $\Delta^{+}_{att} = 0.3$. In this subspace of the full parameter space, the probability of finding a good routing network is approximately 50%, as illustrated in figure 4.6.

For the results so far, the feedback weight w_{fb} was set to 0. The same optimization procedure was performed while allowing w_{fb} to assume values in the range $[0, w_{ff} = 0.5]$. 151 out of 200 optimization attempts were successful in producing an $\tilde{O}^s(X^s_{max}) \geq 0.55$. The parameter distributions of good routers with feedback were qualitatively similar to the ones without feedback shown in figure 4.3, whereas w_{fb} was spread throughout the allowed range. This indicates that w_{fb} is not a crucial parameter for routing. Also, all good routing networks were found to

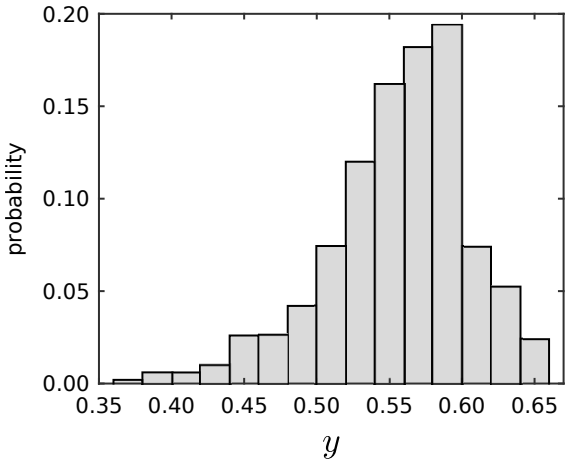

Figure 4.6: Variation of non-crucial parameters does not greatly hinder routing. The probability of observing y is shown when taking 500 random samples from the four dimensional subspace spanned by $w^{ee}_{loc} \in [0, 0.5]$, $w^{ee}_{AB} \in [0, 0.5]$, $w^{ei}_{loc} \in [0, 1]$ and $w^{ei}_{AB} \in [1, 2]$. The other parameters are fixed to $w^{ii}_{loc} = -3$, $w^{ie}_{loc} = -3$, and $\Delta^{+}_{att} = 0.3$. The median lies at 0.558.

express strong oscillations in the gamma range, have a higher activity in A_{exc} than B_{exc}, show anti-phase locking between r^e_A and r^e_B, and similar phase differences between the three populations (see table 4.2).

Furthermore variations of the delay d were considered. Here, the parameter d could have discrete integer values in the range [1 ms, 10 ms]. 110 out of 200 optimization attempts were successful. Also in this case, the distribution of the parameters shown in figure 4.3 was qualitatively reproduced, while the delay was clustered around 5 ms (median: 5 ms, minimal value: 2 ms, maximal value 8 ms). All networks again showed strong oscillations in the γ range, with variations in the peak frequency between ≈ 40 Hz and ≈ 100 Hz, caused by the different delays. Observations of firing rate differences and phase relations were in accordance with the previous results (see table 4.2).

params	default	default and w_{fb}	default and d
$\langle \Delta\phi_{AB} \rangle$ $[\pi]$	0.974 ± 0.044	0.954 ± 0.019	0.968 ± 0.054
$\langle \Delta\phi_{AC} \rangle$ $[\pi]$	0.090 ± 0.026	0.074 ± 0.032	0.089 ± 0.033
$\langle \Delta\phi_{BC} \rangle$ $[\pi]$	-0.881 ± 0.061	-0.879 ± 0.028	-0.884 ± 0.149

Table 4.2: Phase differences of good routers. The values show the mean phase differences $\pm$ one standard deviation between populations averaged over good routers for 3 different optimization settings. Left column: optimization over the default parameters $\{w^{ee}_{loc}, w^{ei}_{loc}, w^{ie}_{loc}, w^{ii}_{loc}, w^{ee}_{AB}, w^{ei}_{AB}, \Delta^{+}_{att}\}$. Middle column: optimization over the default parameters and w_{fb}. Right column: optimization over the default parameters and the delay d.

4.4 DISCUSSION

the RISER model is an optimal selective router

In summary , it was shown that oscillating feed forward networks with lateral interaction in the first layer are optimal for selective information routing. Both pure ING ($w^{ei}_{loc} = 0$) and mixed ING/PING solutions ($w^{ei}_{loc} > 0, w^{ii}_{loc} < 0$) can exhibit similarly high routing selectivity. Pure PING networks ($w^{ei}_{loc} > 0, w^{ii}_{loc} = 0, w^{ie}_{loc} < 0$) were not found. This is probably due to the fact that strong recurrent inhibition is more potent in generating a stable γ rhythm. However, the lack of any pure PING networks among the optimal routers does not mean that these are necessarily "bad" routers, just suboptimal in this setup. The RISER model of chapter 3 is contained in the set of optimal solutions. Networks that employ a routing mechanism not relying on oscillatory activity were not found.

oscillations are crucial

In this simple model, it is theoretically understandable why no other mechanisms exist: pure shifts in the mean rates of the first layer populations, which occur in non-oscillatory regimes, have no effect on selective information transmission, since the chosen objective function is insensitive to the mean. Thus, non-oscillatory solutions would have to make use of the nonlinearity of $h()$ around 0 by silencing the non-attended first layer population. This trivial solution is forbidden, though, by the formulation of the objective function. Under this constraint, selective routing

can only be brought about by oscillations which can modulate the effective influence. Input arriving during a period where $V_C^e < 0$ will fail to elicit a response in r_C^e, and thus have a lower effective influence on B_{exc} than input which arrives during periods where $V_C^e \geq 0$. Also, a bare minimum routing setup, as given by a pure ING network, with no extra recurrent local and lateral excitation is conceivably most effective, since these extra connections would add delayed reverberations of S_A and S_B to the network, which can amount to more "noise" on the representation of the signals in the first layer.

routing networks are robust

Consequently, the most important parameters for information routing are the local recurrent inhibitory weights (w_{loc}^{ii},w_{loc}^{ie}) which generate oscillations, lateral coupling in layer 1 from excitation to inhibition (w_{AB}^{ei}) which sets a suitable phase relation, and the additive attentional rate bias Δ_{att}^{+}. Variations in the other parameters when choosing optimal values for the former do not lead to a dramatic decrease in the routing performance (figure 4.6), showing that a routing network configuration is fairly robust. This was already observed in the spiking network formulation in chapter 3. It is plausible that moderate variations of the parameters not directly involved in the routing mechanism do not hamper performance significantly: Variations in the overall excitation mostly shifts average activations, against which the objective function is insensitive.

4.4.1 *Relation to the RISER model*

In comparison to the model used in the previous chapter 3, the network was simplified to a rate-based formulation. It was confirmed beforehand that both information routing and biased competition can be observed in this reduced description. The network structure was then further simplified by modelling only one second layer population instead of two. This is motivated by the fact that if the feed forward weights from A and B to C are equally strong, as assumed here, both second layer populations essentially behave as one and can thus be conflated. This comes at the expense of biased competition no longer being observable, but as it was argued before, biased competition can be seen as a special

case of information routing. Thus the focus in this study lies on information routing and equality of feed forward weights was chosen to reduce the parameter space.

4.4.2 *Parameter ranges*

The allowed parameter ranges in which networks can be realized (see table 4.2) were chosen for specific reasons. By simulating networks outside this range, it was determined that:

1. Recurrent excitation ($w^{ee}_{loc}, w^{ee}_{AB}, w^{ee}_{fb}$) far above the range can lead to runaway excitation, such that the rate dynamics are stuck at $\max(r) = 1$. This leads to the objective function being 0.

2. Stronger than allowed recurrent inhibition ($w^{ii}_{loc}, w^{ie}_{AB}$) does not change the oscillatory pattern significantly. In addition, sustained oscillations occur at $w^{ii}_{loc} \approx -1.5$ such that the chosen range ensures approximately equal parameter space sizes for oscillatory and non-oscillatory network configurations.

3. If Δ^{+}_{att} becomes too large, the chance of driving r^{e}_{A} to 1 is high, entailing low values of the objective.

4. If both Δ^{+}_{att} and w^{ei}_{AB} are above the chosen range, the chance of silencing the non-attended population grows, which also implies an objective close to 0.

In summary, the parameter ranges were set such that oscillatory and non-oscillatory networks are contained to a similar degree, while making sure that the objective function is different from 0 in a significant hypervolume of X. Were the latter not the case, any optimizer (including a human observer) would struggle to find an optimal solution. However, it is not guaranteed that outside of the boundaries no different routing solutions exist since the parameter space could only be spot-checked due to the computational load. But given the above reasoning, it is unlikely.

4.4.3 *Flexibility*

suppression of non-attended information incurs less flexibility

The objective function contains terms which penalize selective routing setups where the information about the non–attended stimulus is lost in the first layer. These terms were added for several reasons.

Most importantly, losing information about the non-attended stimulus corresponds to strong suppression of activity of B_{exc}. Complete suppression of neuronal activity under attention has not been observed experimentally, so this criterion ensures biological plausibility.

Why this strong suppression does not occur can be motivated from the model dynamics: In additional simulations it was found that if B_{exc} is strongly suppressed, the time after which information from S_B is represented in r^e_C after an attentional switch from stimulus S_A to S_B is longer than when population B_{exc} is active during attention to S_A. Thus it hinders flexible switching of attention.

Secondly, it has to be borne in mind that although the three population network is studied in isolation here for simplicity, in reality this motif is embedded in a much bigger network. If information about the non-attended stimulus is now completely lost through the interactions in layer 1, it is also not accessible for other possibly parallel computations for which the information might necessary. So being completely "blind" to non-attended information seems suboptimal from a computational point of view.

4.4.4 *Certainty of optimality*

Bayesian optimization with Gaussian processes is a relatively new optimization technique, and the claim of minimization of necessary function evaluations to reach an optimum is not yet analytically supported. Also, many variations exist. The algorithm used here was chosen since average convergence for a similar type has been recently proven [Sri+09]. Additionally, it was shown empirically that Bayesian optimization outperforms other state of the art black box function optimizers if only $10D$ to $100D$ function evaluations are allowed [HHL13], with D being the dimension

of the parameter space. Thus the confidence that the results indeed show solutions close to the global maximum of the objective within the allowed parameter range is high.

4.4.5 *Dependence on modelling choices*

simple spiking networks have not shown non-oscillatory routing

The activities of the networks investigated in this chapter were described by population average firing rates to reduce computation time. Where the rejection of selective routing schemes without oscillations for networks in this formulation seems sound, it is still possible that in the simplification step from a spiking description including many neurons per population, necessary properties are lost which would allow for different routing mechanisms. Several studies have investigated stimulus propagation and computation in multi-layer spiking networks in the absence of oscillations, highlighting the theoretical benefits of this network mode [SN98; RTN02]. Especially in [VA05] the neuronal realization of several logical gates and switches in non-synchronized recurrent spiking networks were explored. However, switches that satisfy the requirement of non-vanishing information content of both signals in the first layer were not described.

routing by transient synchronization is suboptimal

A routing scheme that does not rely on ongoing oscillations as the solutions found here, but uses spiking neuronal networks that are "sitting at the edge of synchrony" and show transient, strong γ fluctuations was described in [Pal+17]. Selective routing from two populations which converge onto the same, third population also does not require lateral interaction in the first layer. Similar transiently oscillating network states are possible in the rate-based formulation used in the present study, but were not picked out by the optimization algorithm since they have an objective far below 0.55. Preliminary analysis of the routing selectivity of such transiently oscillating spiking networks without lateral interaction as captured by the spectral coherence – the measure used in the original information routing experiment [Gro+15] and chapter 3 – also revealed that the routing selectivity is worse than for the RISER model from chapter 3 (Udo A. Ernst, personal communication). The authors of [Pal+17] used a different measure to quantify routing not directly comparable to spectral coherence.

non-linear dendrites can support routing without oscillation

If further model assumptions are relaxed, indeed routing networks can be found that neither rely on ongoing nor transient oscillations. In [BSE14], selective routing between two successive layers was demonstrated, reproducing a range of attentional effects on receptive field properties of visual neurons. However, necessary assumptions in this model are the existence of three distinct subpopulations in each population: A stimulus driven subpopulation, a control subpopulation involved in the local computation of complex attentional signals, and a gating subpopulation which features non-linear dendrites. The latter essentially augments the network by a third in-between layer where selective gating can occur. Thus it seems that while non-oscillatory routing can be achieved, it comes at the cost of highly increased complexity. This might seem unnecessary, given that neuronal populations readily express oscillatory behaviour.

5 ATTENTIONAL INFLUENCE IN SELECTIVE ROUTING

5.1 INTRODUCTION

In the previous chapters 3 and 4, network structures supporting selective information routing were exhaustively investigated, where two stimuli are separately processed in a first model layer, but converge onto the same population in the second layer. The focus in this chapter is now on the influence attention directly exerts on the populations in the network.

attentional effects on oscillations and synchrony

The effects attention has on the activity of neuronal populations can be described in terms of synchronization and mean firing rate. The synchronization effects have already been reviewed in chapter 3: an increase of power in the γ frequency band of neurons with a receptive field that overlaps with an attended stimulus has been described [Tay+05; Fri+01; Fri+08]. In a setting where a target and a distractor stimulus are inside the receptive field of a neuron, γ synchronization between this neuron and neurons in earlier areas that exclusively process the target stimulus was found [Gro+12; Bos+12; Gro+15].

attentional rate effects

The rate effect of biased competition, reproduced in chapter 3, concerns the second layer of the routing setup (compare figure 3.1), where several stimuli fall inside of the receptive field of the neuron. When attentional effects on mean firing rates of populations corresponding to the first model layer were investigated, i.e. in a situation with a single stimulus inside the receptive field, while only one stimulus was presented, usually small to moderate firing rate increases were reported in V1 [IG99; Thi+09; Her+08], V2 [Luc+97; Buf+10], and V4 [MM99; RPD00], with a tendency toward higher effect sizes in higher visual areas [Luc+97; Buf+10]. The rate effects of attention on V1, V2, and V4 neurons become more profound if one or more nearby other stimuli are present outside of the classical receptive field of the recorded

neuron [Mot93]. In V1, presentation of a nearby distractor stimulus can strongly enhance attentional rate increases of the population processing the attended stimulus when compared to the single stimulus case [Mot93; LPG04; IG99]. More dramatically, some neurons even show an attentional rate effect only when neighbouring stimuli are present, and none for a single stimulus [IG99]. In addition, attention can evoke an activity suppression on the neurons processing the non-attended stimulus [Hop+06; Syl+08; SA14].

the RISER model causally links both effects

To date the functional relation between effects on the mean firing rate in both model layers and on the temporal properties of the activity in the γ range is unclear. The RISER model necessitates a mean rate increase of the target and at least an indirect suppression of the distractor representation in the first model layer, which causes suitable γ phase relations that support selective information routing. The latter also leads to the biased competition rate effect in the second model layer. As such, it is a promising proposal that unifies effects on mean rates and synchronization and causally links them in one framework.

locus of attentional intervention

Apart from the effects that arise from selective attention, another important question is how these effects are brought about and at which locus attentional processes intervene with neuronal processing. There is substantial debate in this matter. Investigating attentional effects in V4, V2 and V1, it was found that changes in firing rate and γ power were most pronounced and occurred with shortest latency in V4, weaker and later in V2 and least pronounced and latest in V1 [Buf+10]. The authors infer that attentional mechanisms first target higher visual areas and then operate via feedback projections on lower areas. Supporting evidence for this hypothesis was delivered by showing that neurons in the frontal eye field, a putative source for top-down attentional signals, induced an attention-dependent γ frequency coupling with neurons in V4 [Gre+09].

On the other hand, very fast attentional rate effects have been observed already in subcortical LGN neurons that directly project to V1, in fact too fast to be caused by cortical feedback [MCW08; KLZ13]. Furthermore, it is evident that stimuli filling two neighbouring V1 receptive fields, but sharing the same V4 receptive

field, can still be attentionally discriminated [Gro+12; Bos+12; MD85]. The spatial granularity of attention being on the order of V1 receptive field sizes is also supported by a large body of psychophysical literature (e.g. [MLF08; EH72; LaB83; Dow88]). This rather suggests that lower visual areas are natural candidates for attentional mechanisms to target directly.

prediction to test

Following the latter hypothesis, the RISER model incorporates attention as a mean firing rate increase of neurons processing an attended stimulus over neighbouring competing stimuli in a lower area (V1 or V2) that project to the same neuron in a higher area (V4). A prediction of this model is that a higher rate of the lower area neurons processing the attended stimulus than other lower area neurons processing nearby distractor stimuli is the crucial prerequisite for attentional selectivity to occur. At first glance, this is compatible with studies that find increased attentional rate effects in V1 under presentation of additional nearby stimuli [Mot93; LPG04; IG99]. This prediction holds under the assumption that both layer 1 populations drive the layer 2 population equally strongly. For symmetry reasons, such a V4 populations most likely exists for any two V1 or V2 populations.

attention vs. stimulus contrast

If the prediction from the model is correct, stimulus manipulations that have an effect on neurons' firing rates should be able to interfere with the desired attentional rate modulations, which attention has to compensate for in order for routing to function. This was critically tested in an established attention task, where a target stimulus is presented with a similar, nearby distractor stimulus [Gro+12], and activity was measured in V1. The task was modified by presenting the stimuli at different contrasts: the contrast of the target was chosen to be lower than the one of the distractor, creating a stimulus driven rate imbalance in favour of the distractor. It is found that attentional modulations adjust to this situation in a twofold way: the rate difference in favour of the distractor population is overcome as predicted by increasing the target population rate and decreasing the distractor population rate. The same situation was investigated for a target and a distractor with equal low or high contrast. Also here, firing rates show distractor suppression and target facilitation.

In addition, a second effect was discovered. The amount of distractor suppression only depends on the distractor's contrast, whereas target facilitation flexibly adjusts to the task demand. Taken together with different temporal developments of suppression and facilitation, this suggests that they are brought about by distinct mechanisms. This interpretation is supported by accumulating evidence from non-invasive electrophysiology and psychophysics (e.g.[Noo+16; Luc95]). A revised version of the RISER model is presented in the discussion, incorporating these novel findings.

5.2 MATERIALS AND METHODS

5.2.1 *Surgical procedures and training*

Intracortical recordings of V1 neurons were performed on an adult rhesus monkey (Macaca mulatta) that had been already trained on a similar task [Tay+05; Gro+12; Gro+15], and had undergone surgery to place a recording chamber over V1. All procedures and animal care were in accordance with the regulation for the welfare of experimental animals issued by the federal government of Germany and were approved by the local authorities.

5.2.2 *Recording*

Intracortical recordings in the upper layers of visual area V1 were performed using an epoxy-insulated tungsten micro-electrodes (125 μm diameter, 1 - 3 MΩ, FHC Inc., Bowdoin, ME, USA). The electrode signals were amplified 4000 fold (4 fold by a wideband preamplifier MPA32I and 1000 fold by a PGA 64, $1 - 5000$ Hz, both Multi Channel Systems GmbH, Germany) and digitized with 25 kHz sampling rate. The receptive field for each recording session was mapped manually based on MUA and LFP-responses, while the animal performed a fixation task.

5.2.3 *Task*

The monkey performed a variation of an established, highly attention demanding shape tracking task used for previous studies [Tay+05; Gro+12; Gro+15]. Visual stimuli were presented on a 20 inch CRT-monitor with a resolution of 1024 x 768 pixels and a refresh rate of 100 Hz. The screen was placed 90.5 cm in front of the monkey sitting in a custom-made primate chair. Visual stimulation comprised a fixation point and up to four simultaneously presented complex shapes. Figure 5.1 A) shows the sequence of stimuli and events of a single trial: first, the spatial cue appeared, which indicated the position of the relevant stimulus in the upcoming trial. The cue consisted of a ring (1 degree in diameter, 0.04 degree line width) centred over the position of the upcoming target stimulus. After 2.0 s, a central fixation square appeared with a side length of 0.15 degree, requiring the animal to fixate and initiate the trial by pressing a lever. Following the initialization of the trial, the spatial cue disappeared and after 1050 ms, three or four differently shaped stimuli appeared, all at a similar eccentricity between 2.5 and 3.5 degree of visual angle. Either one or two adjacent stimuli, the centres separated by 1.5 degree of visual angle, were located in the lower right visual field quadrant. This quadrant was where receptive fields of recorded neurons were located. The other two stimuli appeared at positions point-mirrored through the screen centre in the upper left visual field quadrant. Stimuli at all locations could serve as target.

The initial complex shapes at each stimulus location were presented statically for 510 ms and subsequently started to morph continuously into other complex shapes. A single morphing cycle, i.e. morphing completely from one shape into another, lasted 1.0 s. All shapes were taken randomly with equal probability out of a set of 12 shapes (figure 5.1 B). The reappearance of the initial shape at the cued stimulus location required the monkeys to release the lever within a time window ranging from 310 ms before the shape's complete reappearance to 400 ms afterwards. The initial shape at the cued location reappeared either as the third, fourth or fifth shape with equal probability. The appearance of the target's initial shape at the distractor locations and

the reappearance of the distractors' own initial shapes had to be ignored. Throughout the whole trial, the eye was monitored by video-oculography (IScan Inc., Woburn, MA, USA) and the inferred direction of gaze was not allowed to deviate from the screen centre by more than 0.5 degree. If the monkey released the lever within the response window, it was rewarded with a small amount of diluted fruit juice. If fixation was broken or a response occurred outside of the response window, trials were aborted without reward.

The stimuli were presented at different contrasts. The background luminance of the screen was set to 0.51 cd/m^2, and the stimuli could be presented at seven higher luminance values yielding Michelson contrasts of 0.12, 0.16, 0.22, 0.26, 0.32, 0.48, and 0.64. In the first part of each recording session, the contrast response function was acquired by presenting only one stimulus in the lower right quadrant inside the receptive field of the recording site, while the monkey performed the shape tracking task on one of two simultaneously presented stimuli in the upper left quadrant. The stimuli in the upper left quadrant were presented at maximal contrast of 0.64, while the stimulus inside the recorded receptive field was presented at all possible contrasts. From the resulting contrast tuning curve, two contrasts were chosen for the high contrast and the low contrast stimulus, such that the low contrast stimulus elicited $\approx$ 50% of the response to the high contrast stimulus. The choices of the low contrast ranged between 0.22 and 0.32. The high contrast was always either 0.48 or 0.64.

The second part of each session comprised a number of different stimulus configurations, where one stimulus or two stimuli with different contrast combinations were shown in the lower right quadrant, while always two stimuli were presented in the upper left one. The details can be found in table 5.1.

Only one V1 site was recorded at a time. The purpose of the experiments was to investigate the behaviour of two V1 populations at the same time, one with a receptive field centred on the target and one on the nearby distractor stimulus. To generate statements about this situation, it is assumed that V1 populations have equal properties at different locations, and thus the

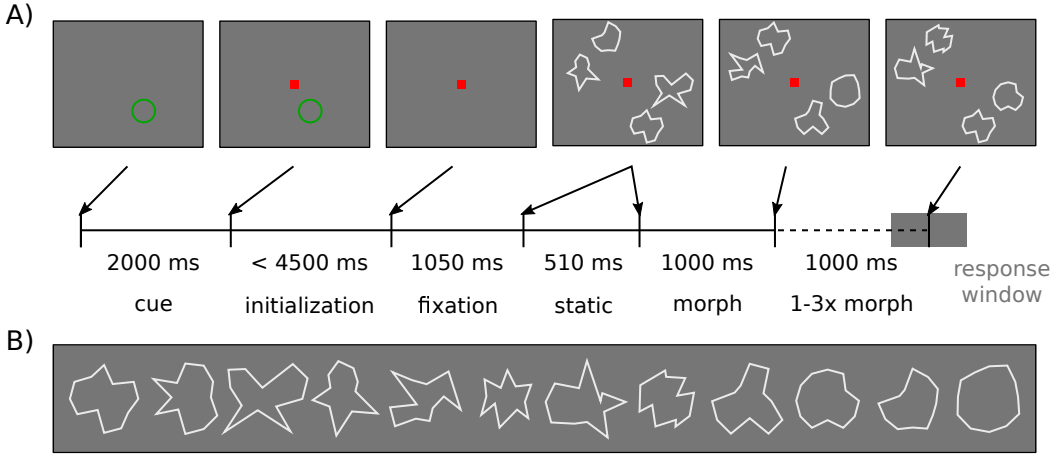

Figure 5.1: Task organization. **A)** In the beginning a location cue (green circle) is presented to guide attention to this location. The cue then disappeared and the monkey was required to accept the trial by pressing a lever within the initialization period. After that, fixation on the the red dot in the centre or the screen had to be held throughout the trial. In the following static period, four shapes at four locations with equal eccentricity (3° at shape centre) to the screen centre were presented. These shapes started morphing, and after 1 second completed the morphing into a new shape. This was repeated up to four times. The monkey had to report the reappearance of the initial shape presented during the static period at the cued location. Thus, the first morph cycle is not attentionally relevant since the initial shape could not reappear here. Drawings are not to scale. **B)** All 12 shapes used for stimulation.

location	rec	nei	rec	nei	rec	nei	rec	nei	rec	nei
target	no	no	no	no	no	no	no	no	yes	no
contrast	low	low	high	high	low	high	high	low	low	high
rate	ll		hh		lh		hl		$\bar{l}h$	

location	rec	nei	rec	nei	rec	nei	rec	nei	rec	nei
target	yes	no	no	yes	yes	no	no	yes	no	yes
contrast	low	low	low	low	high	high	high	high	high	low
rate	$\bar{l}l$		$l\bar{l}$		$\bar{h}h$		$h\bar{h}$		$h\bar{l}$	

Table 5.1: Task conditions and recorded stimulus configurations. In total, 10 stimuli configurations were recorded. The first three rows detail the stimulus configuration at the recorded (rec) and neighbouring (nei) site and whether one of the stimuli was the target stimulus. The fourth row shows the symbol for the firing rate recorded for the different configurations. By convention, the first letter denotes the contrast of the recorded site, the second the contrast at the neighbouring site, and the bar denotes which stimulus was the target. The other two stimuli in the non - recorded upper left visual quadrant were always shown at maximum contrast and one of them was the target, if neither the recorded or the neighbouring stimulus was.

behaviour of neurons for a specific condition at the neighbouring site is similar to the behaviour of neurons at the recorded site when recorded and neighbouring stimuli and target assignments are switched. Consequently, firing rates observed in the stimulus configurations in table 5.1 can be combined to represent the behaviour of both recorded and neighbouring site at the same time. For example, while $\bar{l}h$ is the rate observed for a low contrast target with a high contrast nearby distractor, the population active at the distractor site is assumed to fire at the rate $h\bar{l}$ at the same time.

In addition to these stimulus configurations, another condition was recorded to make sure that the neighbouring stimulus did not lie within the classical receptive field of the recorded neurons. For this purpose, a high contrast target was presented at the neighbouring location. For all results, only sessions were analysed where this condition did not elicit a firing rate significantly different ($p < 0.05$, Wilcoxon rank-sum test) from the baseline rate, which was extracted from the fixation period.

The firing rates in table 5.1, e.g. $\bar{l}h$, by default refer to mean rates of this stimulus configuration over a recording session. If all sessions are combined and stimulus configuration means over all pooled trials are taken, it is denoted by capital letters, e.g. $\overline{L}H$.

5.2.4 *Data analysis*

The raw data, recorded at a sampling rate of 25 kHz, was band pass filtered between 300 Hz and 12 kHz, after 50 Hz power line noise was removed. To obtain the MUA, putative action potentials recorded by the electrode were detected with a simple thresholding method [QNB04]. First, the noise standard deviation σ_n of the signal x, consisting of all concatenated recordings from the whole session, was estimated as

$$\sigma_n = m\left(\frac{|x|}{0.6745}\right) \quad ,$$

where m denotes the median. This estimation was performed for each recording session separately. The data was then thresholded at $k\sigma_n$, where k was chosen for each session by visually assess-

ing the signal to noise ratio. Peaks above this threshold were counted as action potentials. A peak at time index t occurred when $x_{t-1} < x_t \wedge x_t > x_{t+1}$. When multiple peaks were separated by less than 1 ms, the peak with the highest x_t was counted and the others rejected. Each morph cycle was considered as one trial. If not stated otherwise, firing rates of morph cycles 2 and 3 were pooled, since the target could not reappear at the end of the first morph cycle and thus no attentional allocation was needed during this period.

5.3 RESULTS

The mean firing rates of every recoding session of the V1 populations for the condition of low and high contrast nearby stimuli without target allocation (lh and hl), and with the low contrast stimulus being the target ($\bar{l}h$ and $h\bar{l}$) are shown in figure 5.2. In the former, the rate hl was higher than lh in all 27 sessions (diamonds), as expected from the single-stimulus contrast tuning. This difference is significant in all cases ($p < 0.05$, Wilcoxon rank-sum test). If the low contrast stimulus was the target, an increase of the population rate $\bar{l}h$ was observed in all sessions compared to lh. This rate increase co-occurred with a decrease of the distractor population rate $h\bar{l}$ compared to hl in all but one case (circles). This leads to 25 out of 27 sessions showing either no significant rate difference between $\bar{l}h$ and $h\bar{l}$, or a significantly higher target population rate $\bar{l}h$ (6 cases). In two cases, a significantly higher distractor population rate $h\bar{l}$ was observed. Pooling over all sessions shows that the grand average target rate $\overline{L}H$ is significantly larger than the grand average distractor rate $H\overline{L}$ ($p < 0.001$, Wilcoxon rank-sum test). With $LH = 17.63$ Hz, $HL = 32.42$ Hz, $\overline{L}H = 28.59$ Hz, and $H\overline{L} = 26.44$ Hz, the absolute mean target rate increase is $\overline{L}H - LH = 10.96$ Hz respectively the relative increase $\overline{L}H/LH - 1 = 62.16\%$. Effect sizes of distractor modulation are $H\overline{L} - HL = -5.98$ Hz and $H\overline{L}/HL - 1 = -18.44\%$. The bar plot insets display LH, HL, $\overline{L}H$ and $H\overline{L}$.

Figure 5.3 shows the same as figure 5.2 for two neighbouring low contrast stimuli. In the absence of a target, the rates ll for both populations are exactly on the equality line, since

these rates were only measured once and the neighbouring population was assumed to show the same firing rate. For a low contrast target and a low contrast distractor, the target population rate $\bar{l}l$ increased in 26 out of 27 sessions compared to ll, and the distractor population rate $l\bar{l}$ decreased in 23 of 27 sessions. All sessions show a higher rate of the target population than the distractor population, where the difference is significant in 24 of 27 cases ($p < 0.05$, Wilcoxon rank-sum test). The whole population average rates are $LL = 20.78$ Hz, $\bar{L}L = 29.01$ Hz, and $L\bar{L} = 18.58$ Hz. The effect sizes of target allocation are $\bar{L}L - LL = 8.23$ Hz or $\bar{L}L/LL - 1 = 39.60\%$, and $L\bar{L} - LL = -2.22$ Hz or $L\bar{L}/LL - 1 = -10.68\%$.

In figure 5.4 again the the same is plotted, but for two neighbouring high contrast stimuli. When a high contrast target was paired with a high contrast distractor, the target population rate $\bar{h}h$ increased compared to hh in 13 of 17 sessions, and the distractor population rate $h\bar{h}$ decreased compared to hh in 16 of 17 sessions. The target rate was always higher than the distractor rate, significantly so in 16 out of 17 cases ($p < 0.05$, Wilcoxon rank-sum test). The whole population average rates are $HH = 31.12$ Hz, $\overline{H}H = 38.21$ Hz, and $H\overline{H} = 24.72$ Hz. The absolute and relative effect sizes of target allocation are $\overline{H}H - HH = 7.09$ Hz, $\overline{H}H/HH - 1 = 22.78\%$, $H\overline{H} - HH = -6.40$ Hz, and $H\overline{H}/HH - 1 = -20.56\%$.

The rate changes between target and corresponding non-target conditions are further analysed in figure 5.5. The absolute rate increase a population experiences when becoming a target is significantly higher in the low contrast target, high contrast distractor condition compared to the equal contrast conditions ($p < 0.05$, ANOVA with Tukey-Kramer criterion). The rate decrease on the population that became the distractor was significantly higher when the corresponding stimulus had a high contrast than when it had a low contrast ($p < 0.05$, ANOVA with Tukey-Kramer criterion).

In a second step, the firing rates in trials where errors occurred were analysed. In error trials, a response was given outside of the response window, either too early or too late, indicating incorrect attentional allocation leading to a wrong or missed

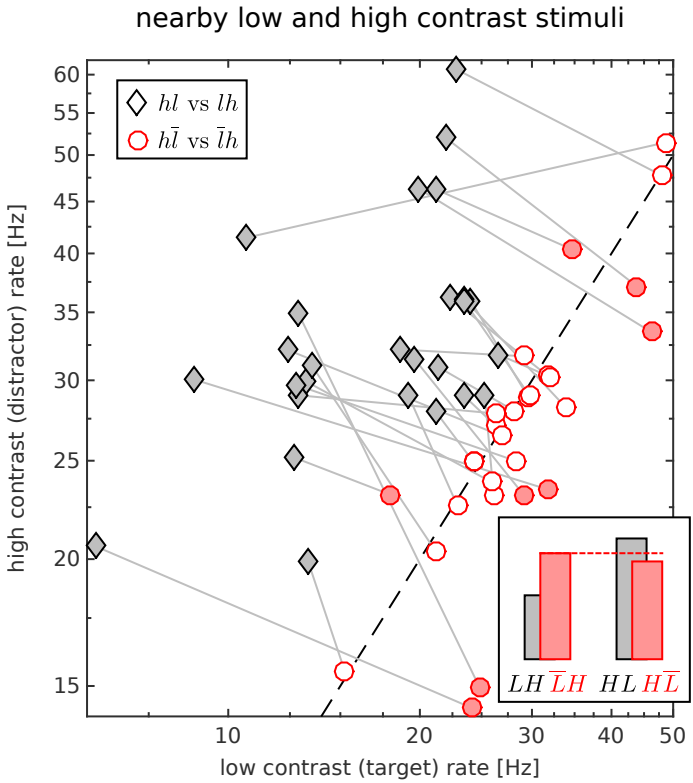

Figure 5.2: Firing rates of V1 neurons representing neighbouring low and high contrast stimuli. The abscissa represents the rates of populations with a low contrast stimulus in its receptive fields (lh, diamonds), and the rates in the same situation when the low contrast stimulus was the target ($\bar{l}h$, circles). Same recording sessions are linked by grey lines. The ordinate show the corresponding rates hl and $\overline{h}l$ of the population directly stimulated by the high contrast stimulus from the same recording session. In all sessions, hl was bigger than lh (diamonds, one per recording session). The dotted line marks rate equality. If the low contrast stimulus became the target, $\bar{l}h$ was increased in all sessions when compared to lh and the distractor population rate decreased in all but one session (circles). A filled symbol denotes a significant difference between lh and hl respectively $\bar{l}h$ and $h\bar{l}$ ($p < 0.05$, Wilcoxon rank-sum test). The grey bars in the inset represent LH and HL and the red ones $\overline{L}H$ and $H\overline{L}$. The average low contrast target population rate $\overline{L}H$ is significantly higher than the high contrast distractor population rate $H\overline{L}$ ($p < 0.001$, Wilcoxon rank-sum test).

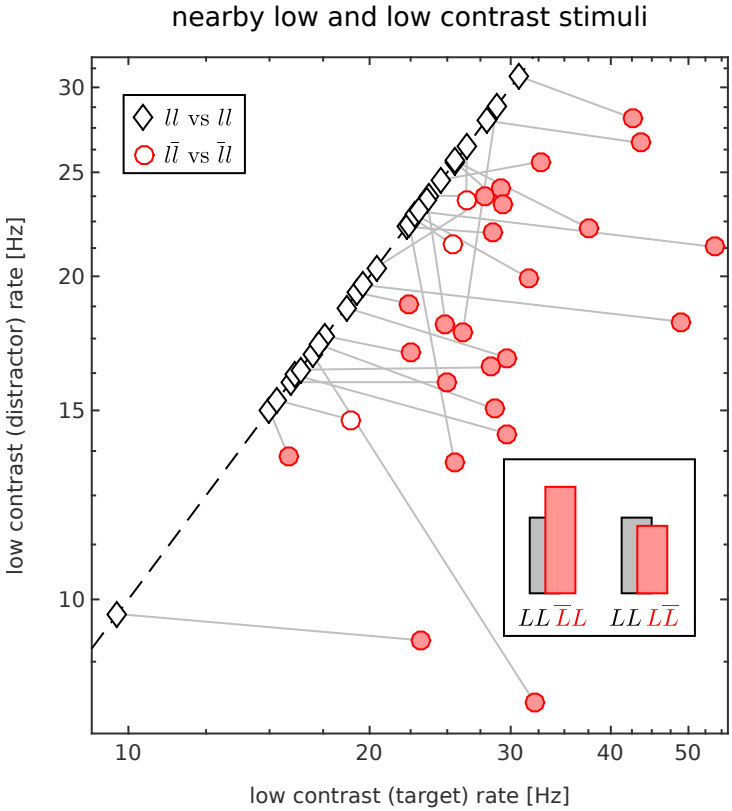

Figure 5.3: Firing rates of V1 neurons representing neighbouring low contrast stimuli. The dotted line marks rate equality. Without target, the rates were equal by definition (diamonds). If one stimulus became the target, the target rate $\bar{l}l$ was increased in all sites but one session and the distractor rate $l\bar{l}$ decreased compared to ll in 23 of 27 sessions (circles). A filled symbol denotes a significant difference between rates ($p < 0.05$, Wilcoxon rank-sum test). The grey bars in the inset represent LL and the red ones $\bar{L}L$ and $L\bar{L}$, obtained by pooling over recording sessions.

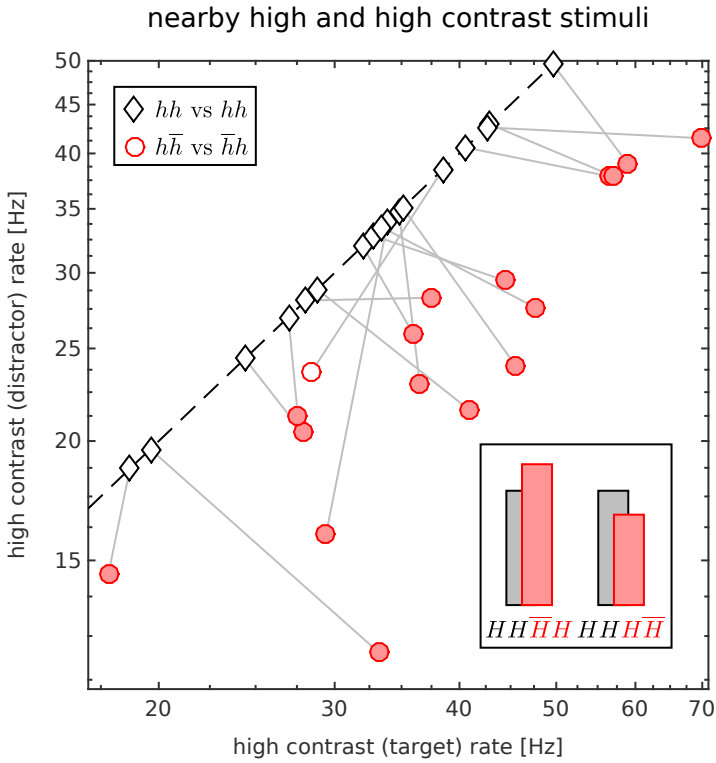

Figure 5.4: Firing rates of V1 neurons representing neighbouring high contrast stimuli. The dotted line marks rate equality. Without target, the rates were equal by definition. If one stimulus became the target, the target rate $\overline{h}h$ was increased in 13 of 17 sessions and the distractor rate $h\overline{h}$ decreased compared to hh in 16 of 17 cases (circles). A filled symbol denotes a significant difference between rates ($p < 0.05$, Wilcoxon rank-sum test). The grey bars in the inset represent HH and the red ones $\overline{H}H$ and $H\overline{H}$, obtained by pooling over recording sessions.

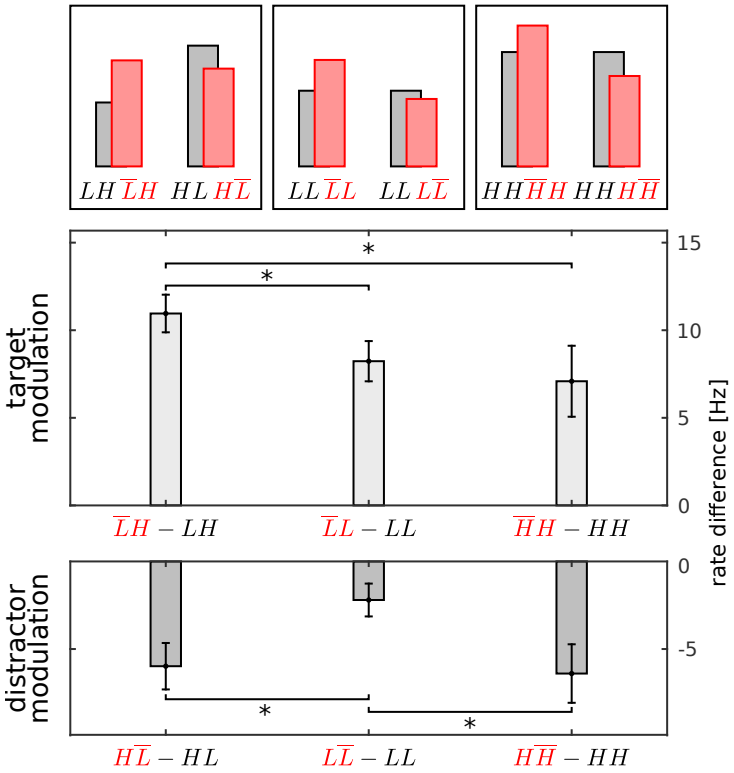

Figure 5.5: Average rate effect of target allocation on populations rates in all contrast combinations. The bars on the top show the average rates in all conditions. The light grey upward bars show the rate increase a population experienced when becoming a target. The darker grey downward bars represent the average rate decrease a population was subjected to when becoming a distractor. Error bars show $1.96 \times$ SEM and the asterisks denote significant differences. The target population rate increase is significantly higher in the low contrast target, high contrast distractor condition (left) when compared to the other two ($p < 0.05$, ANOVA with Tukey-Kramer criterion). The rate decrease of a population when becoming a distractor is significantly lower when the corresponding stimulus is of low contrast (middle) ($p < 0.05$, ANOVA with Tukey-Kramer criterion).

detection. Figure 5.6 reveals that the rate relation between $\bar{l}h$ and $h\bar{l}$ follows a different pattern than when the trial was successful. In total, only 7 out of 27 sessions showed a higher rate for the target stimulus population, which is the reverse of the pattern observed in successful trials (compare circles in figure 5.2). In 3 sessions, the distractor population rate $h\bar{l}$ was significantly higher than the target population rate $\bar{l}h$. Pooling data from all sessions, the average rate of the distractor population $H\overline{L}$ was significantly higher than the target population rate $\overline{L}H$ ($p < 0.001$, Wilcoxon rank-sum test). This pattern inversion does not occur in the cases of equal target and distractor contrast. Here, all target population rates were still higher than the distractor population rates. For this error analysis, only the firing rate in the last cycle before the response was used since only during this cycle the attentional allocation has surely failed.

The error frequency, i.e. the # of error trials / # of all trials, averaged over recording sites, was 33.11% in the low contrast target, high contrast distractor trials, 31.65% in the low contrast target, low contrast distractor trials, and 16.82% in the high contrast target, high contrast distractor trials. The latter is significantly different from the former two ($p < 0.05$, Wilcoxon rank-sum test with Bonferroni correction). Only considering error trials where the monkey responded before the response window, corresponding to an active wrong decision, the error frequencies were 17.58% (low contrast target, high contrast distractor), 13.89% (low contrast target, low contrast distractor), and 8.00% (high contrast target, high contrast distractor). Here, all pairwise comparisons reveal a significant difference.

Up to this point, firing rates from task cycles 2 and 3 were combined for the analyses. Now the focus is on the temporal evolution of the attentional rate effects during a trial.

First the average firing rate differences between the two stimulated V1 populations was investigated with and without target allocation. The time periods of the trial are defined in figure 5.1. Figure 5.7 A) shows $LH - HL$ and $\overline{L}H - H\overline{L}$ during the static periods and cycles 1 to 3. The rate difference was highest and in favour of the high contrast stimulus for both $LH - HL$ and $\overline{L}H - H\overline{L}$ during the static period. However, whereas without tar-

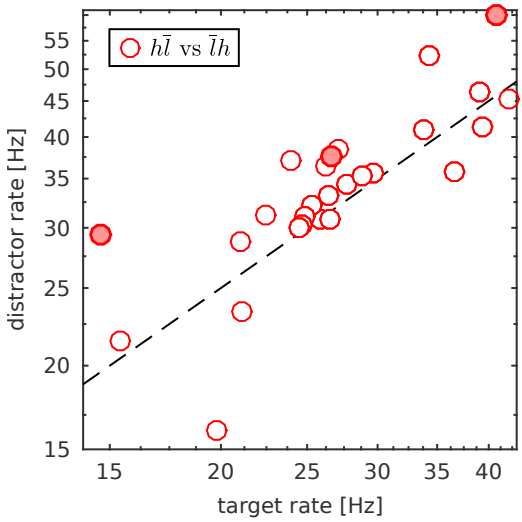

Figure 5.6: Firing rates of V1 neurons in the low contrast target, high contrast distractor condition in trials where errors occurred. The dotted line marks rate equality. the target rate $\bar{l}h$ was higher than the distractor rate $h\bar{l}$ in 7 out of 27 cases. A filled symbol denotes a significant difference between rates ($p < 0.05$, Wilcoxon rank-sum test). The three significant rate differences all show a higher distractor rate. Compare circles in figure 5.2 for the same rates in successful trials.

get allocation the rate difference $LH - HL$ did not change from cycle 1 to 3 and stayed in favour of the high contrast stimulus, in the low contrast target, high contrast distractor condition the rate difference $\overline{L}H - H\overline{L}$ did, to a point where the target population rate $\overline{L}H$ was above the distractor population rate $H\overline{L}$ during cycle 2 and 3. Notably, these are the periods in which attentional intervention would be expected most since the target shape cannot reappear at the end of cycle 1. As shown in figure 5.7 B), the target population rate was always above the distractor population rate for equal contrast conditions, with an increased difference in cycle 2 and 3.

Also during the static period a higher rate of the target population could be expected to route the target stimulus to a higher visual area for later comparison. However, this was not the case

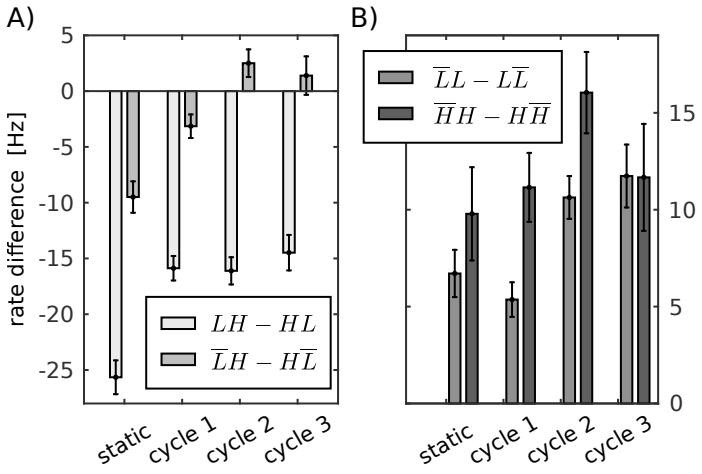

Figure 5.7: Temporal evolution of rate differences between neighbouring populations. **A)** shows the rate differences between the low contrast population and the high contrast population ($LH - HL$, brighter grey), and the same differences with target allocation to the low contrast stimulus ($\overline{L}H - H\overline{L}$, darker grey), temporally broken down into static period and cycles 1 to 3. Error bars show $1.96 \times$ SEM (standard error of the mean). In both conditions, the rate imbalance in favour of the high contrast population was highest in the onset period. The difference $LH - HL$ did not change between the cycles. However, it did when the low contrast stimulus was the target ($\overline{L}H - H\overline{L}$) to a degree where the target rate was greater than the distractor rate in cycles 2 and 3. **B)** Rate differences in the low contrast target, low contrast distractor condition (middle grey) and in the high contrast target, high contrast distractor condition (dark grey). Both conditions show a rate difference in favour of the target population throughout the periods, with the tendency to an increased difference in cycles 2 and 3.

in the low contrast target, high contrast distractor condition (figure 5.7 A). Figure 5.8 shows the time course of average pooled firing rates within the static period for this condition. A first transient response is followed by a period of insignificant difference between target and distractor population rates. Thus, the high difference between $H\overline{L}$ and $\overline{L}H$ in favour of the distractor popu-

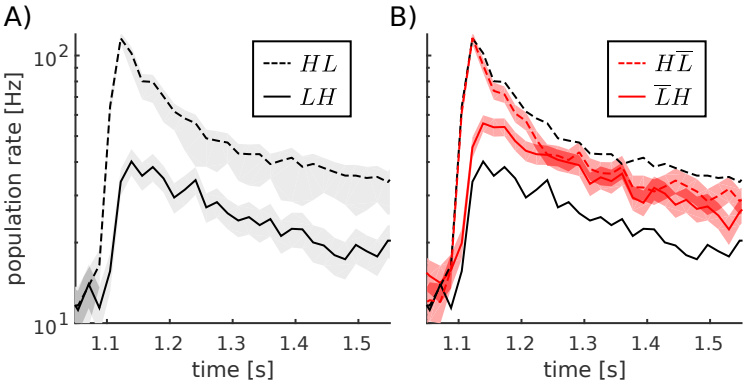

Figure 5.8: Average time courses of low contrast and hight contrast population rates during the static period. **A)** The high contrast population rate HL is higher than the low contrast rate LH during the whole static period. **B)** Rates $\overline{L}H$ and $H\overline{L}$ during the static period. Averaged over the full static period from 1.05 to 1.55 s, the distractor population rate is higher than the target rate (figure 5.7 A)). However, this higher resolution plot shows that there is a period of insignificant difference between $\overline{L}H$ and $H\overline{L}$ after the initial transient response. The red shadings represent 1.96 × SEM. The black lines show the mean rates LH and HL from A).

lation averaged over the whole static period is mostly due to the strong stimulus onset transient.

Also, the development of rate effects over time between target and corresponding non-target conditions was investigated. Figure 5.9 shows a time resolved version of figure 5.5. The contrast dependent distractor population suppression did not change significantly over time in all contrast pairings and was already present during the static period. The same does not hold for the target population rate increase: In the low contrast target, high contrast distractor condition, a high rate increase during the static period was followed by a significantly lower increase in cycle 1, in turn again followed by a significantly higher rate increase in cycle 2. In the low contrast target, low contrast distractor condition, the rate increase was relatively low during onset and cycle 1, and then grew significantly towards cycle 2 and 3. A similar pattern

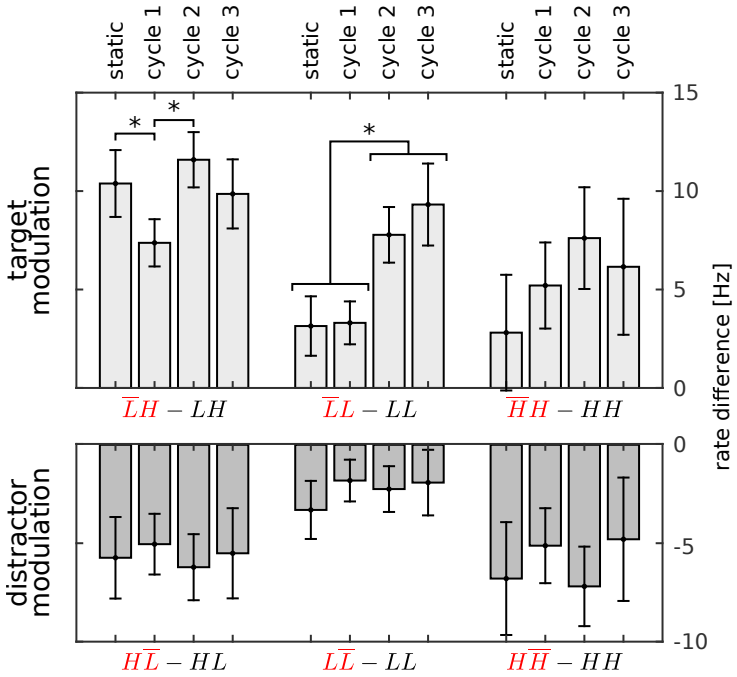

Figure 5.9: Time resolved version of figure 5.5. Error bars represent $1.96 \times$ SEM and asterisks denote significant differences. The distractor population rate suppression exhibited no significant changes across time periods. The attentional rate increase was high during the onset period in the low contrast target, high contrast distractor condition (left), decreased significantly during cycle 1 and increased again to cycle 2 and 3 ($p < 0.05$, ANOVA with Tukey-Kramer criterion). For an equal low contrast distractor, rate increases were relatively low during onset and cycle 1, and increased significantly to cycle 2 and 3 ($p < 0.05$, ANOVA with Tukey-Kramer criterion). A similar patter could underlie the observation in the high contrast target, high contrast distractor case, but the rate differences are not significant.

can be seen in the high contrast target, high contrast distractor condition, albeit without significant differences.

5.4 DISCUSSION

target rate > distractor rate

The results show that under attention the rate of the target population is increased while the rate of the distractor population is decreased such that the target population rate lies above the distractor population rate in all tested conditions. If the difference to the distractor population rate is high (low contrast target, high contrast distractor), the attentional rate increase of the target population is higher than when the distractor population has a similar rate. In the low contrast target, high contrast distractor condition, errors occur when the prerequisite of the target population rate being larger that the distractor population rate is not fulfilled.

populations, not single neurons matter

At this point, it shall be noted that the predicted rate imbalance in favour of the target population is an explicit requirement on populations, not single neurons. Consider the case of a strongly colour tuned neuron with a preference for red and no response to blue. Further assume that a blue stimulus is presented at the recorded and a red stimulus at the neighbouring site. This neuron, when part of the recorded ensemble will show no response. If it is in the neighbouring populations, it will be maximally active. It is not plausible that this neuron is now elevated by attention at the recorded site and suppressed at the neighbouring site such that it is more active at the recorded site. It is more plausible that the rate imbalance is defined between the subsets of neurons that encode the respective stimuli well, i.e. the blue preferring neurons in the recorded ensemble and the red preferring neurons in the neighbouring ensemble. This also explains why finding some sessions where $\bar{l}h < h\bar{l}$ is not detrimental to the statement, and rather the pooled average is of importance. It might have occurred that a small set of neurons was recorded in a session which were suboptimally driven by the stimuli.

To alleviate this potential confound as much as possible, the contrast feature was chosen to manipulate neurons' responses. Contrast tuning is one of the most thoroughly investigated tuning properties in early visual cortex. Although there is heterogeneity from cell to cell, the overwhelming majority of neurons

described so far shows a monotonically increasing response with contrast irrespective of other tuning properties (e.g. [Alb95; AH82; Bon91; CH94; Alb+02]). This pattern is unaffected by the mean luminance of the stimulus [GAC07]. Thus, if a reasonably active subset of neurons was recorded, it is to be assumed that these neurons also participate in the population response to a stimulus which differs in contrast.

data is consistent with RISER model prediction

Taken together, this data is consistent with the prediction that a higher rate of the target population is a prerequisite for selective attentional information routing. A further case in point is that the contrast difference was chosen close to the point at which the monkey refused to perform the task, presumably because it got too difficult for larger contrast differences. Incidentally in the low contrast target, high contrast distractor condition, the target population rate is barely larger than the distractor population's, whereas the difference is much bigger for equal contrast conditions. Also, the error frequency is highest for the low contrast target, high contrast distractor condition. Thus it is plausible that the point at which the monkey refuses to perform the task is the point where the target population rate cannot be increased over the distractor population rate. An error frequency difference observed between the low contrast target, low contrast distractor and the high contrast target, high contrast distractor conditions could be explained by the task with low contrast stimuli being naturally harder: not only the distractor has to be gated out, but also the target perceived in the first place, which was close to the perceptual threshold for the human experimenter.

However, establishing a rate excess of the target population over the distractor population does not guarantee successful task performance: errors in equal contrast conditions can also occur when a higher target population rate is given. Possibly the respective errors have different causes in processing stages beyond V1. A possible explanation in the CTC framework would be that although the mean rates in V1 were properly set up for selective routing, the phase relations between communicating V1 and V4 populations were temporarily disturbed, possibly by transient rate fluctuations. Conversely, successful performance does entail

a higher target rate, consistent with the assumption that the rate imbalance in favour of the target is a necessary condition.

suppression and facilitation stem from different processes

The rate changes brought about by attention, which are necessary to abolish the rate differences between target and distractor population, exhibit a quite complex pattern. During the task relevant cycles 2 and 3 (figure 5.5), attentional suppression of the distractor population seems to be independent of the firing rate of the target population, and to scale with the firing rate of the distractor population in the absence of attention. The rate increase of the target population does not scale with target or distractor rate, rather with the difference between the two: it is higher in the low contrast target, high contrast distractor condition than when target and distractor are of the same contrast. Importantly, the absolute effect sizes of target facilitation and distractor suppression do not co-vary, which would be expected if they are signatures of the same mechanism. Furthermore, when analysing the temporal evolution of attentional rate effects, it is found that the suppression is constant over a trial and already present during the static period. On the other hand, in the low contrast target, low contrast distractor condition the attentional increase develops over the trial such that it is maximal during cycles 2 and 3 when the target stimulus can possibly reappear (figure 5.9). In the low contrast target, high contrast distractor condition, a similar pattern shows, with the difference that a strong attentional rate increase occurs during the static period. This extra increase could be explained by the need to route the target shape for later comparison to determine the shape's reappearance. Indeed, within the static period, a time window exists with insignificant difference between target and distractor rate where this routing could take place (figure 5.8).

Taking both the different time courses and the different patterns of effect sizes into account, they imply that suppressive and facilitatory attentional effects originate from two distinct processes: A stimulus driven, 'bottom up' distractor suppression, and a cue driven, 'top-down' target facilitation which can be flexibly adjusted to task demands. This finding, to my knowledge, has not been described on a population rate basis before.

supporting evidence from other studies

However, there is accumulating supporting evidence for this hypothesis from behavioural studies and non-invasive electrophysiology. In a psychophysical study trying to dissociate stimulus driven and cue driven attentional effects [CG98], it was found that especially stimulus driven attention leads to a decrease of discrimination performance at distractor locations close to the target. The authors hypothesize that this is due to a lower saliency of the distractor stimulus. A lower distractor saliency is akin to a lower perceived contrast brought about by a decreased distractor population rate. Results from [Luc95] also suggest that facilitation and suppression are functionally distinct processes: Investigating the effects of distractor suppression and target facilitation in electroencephalography (EEG) time courses, it was found that both appear at different times and can be observed in the absence of one another. A further recent result from EEG studies is that inter-individual differences in the visual working memory capacity between subjects correlate with different attentional strategies. Subjects with a high visual working memory capacity both employ target facilitation and distractor suppression, whereas subjects with a low capacity fail to show distractor suppression, and only target representations are enhanced [Gas+16; Gul+14]. This is a further case in point that distractor suppression and target facilitation arise from different mechanisms, and are not two sides of the same coin. Recently, by investigating attentional changes on reaction times, it was reported that target cueing leads to flexible target facilitation, whereas distractor cueing is ineffective for distractor suppression, suggesting that target facilitation is a top-down process, whereas distractor suppression is not [Noo+16].

RISER model is compatible with independent suppression and facilitation

The RISER model developed in chapter 3 can be updated to incorporate independent attentional facilitation and suppression. In addition to the Δ^{+}_{add} applied to the excitatory and inhibitory subpopulation of the first layer population processing the attended stimulus, Δ^{-}_{add} is subtracted from the excitatory and inhibitory drive to the population representing the non-attended stimulus. Figure 5.10 shows attentional rate effects corresponding to figure 5.5 with $\Delta^{+}_{att} = 0.3$ in the low contrast target, high contrast distractor condition and 0.2 in the other conditions, and $\Delta^{-}_{att} = -0.15$ in the conditions with a high contrast distractor and -0.075 in the

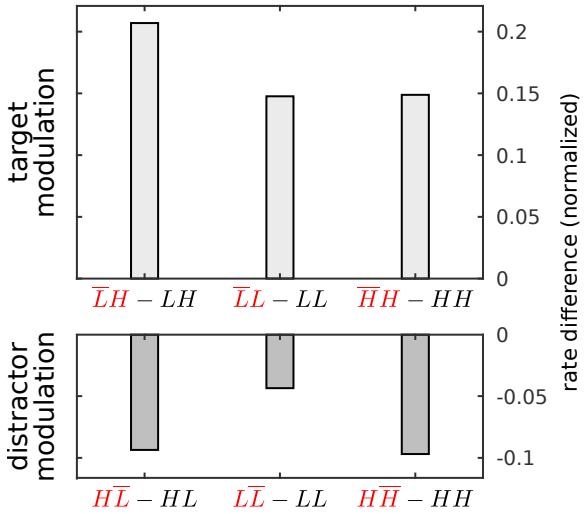

Figure 5.10: Reproduction of effects on target and distractor population rates in the RISER model. According to the experimental findings, an additional attentional signal Δ^{-}_{att} was incorporated which decreases the input to the input to the first layer population processing the non attended stimulus. The parameters are $w^{ii}_{loc} = w^{ie}_{loc} = -3$, $w^{ei}_{AB} = 1.5$, and $w_{ff} = 0.5$. Furthermore $\Delta^{+}_{att} = 0.3$ in the low contrast target, high contrast distractor condition and $\Delta^{+}_{att} = 0.2$ in the others. Δ^{-}_{att} was set to -0.15 in the conditions with a high contrast distractor and -0.075 in the low contrast target, low contrast distractor condition. For the model setup, please refer to figure 4.1.

low contrast target, low contrast distractor condition. A simplified pure ING rate-based model (chapter 4) was simulated with $w^{ii}_{loc} = w^{ie}_{loc} = -3$, $w^{ei}_{AB} = 1.5$, $w_{ff} = 0.5$ and all other weights set to 0.

5.4.1 *Locus of attentional intervention*

A further point of consideration is the target of the attentional signal. This study highlights the importance of firing rate relations at a cortical stage where stimuli are still separately processed (here V1), but project to the same population in a higher area (here

locus of attentional intervention might depend on stimulus configuration

V4), consistent with the RISER model. However, as already noted in the introduction, a backward progression of latency and size of attentional effects was described, being fastest and strongest in V4, and latest and weakest in V1 [Buf+10]. This apparent disagreement can be potentially explained by differences in tasks: The stimuli used in the experiment in [Buf+10] were set up far apart as to *not* cause biased competition in the recorded V4 neurons. However, I conceptualize attention as a mechanism to resolve competition. An attentional signal would be most effective at the stage of strongest competition, i.e. layer 1 in the model, before both stimuli are integrated in a successive processing stage. Thus it follows that also the area attention targets depends on the stimulus and task configuration. Given the specific stimulus setup, this target area could well be V4 in the aforementioned experiments, meaning that V4 would correspond to layer 1 in the RISER model. Whereas the results from the present study advocate the validity of this rationale for the locus of attentional intervention, it cannot be completely clarified. For this purpose, recordings with different stimulus distances would be necessary to probe whether increasing distance leads to similar rate effects as observed here in higher cortical areas.

5.4.2 *Previous studies on contrast and attention*

The idea that contrast and attention might apply similar mechanisms is founded on the observations that sustained attention can be described as a contrast gain in a range of studies [LC06; MT02; RD03; CLR04]. Notably, in [RD03] a similar situation was investigated where two stimuli of different contrast were placed in the receptive field of a V4 neuron. The authors report that the stimulus mainly represented in the V4 neuron is the one of higher contrast and hypothesized that the visual system is hard-wired to select the stimulus of highest contrast among competitors. They further hypothesize that attention has to overcome the contrast imbalance such that the lower contrast stimulus can be preferentially processed. The present study complements this finding in showing that, indeed, attention compensates for the contrast imbalance by intervening in V1 populations that project to V4.

5.4.3 *Alternative routing mechanisms*

In [Fri15] it was hypothesized that an increase in γ synchrony paired with a slight increase in the population γ frequency of the attended layer 1 population is sufficient to selectively route information. The attended population would then transiently entrain the layer 2 population and during these locked phases slightly precede activity incoming from the non-attended layer 1 population. Whereas a slight increase of the γ frequency in a local population in the RISER model due to attentional facilitation could also occur, this alternative routing scheme would need only little attentional mean rate effects in layer 1. Especially, a necessary increase of the target population rate over the distractor population rate would not be expected. However, the present results showing strong rate effects in layer 1 are consistent with the rate imbalance in favour of the target population being a necessary condition for selective routing, and thus rather fit into the framework of the RISER model.

5.5 ACKNOWLEDGEMENTS

The study was designed by me and Udo A. Ernst, in collaboration with the lab of Andreas K. Kreiter. The recordings were performed by Lukas Rausch and Eric Drebitz in the Kreiter lab. I analysed and visualized the data.

6 CONCLUSION AND OUTLOOK

This concludes the part on attentional mechanisms of selective routing. The purpose was to develop a network model and investigate dynamics on these networks that could subserve flexible information processing. As a simple example of this flexible processing, a experimentally well researched attention paradigm was used as a starting point where three or four distinct populations, two in a lower network layer and one or two in a higher layer, engage in collective dynamics.

In chapter 3 a model for selective routing, dubbed the RISER model, was proposed and shown to reproduce several electrophysiological effects of attention that were not reconciled before. It implements the previously formulated CTC hypothesis in a specific way such that flexible routing configurations self-organize and can be guided by simple attentional signals. This model was shown to be optimal for selective information routing. A model specific prediction was tested and confirmed in a specifically designed electrophysiological experiment.

Additionally, details of how attention influences activity in this network could be disentangled. The data from the experiment suggests that target facilitation and distractor suppression are mediated by distinct mechanisms, which is supported by psychophysical and EEG studies. The RISER model can incorporate such a dual mechanism without problems.

In short, a good candidate circuit underlying flexible processing in the primate brain was developed and the understanding of attentional mechanisms furthered. Since no striking counter arguments are known to me, I would even speculate that the RISER model is a canonical switch motif that might be implemented throughout the brain. Chapter 3 was published in a similar form [HEP15], and chapters 4 and 5 are being prepared for publication. For the latter, data from a second monkey is needed which is being recorded at the moment.

The validity of the RISER model could further be probed by investigating γ-phase relations between competing first layer populations. A clear prediction is that populations close enough to enter competition under attention will engage in anti-phase γ-oscillations, whereas populations further apart will not. To date, such distance dependent phase relations have not been investigated to my knowledge. Along the same lines, also target facilitation and distractor suppression should occur at different cortical stages for different stimuli distances. This could be tested in the same experiment.

For future work, the RISER model can be used as a testbed to determine the effects of artificial stimulation in the cortex. In an ongoing study, the effect of stimulation with short electrical pulses in different subpopulations is investigated to determine whether they can be used to interfere with attentional allocation [LE16]. Similar experiments are being performed in primate studies in the Kreiter lab, which, if successful, would further bolster the oscillation phase based routing theories.

After the local circuitry for routing has been established, it remains to be shown that several RISER modules can be combined in parallel to form a coherent model of two successive visual areas with receptive fields spanning the whole visual field. I hypothesize that the network can be set up by a distance dependent lateral interaction profile such that strong oscillations are only evoked in locations where different stimuli are close to each other. At other locations, where single stimuli are present and no competition has to be resolved, the network would operate in a rather stochastic spiking mode. Hence, oscillations could be used as much as necessary and as little as possible. Such a mode of operation is favourable since strong stereotypic intrinsic dynamics can be detrimental for information representation.

Part II

THEORY OF EFFECTIVE INFLUENCE

OUTLINE PART II

This second part of the thesis delves deeper into the question of how effective influence can be defined in a complex dynamical system, resulting in the presentation of the novel theory of topological causality.

First, chapter 7 lays out the mathematical foundation for the following. Chapter 8 then contains the theory of the novel definition of effective influence and showcases some applications, as well as discusses important properties. This chapter was published in similar form [Har+17]. The second part ends in chapter 9, which gives a conclusion and an outlook.

7 MATHEMATICAL FOUNDATIONS

The novel theory of effective influence will be related to the established methods of Granger causality and convergent cross-mapping. These two measures are introduced in the following, along with Takens' theorem, which is the foundation for both convergent cross-mapping and the newly developed measure.

7.1 GRANGER CAUSALITY

Granger causality (GC) [Gra69] is a statistical concept for the detection of causal links between two stochastic variables x_1 and x_2 which generate discrete time series. It holds for stochastic linear systems, i.e. it assumes that the time series of x_1, x_2 can be written as a linear autoregressive model of the form

$$\begin{aligned} x_1(t) &= \sum_{k=1}^{p} w_{11,k} x_1(t-k) + \sum_{k=1}^{p} w_{12,k} x_2(t-k) + \eta_1(t) \\ x_2(t) &= \sum_{k=1}^{p} w_{22,k} x_2(t-k) + \sum_{k=1}^{p} w_{21,k} x_1(t-k) + \eta_2(t) \quad , \end{aligned}$$

where p is the model order. The residuals η_1, η_2 are assumed to be noise processes. The idea is that if knowledge about the past of x_2 yields an advantage for predicting $x_1(t)$ over a prediction based on the past of x_1 alone, then x_2 has an influence on x_1. Or 'x_2 Granger causes x_1'. Given that some coefficients in $w_{12,k}$ are non-zero, the rationale is easy to see: not knowing about influence from x_2 to x_1, i.e. assuming that the coefficients $w_{12,k}$ are 0 while they are not, effectively absorbs $\sum_{k=1}^{p} w_{12,k} x_2(t-k)$ into $\eta_1(t)$, leading to a larger non-deterministic residual $\eta_1(t)$ and hence a worse prediction of $x_1(t)$. The coefficients being non-zero can be tested by an F-test against the null-hypothesis, where the logarithm of the statistic is related to the size of coefficients [Gew82].

Specifically, let ${}^e x_i(t)|x_i(t-1),...,x_i(t-p)$ be the prediction of $x_i(t)$ based on the past values of x_i alone, and ${}^e x_i(t)|x_i(t-1),...,x_i(t-p),x_j(t-1),...,x_j(t-p)$ the prediction including the past of x_j. The prediction errors $E(x_i)$ and $E(x_i,x_j)$ averaged over the whole time series can then be written as

$$E(x_i) = \langle [x_i(t+1) -^e x_i(t+1)|x_i(t),...,x_i(t-p)]^2 \rangle_t \quad ,$$

and

$$E(x_i,x_j) = \\ \langle [x_i(t+1) -^e x_i(t+1)|x_i(t),...,x_i(t-p),x_j(t),...,x_j(t-p)]^2 \rangle_t$$

Granger causality, following [Gew82], is then quantified as

$$C^G_{j\to i} = \log\left(\frac{E(x_i)}{E(x_i,x_j)}\right) \quad . \tag{7.1}$$

Now for a simple example of a autoregressive system with order $p = 1$, the model reads

$$\begin{aligned} x_1(t) &= w_{11}x_1(t-1) + w_{12}x_2(t-1) + \eta_1(t) \\ x_2(t) &= w_{22}x_2(t-1) + w_{21}x_1(t-1) + \eta_2(t) \quad . \end{aligned}$$

Since everything is linear and the influence from system 2 to system 1 is mediated by the coupling weight w_{12}, $C^G_{2\to 1}$ will be proportional to this coupling weight: The larger the weight, the more the prediction accuracy for $x_1(t)$ will increase through knowledge of $x_2(t-1)$.

GC requires separability

The concept of separability is fundamental here. It reflects in the notion that possible explanations, or causes, for x_1, can be either considered or not for prediction of $x_1(t)$. Mathematically, this is due to the linear stochastic nature of the system which allows $w_{12}x_2(t-1)$ to be absorbed in the noise $\eta_1(t)$. This means that a coupled system $\{x_1,x_2\}$ can be decomposed into its components, such that each component x_1, x_2 on its own has less information than the full system.

For simplicity only 2 coupled variables were used here, but the concept holds for an arbitrary amount.

7.1.1 *Transfer entropy*

Transfer entropy [Schoo] can be seen as an information theoretic formulation of GC for auto-regressive processes [BBS09]. However, it is more general in the sense that it does not require linearity of the system. As GC, however, it relies on separability and the components being random variables.

7.2 TAKENS' THEOREM

Takens' theorem [Tak81] makes a statement about attractor reconstructability of dynamical systems from the individual system components. It can be phrased as follows:

Let the (multidimensional) states $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of two components of a deterministic dynamical system be governed by

$$\begin{aligned}\dot{\boldsymbol{x}}_1 &= f_1(\boldsymbol{x}_1, w_{12}\mu_2(\boldsymbol{x}_2))\\ \dot{\boldsymbol{x}}_2 &= f_2(\boldsymbol{x}_2, w_{21}\mu_1(\boldsymbol{x}_1))\end{aligned}$$

where $\mu_i(\boldsymbol{x}_i)$ denote fixed scalar functions and w_{ij} coupling constants. A component here refers to a mutually coupled set of variables. The trajectories $(\boldsymbol{x}_1(t), \boldsymbol{x}_2(t))$ form an invariant manifold, an attractor, in the phase space of the joint dynamical system. A manifold in a delay coordinate space visited by

$$\boldsymbol{r}^{x_i}_{\phi_i}(t) = (\phi_i(\boldsymbol{x}_i(t)), \phi_i(\boldsymbol{x}_i(t+\tau)), \ldots, \phi_i(\boldsymbol{x}_i(t+(m-1)\tau)))$$

is topologically equivalent, meaning that a homeomorphic mapping between both manifolds exists, if $w_{ij} \neq 0$ and the embedding dimension m is sufficient. If $m > 2d$, where d is the dimensionality of the attractor of the joint dynamical system, it is certainly sufficient. Here, ϕ_i is a measurement function depending on a scalar component of $\boldsymbol{x}_i$, which will be omitted to simplify the notation: $\boldsymbol{r}^{x_i} := \boldsymbol{r}^{x_i}_{\phi_i}$. A recent addition to the theorem shows that by transitivity, a surjective smooth mapping from $\boldsymbol{r}^{x_i}(t)$ to $\boldsymbol{r}^{x_j}(t)$, denoted by $M^t_{i\to j}$, exists *iff* $w_{ij} \neq 0$ [CGS15].

This means that the attractor of a mutually coupled dynamical system can be reconstructed by observing dynamics in a time-delay coordinate space from each single component. Crucially,

dynamical systems are generically not separable

the system is in general *not* separable, since each component contains the information of the full system attractor. Thus, homeomorphic mappings between the reconstructions and the original attractor and amongst each other exist. Also here, only two components are used for simplicity, but the theory encompasses arbitrarily many components.

As an example consider the famous Lorenz system [Lor63] given by

$$\begin{aligned} \dot{x}_1(t) &= 10(x_2 - x_1) \\ \dot{x}_2(t) &= x_1(28 - x_3) - x_2 \\ \dot{x}_3(t) &= x_1 x_2 - \frac{8}{3} x_3 \end{aligned} \tag{7.2}$$

Figure 7.1 shows a trajectory in the phase space of the system and trajectories in time-delay coordinate systems based on x_1 and x_2. The characteristic "butterfly" shape of the original attractor shows as well in the reconstructions, albeit a bit warped and bent. The homeomorphic map between all manifolds ensures that they can be morphed into each other only by "stretching and squeezing" and without "ripping", figuratively speaking.

7.3 CONVERGENT CROSS-MAPPING

The recently proposed procedure of convergent cross-mapping (CCM) [Sug+12] checks for the existence of a topology preserving mapping $M^t_{i \to j}$ between two reconstructions r^{x_i} and r^{x_j} to infer the existence of a causal link $i \to j$. It is stated as follows:

For a given time point t one finds the set of k nearest neighbours

$$\{\boldsymbol{r}^{x_i}(t_1^{x_i}), ..., \boldsymbol{r}^{x_i}(t_k^{x_i})\}$$

to $\boldsymbol{r}^{x_i}(t)$. The time indices $\{t_1^{x_i}, ..., t_k^{x_i}\}$ of these points are used to determine the corresponding points

$$\{\boldsymbol{r}^{x_j}(t_1^{x_i}), ..., \boldsymbol{r}^{x_j}(t_k^{x_i})\}$$

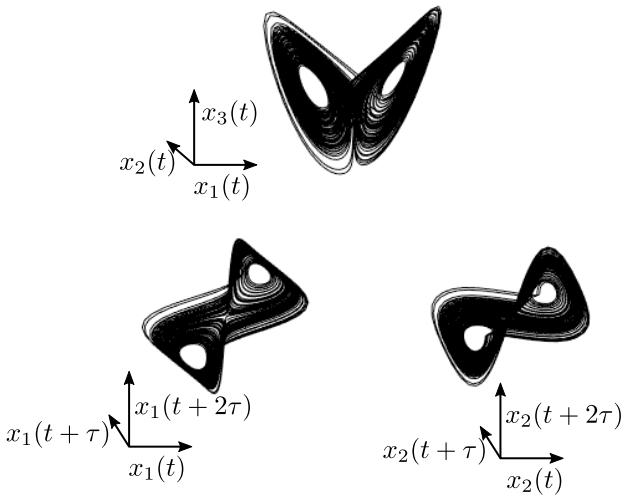

Figure 7.1: Illustration of Takens' theorem. The trajectory in the original phase space of a Lorenz system (eqs 7.2) (up) visits the distinct "butterfly" shape of the attractor. Trajectories in time lagged coordinate spaces of x_1 (down left) and x_2 (down right) reveal a topologically equivalent shape.

on $\boldsymbol{r}^{x_j}$. If a topology preserving mapping $M^t_{i\to j}$ exists, these latter points should lie close to $\boldsymbol{r}^{x_j}(t)$. To test this, a linear prediction

$$ {}^{e}\boldsymbol{r}^{x_j}(t) = \sum_{l=1}^{k} a_l \boldsymbol{r}^{x_j}(t_l^{x_i}) $$

is generated, where the coefficients a_l are monotonously decreasing functions of the distance $|\boldsymbol{r}^{x_i}(t) - \boldsymbol{r}^{x_i}(t_l^{x_i})|$. If now the number of available data points – or as the authors put it: the library length L – increases and the correlation $\rho_{j\to i}$ between predictions ${}^{e}\boldsymbol{r}^{x_j}(t)$ and true values $\boldsymbol{r}^{x_j}(t)$ converges to 1, a causal link from x_j to x_i is assumed. The authors also observe that stronger coupling seems to imply higher convergence speed [Sug+12].

The rationale is that if a mapping exists, projecting neighbourhood onto neighbourhood, the prediction of points on one reconstructions from neighbourhoods in the other should improve as the attractor "fills in" with data points.

Inferring the existence of a causal in link in the direction $j \to i$ the from the existence of a mapping $i \to j$ in the opposite direction might seem counter-intuitive at first. To clarify one might consider the following argument. If a coupling w_{ij} from x_j to x_i exists, then information about the trajectory of x_j is contained in x_i. This entails that $\boldsymbol{r}^{x_j}$ can be reconstructed from $\boldsymbol{r}^{x_i}$, i.e a mapping $M^t_{i\to j}$ exists.

8 TOPOLOGICAL CAUSALITY

8.1 INTRODUCTION

state dependency of effective influence

Part I of this thesis showed how modulations of effective influences in neuronal networks are crucial for selective information processing. Neuronal networks can be modelled as non-linear dynamical systems. The CTC mechanism is a good example of how this non-linearity allows for modulations of effective influences that would not be possible in linear systems: due to a non-linear activation function mapping membrane potential to firing rate, a neuron can be more or less receptive for incoming input. If the neuron's membrane potential is close to the firing threshold, an input might cause a spike which otherwise would not occur if the membrane potential is further away from threshold. Thus it is clear that the influence exerted across a synapse does not only depend on the strength of the synapse, but also on the state of the system. Or differently put, the effective influence is generated from the interplay of the system's structure with its dynamics.

analyses of non-linear systems often use unfit measures

This state dependency is a generic property of non-linear dynamical systems and is at odds with the assumptions of Granger causality (GC). However, many studies, also discussed previously in part I, that investigate influences between neuronal populations resort to linear stochastic measures such as GC (e.g. [Bas+15; Bos+12; Mic+16]), probably due to the lack of a viable alternative. A linear measure can be assumed to be reasonable if the causality is certain to be unidirectional, such as when the influence of an external signal to neuronal populations is investigated (Spectral coherence used in [Gro+15] and chapters 3 and 4). However, in the general case GC was shown to produce erroneous detection of influences from electrophysiological data [SP17a], which recently incited a vigorous debate [FSM17; BBS17; SP17b]. I want to address in the following how a suitable alternative measure for effective influence in generic non-linear dynamical systems

can be formulated that does not suffer from the shortcomings of applying a measure to a situation for which it is not designed.

The remainder of this chapter is adapted with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

"cause" and "effect" are generically entangled ...

Historically, influence between system components and the concept of causality are tightly connected. Discussions about the nature of causality range back to the beginnings of natural philosophy [AriBCa]. In recent formalizations it refers to situations where states $\boldsymbol{x}_1$ of one system part influence states $\boldsymbol{x}_2$ of another part [AP08]. It is often assumed that some aspects of $\boldsymbol{x}_1$ vary independently of $\boldsymbol{x}_2$, and that the flow of information in the overall system is essentially unidirectional. This premise is at odds with non-linear dynamical systems studied in e.g. physics, ecology, economy and neuroscience: generally, two system parts, e.g. two brain areas, will have bidirectional interaction and cyclic information flow [BWD12]. The classical notion of causality becomes problematic here since cause and effect are entangled.

... proven by Takens' theorem

This entanglement is reflected in Takens' theorem [Tak81; Pac+80], which proves that in generic smooth deterministic dynamical systems, the overall state is reconstructable from any measured observable. In other words, if $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ interact bidirectionally, each $\boldsymbol{x}_1(t)$ and $\boldsymbol{x}_2(t)$ alone contain the full information about the whole system constituted by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. That is, the system cannot be separated into subsystems and rather behaves as a whole. In consequence, the question for causal relations can not be answered by classifying component systems as "cause" or "effect", but rather asks for the asymmetry and strength of influence among these components.

Mathematically sound and agreed upon definitions of these quantities in dynamical systems are lacking. A novel approach tailored to entangled dynamical systems, introduced in the next section, provides the foundation for exactly that. As a key insight it was discovered that local distortions in the mappings between state space reconstructions based on different observables reflect the time or state dependent efficacy of causal links among the

underlying system components. Measures of effective influence derived from this relation are analytically accessible for simple systems and for more complicated ones can be estimated in a model free, data driven manner.

8.2 THEORY OF TOPOLOGICAL CAUSALITY

Let the (multidimensional) states $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of two system components be governed by

$$\begin{aligned}\dot{\boldsymbol{x}}_1 &= f_1(\boldsymbol{x}_1, w_{12}\mu_2(\boldsymbol{x}_2)) \\ \dot{\boldsymbol{x}}_2 &= f_2(\boldsymbol{x}_2, w_{21}\mu_1(\boldsymbol{x}_1)) \quad ,\end{aligned}$$

where $\mu_i(\boldsymbol{x}_i)$ denote fixed scalar functions and w_{ij} coupling constants. Takens' theorem states that a unique mapping $M_{i\to j}$ from reconstruction $\boldsymbol{r}^{x_i}$ to reconstruction $\boldsymbol{r}^{x_j}$ exists *iff* $w_{ij} \neq 0$ [Tak81; CGS15].

Assuming that these mappings are differentiable, $M^t_{i\to j}$ denotes the local linearisation (Jacobian Matrix) of $M_{i\to j}$ at the reference point t: Given that $\{t_1^{x_i}, ..., t_k^{x_i}\}$ are the time indices of the nearest neighbours on $\boldsymbol{r}^{x_i}$ to the reference point $\boldsymbol{r}^{x_i}(t)$, $M^t_{i\to j}$ is the linear approximation of the mapping which projects

$$\{\boldsymbol{r}^{x_i}\left(t_1^{x_i}\right), ..., \boldsymbol{r}^{x_i}\left(t_k^{x_i}\right)\} \quad \text{to} \quad \{\boldsymbol{r}^{x_j}\left(t_1^{x_i}\right), ..., \boldsymbol{r}^{x_j}\left(t_k^{x_i}\right)\} \quad .$$

In practice, the expansion $e^t_{i\to j}$ of $M^t_{i\to j}$ will be analysed, which is determined by the singular values $\sigma^t_k(M^t_{i\to j})$ of $M^t_{i\to j}$ larger than one:

$$e^t_{i\to j} = \prod_k \max\left(1, \sigma^t_k(M^t_{i\to j})\right) \quad .$$

expansions between reconstructions relate to the effective influence

To illustrate how the expansions of these mappings between reconstructions relate to directed effective influence, consider the following thought experiment: First, a system with unidirectional interaction is observed, i.e. $w_{21} \neq 0; w_{12} = 0$. By virtue of Takens' theorem, a unique mapping $M_{2\to 1}$ from reconstruction $\boldsymbol{r}^{x_2}$ to $\boldsymbol{r}^{x_1}$ exists. However, $M_{1\to 2}$ does not exist, since $\boldsymbol{x}_1$ has no information on $\boldsymbol{x}_2$. This is illustrated in figure 8.1 A) by a joint manifold $(\boldsymbol{r}^{x_1}, \boldsymbol{r}^{x_2})$ lying 'folded' over $\boldsymbol{r}^{x_1}$ but uniquely over $\boldsymbol{r}^{x_2}$. Note here

that maybe counter-intuitively the influence from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ is reflected in the 'backward' mapping $M_{2\to1}$: the existence of $M_{2\to1}$ implies coupling from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$.

Now increasing w_{12} while keeping $w_{21} > w_{12}$ leads to mutual but asymmetric interaction. Reconstructions $\boldsymbol{r}^{x_1}$ and $\boldsymbol{r}^{x_2}$ will both reveal the same global system state. However, the weaker coupling from x_2 to x_1 implies that the region of $\boldsymbol{r}^{x_2}$ states consistent with a small region of $\boldsymbol{r}^{x_1}$ states around $\boldsymbol{r}^{x_1}(t)$ is larger than vice versa at most reference points t: Both $\boldsymbol{r}^{x_1}$ and $\boldsymbol{r}^{x_2}$ are driven away from their state at time t by a combination of internal dynamics and the external influence from the other variable, but $\boldsymbol{r}^{x_2}$ is more so due to the stronger coupling w_{21}. This entails that $e^t_{1\to2} > e^t_{2\to1}$, and a joint manifold $(\boldsymbol{r}^{x_1}, \boldsymbol{r}^{x_2})$ lying uniquely over both reconstruction spaces, but more "steeply" over $\boldsymbol{r}^{x_1}$ (figure 8.1 B).

If w_{12} is now decreased again to approach zero, $e^t_{1\to2}$ will increase until it diverges at the point where $(\boldsymbol{r}^{x_1}, \boldsymbol{r}^{x_2})$ folds in on itself as seen from $\boldsymbol{r}^{x_1}$ (compare figure 8.1 A). This happens at $w_{12} = 0$, where the mapping $M^t_{1\to2}$ looses uniqueness and corresponding points to neighbours in $\boldsymbol{r}^{x_1}$ lie scattered over the whole dynamical range of $\boldsymbol{r}^{x_2}$. Thus infinite expansion equates to the non-existence of the corresponding mapping.

Consequently, when the couplings among $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ vanish altogether, both component systems will behave independently and the density of the resulting joint manifold factorizes. When observed from reference states $\boldsymbol{r}^{x_1}(t)$ and $\boldsymbol{r}^{x_2}(t)$, the mappings can be considered infinitely expanding in both directions, since for most reference points close neighbours correspond to distant points in the respective other space. This situation is illustrated in figure 8.1 C).

expansion invariance through quantile transformation

In these expositions it was assumed that the scalar observables $x_i = \phi_i(\boldsymbol{x}_i)$ have been transformed to their quantiles $q(x_i) = F(x_i)$ prior to time-delay embedding, where $F(x_i) = P[X_i \leq x_i]$ is the cumulative density function for the invariant measure of x_i. This eliminates expansions not caused by directed influences but arising from the numerical representation of the scalar time series or from the measurement function ϕ. Using the fact that smooth scalar injective transformations ϕ are either monotoni-

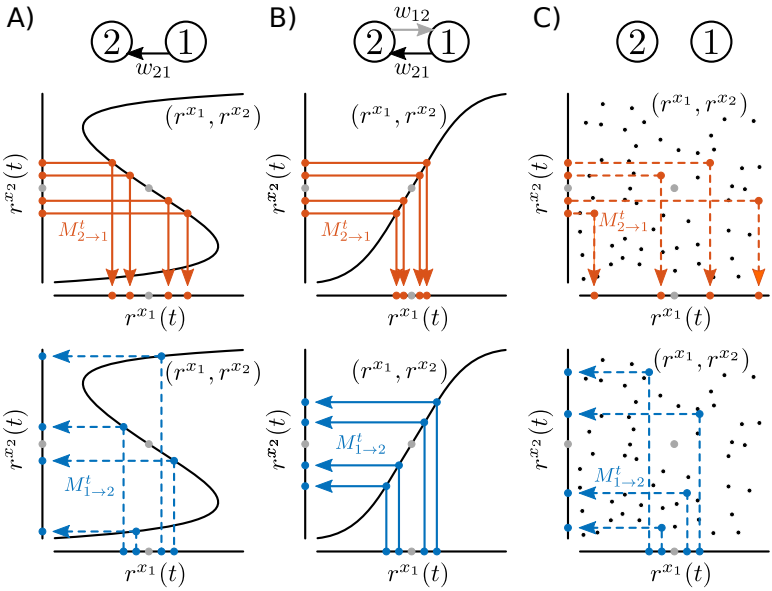

Figure 8.1: The relation of points $\boldsymbol{r}^{x_1}(t)$ and $\boldsymbol{r}^{x_2}(t)$ on multidimensional manifolds illustrated in 1-D. The joint manifold represented by $(\boldsymbol{r}^{x_1}, \boldsymbol{r}^{x_2})$ can be interpreted as the function mediating the mappings $M_{i\to j}$ between both spaces, and local linearisation $M^t_{i\to j}$ of the mappings as the slope around a reference point. **A)** When only $w_{21} \neq 0$, a one-to-one mapping $M_{2\to1}$ from $\boldsymbol{r}^{x_2}$ to $\boldsymbol{r}^{x_1}$ exists, but not in the reverse direction: $\boldsymbol{r}^{x_2}(t)$ is not uniquely determined for all states $\boldsymbol{r}^{x_1}(t)$. Locally, $M^t_{1\to2}$ can be attributed a diverging expansion property, since close neighbours of a given point $\boldsymbol{r}^{x_1}(t)$ correspond to distant parts of the joint density $(\boldsymbol{r}^{x_1}, \boldsymbol{r}^{x_2})$. The dashed lines visualize non-uniqueness. **B)** Here, both couplings are non-zero, but $w_{21} > w_{12}$. Larger independence of x_1 implies a stronger expansion by $M^t_{1\to2}$ than by $M^t_{2\to1}$ at most reference points, which is indicated by the higher slope of $(\boldsymbol{r}^{x_1}, \boldsymbol{r}^{x_2})$ when seen from $\boldsymbol{r}^{x_1}$. **C)** If no coupling exists, expansion diverges in both directions. Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

cally increasing or decreasing, they have no effect on the singular values of the Jacobian based on the quantile transformed observables. Thus, in the following, quantile transformations are routinely performed by kernel density estimations of the invariant densities $p(x_i)$ [BA97].

Following these topological considerations, I hypothesize that local expansions of the mappings between reconstruction manifolds of two observables can be utilized for graded measures of the directed causal effective influences between component systems represented by these observables, where $e^t_{i\to j}$ is inversely related to the strength of the effective influence $j \to i$.

a simple analytical example

The relation between expansion and effective influence can be made fully explicit with coupled one dimensional time-discrete maps for which two dimensional time-delay embeddings are sufficient. Consider the system given by

$$\begin{aligned} x_1(t+1) &= \overbrace{x_1(t)[r_1(1-x_1(t))}^{g_1(t)} - w_{12}x_2(t)] \\ x_2(t+1) &= \underbrace{x_2(t)[r_2(1-x_2(t))}_{g_2(t)} - w_{21}x_1(t)] \end{aligned} \tag{8.1}$$

which has served as a model of ecological systems [Sug+12]. Here, $M^t_{i\to j}$ based on raw data is accessible analytically for every point in state space.

The linearised mapping $W^t_{i\to j}$ of small perturbations around $(x_i(t), x_i(t+1))^T$ to $\mathbf{r}^{x_j}(t)$, such that

$$(\Delta x_j(t), \Delta x_j(t+1))^T = W^t_{i\to j}(\Delta x_i(t), \Delta x_i(t+1))^T$$

is

$$W^t_{i\to j} = \tfrac{1}{w_{ij}x_j(t)} \begin{pmatrix} g_i'(t) - w_{ij}x_j(t) & -1 \\ (g_i'(t) - w_{ij}x_i(t))(g_j'(t) - w_{ji}x_j(t)) - w_{ij}w_{ji}x_i(t)x_j(t) & -(g_j'(t) - w_{ji}x_i(t)) \end{pmatrix}$$

with $g_i(t)' = r_i(1-2x_i(t))$ being the derivative of the internal dynamics of x_i w.r.t. $x_i(t)$. This leads to one non-zero singular value

$$\sigma = \frac{1}{|w_{ij}x_j(t)|}\sqrt{([g_i'(t) - w_{ij}x_j(t)]^2 + 1)([g_j'(t) - w_{ji}x_i(t)]^2 + 1)}$$

The effect of quantile transformation is captured by modifying $W^t_{i\to j}$ to

$$M^t_{i\to j} = \begin{pmatrix} p(x_j(t)) & 0 \\ 0 & p(x_j(t+1)) \end{pmatrix} W^t_{i\to j} \begin{pmatrix} p(x_i(t))^{-1} & 0 \\ 0 & p(x_i(t+1))^{-1} \end{pmatrix}$$

with $p()$ denoting probability density. The non-zero singular value of $M^t_{i\to j}$ then becomes

$$\sigma^t\left(M^t_{i\to j}\right) = \frac{1}{|w_{ij}x_j(t)|}\chi \tag{8.2}$$

$$\chi = \frac{1}{|p(x_i(t))p(x_i(t+1))|} \sqrt{[p(x_j(t+1))(g_i'(t) - w_{ij}x_j(t))]^2 + p(x_j(t))^2} \sqrt{[p(x_i(t+1))(g_j'(t) - w_{ji}x_i(t))]^2 + p(x_i(t))^2}$$

Thus, the expansion depends explicitly, and implicitly via the derivations of the internal dynamics, on system states $x_i(t)$ and $x_j(t)$. For small w_{ij}, the factor χ converges to a fixed value, such that the whole expression is dominated by $1/|w_{ij}x_j(t)|$.

This example illustrates two important points. First, the weights are inversely proportional to the respective expansions and $\lim_{w_{ij}\to 0} e^t_{i\to j} = \infty$, in line with the heuristic considerations. And second, the effective influence will not only depend on the coupling weights, but also on the system's state: $M^t_{1\to 2}$ is strongly expansive for low x_1 values, and $M^t_{2\to 1}$ for low values of x_2, so that different regions in state space could be characterized by different influence dominance.

state dependent asymmetry of effective influences

To measure such state dependent asymmetry of effective influence let the index $-1 \leq \alpha^t \leq 1$ be defined as

$$\alpha^t = \frac{\log(e^t_{1\to 2}) - \log(e^t_{2\to 1})}{\log(e^t_{1\to 2}) + \log(e^t_{2\to 1})}$$

This definition is motivated by the relation of the log expansions to loss of certainty in information theoretical terms (further discussed in section 8.5.1). Note, however, that depending on the

particular system and the purpose of analysis other choices can be more useful. Figure 8.2 A) shows that the asymmetry index α^t in this example fluctuates considerably over time as the system explores the state space. This change of influence dominance gives rise to various dynamical regimes among the time courses of $\boldsymbol{r}^{x_1}$ and $\boldsymbol{r}^{x_2}$, which are also obvious to see in $x_1(t)$ and $x_2(t)$ proper since the dimensionality of the system is low (figure 8.2 B)). Specifically, when e.g. the influence from x_1 to x_2 is stronger than in the reverse direction (blue region), i.e. $e^t_{1\to2} > e^t_{2\to1}$, the trajectory of x_2 shows stronger fluctuations than the one of x_1.

Averaging over states visited during the dynamics yields a mean asymmetry index:

$$\alpha = \frac{\langle \log(e^t_{1\to2}) - \log(e^t_{2\to1})\rangle_t}{\langle \log(e^t_{1\to2}) + \log(e^t_{2\to1})\rangle_t} \quad .$$

For the case of bidirectional coupling, which enforces $M^t_{1\to2} = (M^t_{2\to1})^{-1}$, this expression simplifies to

$$\alpha = \frac{\langle \log(\det(M^t_{1\to2}(M^t_{1\to2})^T))\rangle_t}{2\sum_k \langle |\log(\sigma^t_k(M^t_{1\to2}))|\rangle_t} \quad ,$$

which underlines the dissimilarity to approaches using the logarithm of $|\det(M)|$ [Jan+12]. Figure 8.2 C) shows α for different combinations of coupling weights. The fact that $\alpha \neq 0$ for $w_{12} = w_{21}$ reflects the difference between the dynamical equations for x_1 and x_2 and highlights again that the expansion is not a mere proxy of the coupling weight, but actually measures the effective influence exerted along the causal link.

state dependent effective influence

However, the log determinant in α does not differentiate between the qualitatively different situations of balanced strong and balanced weak influence. For this purpose the topological causality (TC) is defined as

$$C^t_{i\to j} = \frac{1}{1 + \log\left(e^t_{j\to i}\right)}$$

$$C_{i\to j} = \frac{1}{1 + \langle \log\left(e^t_{j\to i}\right)\rangle_t} \quad .$$

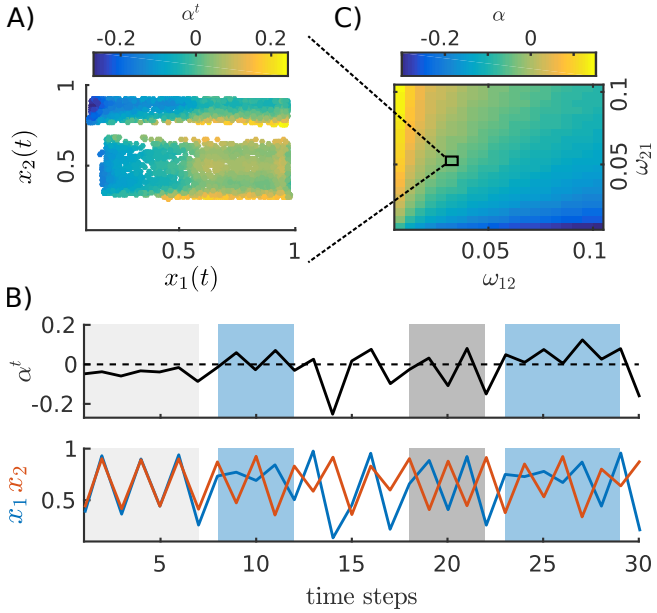

Figure 8.2: Example of state dependent asymmetry. **A)** The state space dependent α^t in a system given by equations (8.1) is shown for $w_{21} = 0.05; w_{12} = 0.02$ and $r_1 = 3.8; r_2 = 3.6$. **B)** A segment of α^t and the corresponding time courses of x_1 and x_2 for the same coupling weights. Different regimes of dominant influence direction give rise to different dynamical motifs. If α^t is close to 0 for subsequent time points (light grey), x_1 and x_2 synchronize. If α^t varies strongly around 0 (dark grey), x_1 and x_2 desynchronize. When the causal effective influence from one variable to the other is dominant, here from 1 to 2 (blue), the trajectory of x_2 shows higher amplitude excursions than the one of x_1. **C)** The mean asymmetry index α for the same system with varying coupling strengths. Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

$C \in [0,1]$ satisfies the following fundamental intuitions about causality: TC from component system i to j vanishes if no causal link exists ($w_{ji} = 0$), and for small couplings it is a monotonous function of the coupling weight. Also here alternative definitions

with the same properties are possible. In general, C and C^t depend on the coupling weights as well as on the current state of the system.

8.3 NUMERICAL METHODS

In cases where the dynamical system model does not allow for an analytical linearisation of the mappings between reconstructed spaces, or the model itself is not known, the local mappings and hence their expansion can be estimated in a purely data-driven manner. After optimal embedding parameters m and τ are estimated for the quantile-transformed time series, for which a range of methods exists [BK15a], the k neighbouring points

$$\{\boldsymbol{r}^{x_1}(t_1^{x_1}), ..., \boldsymbol{r}^{x_1}(t_k^{x_1})\}$$

to a reference point $\boldsymbol{r}^{x_1}(t)$ are found under exclusion of the temporal neighbours within $\tau(m-1)$. Estimations of the expansions are then obtained by two consecutive principal component analyses to increase numerical robustness.

A first principal component analysis is used to reduce the dimension of the set of points of nearest neighbours

$$\{\boldsymbol{r}^{x_1}(t_1^{x_1}), ..., \boldsymbol{r}^{x_1}(t_k^{x_1})\}$$

and the set of corresponding points

$$\{\boldsymbol{r}^{x_2}(t_1^{x_1}), ..., \boldsymbol{r}^{x_2}(t_k^{x_1})\}$$

individually: Let m_{pca}^1 and m_{pca}^2 be the number of components contributing 99% to the explained variance for each set of points and $m_{pca} = \max(m_{pca}^1, m_{pca}^2)$. Then $\boldsymbol{X}_1^{x_1}$ and $\boldsymbol{X}_2^{x_1}$ denote the $k \times m_{pca}$ matrices of representations of

$$\{\boldsymbol{r}^{x_1}(t_1^{x_1}), ..., \boldsymbol{r}^{x_1}(t_k^{x_1})\} \text{ and } \{\boldsymbol{r}^{x_2}(t_1^{x_1}), ..., \boldsymbol{r}^{x_2}(t_k^{x_1})\}$$

in the individual component spaces.

In a second step, a principle component analysis is performed on the combined $k \times 2m_{pca}$ reconstruction $[\boldsymbol{X}_1^{x_1}, \boldsymbol{X}_2^{x_1}]$. Let $\boldsymbol{P}$ denote the $2m_{pca} \times 2m_{pca}$ matrix containing the principal compo-

nent coefficients. $M^t_{2\to1}$ is then estimated by solving the equation system

$$M^t_{2\to1} = \underbrace{\begin{pmatrix} \boldsymbol{P}_{1,1} & \dots & \boldsymbol{P}_{1,m_{pca}} \\ \vdots & \vdots & \vdots \\ \boldsymbol{P}_{m_{pca},1} & \dots & \boldsymbol{P}_{m_{pca},m_{pca}} \end{pmatrix}}_{\boldsymbol{P}_{x_1}} \underbrace{\begin{pmatrix} \boldsymbol{P}_{m_{pca}+1,1} & \dots & \boldsymbol{P}_{m_{pca}+1,m_{pca}} \\ \vdots & \vdots & \vdots \\ \boldsymbol{P}_{2m_{pca},1} & \dots & \boldsymbol{P}_{2m_{pca},m_{pca}} \end{pmatrix}^{-1}}_{\boldsymbol{P}^{-1}_{x_2}}$$

where P_{x_i} contains the x_i projections of the m_{pca} components of $\boldsymbol{P}$ that carry the largest variance. The singular values, and thus the expansion can now be derived from $M^t_{2\to1}$, and $M^t_{1\to2}$ is found analogously.

For estimating a chance level the neighbours

$$\{\boldsymbol{r}^{x_1}(t^{x_1}_1), ..., \boldsymbol{r}^{x_1}(t^{x_1}_k)\}$$

in the initial reconstruction are mapped N times to a set of randomly chosen points

$$\{\boldsymbol{r}^{x_2}(t^{rand}_1), ..., \boldsymbol{r}^{x_2}(t^{rand}_k)\}$$

in the other reconstruction. The 5%-quantile of the expansions found in these N trials is taken as the chance-level expansion.

Note that the principal component analysis steps, which increase numerical stability, make the procedure look quite involved, whereas it essentially just estimates the linear mapping between two sets of data points.

8.4 APPLICATION EXAMPLES

time-resolved effective influence

As an example of a more complex case that is not analytically tractable consider a system of coupled Rössler equations [Rös76] described by

$$\begin{aligned} \dot{x}_i(t) &= -f_i y_i(t) - z_i(t) + \sum_j \Omega_{ij} \\ \dot{y}_i(t) &= f_i x_i(t) + 0.1 y_i(t) \\ \dot{z}_i(t) &= 0.1 + z_i(t)(x_i(t) - 14) \quad ; \quad i = 1, \ldots, n \end{aligned} \tag{8.3}$$

with coupling functions Ω_{ij}. If not stated otherwise, $\{f_1, f_2, f_3\}$ were set to $\{0.99, 0.85, 0.67\}$. As measurements from the individual systems $q(y_i)$ were used. Figure 8.3 A) and B) show causality measures for a bidirectionally coupled system ($n = 2$) with $\Omega_{ij} = w_{ij} z_j(t)$. When choosing the coupling function in this way, strong causal influence $i \to j$ is only expected if the driving z_i component deviates from 0. Both $\hat{\alpha}^t$ and $\hat{C}^t_{i\to j}$ capture this temporal structure, which is not obvious from the time courses of the used measurements y_1, y_2 (figure 8.3 C)).

causal transitivity is guaranteed

In order to serve as satisfactory definitions, the proposed indices must meet fundamental requirements of causality that can be demonstrated by examining simple network motifs. One prerequisite is transitivity, meaning that "if 1 influences 2 and 2 influences 3, then 1 influences 3". Since $M^t_{3\to1} = M^t_{2\to1} M^t_{3\to2}$, it can be shown that

$$\begin{aligned} C^t_{1\to3} &\geq C^t_{1\to2} C^t_{2\to3} \qquad \text{if} \quad w_{21} \neq 0 \wedge w_{32} \neq 0 \\ C^t_{1\to3} &= 0 \qquad\qquad\qquad\quad \text{else} \end{aligned}$$

meaning that transitivity is mathematically guaranteed. Figure 8.4 A) shows numerically estimated $\hat{C}_{1\to3}$ for a system of 3 coupled Rössler equations. Only w_{21} and w_{32} were varied and other coupling weights fixed to zero.

shared input is not confused with true interaction

Another required property is the ability to distinguish shared input from true interaction. This is formally guaranteed if the receiving systems both retain independent degrees of freedom, i.e. do not synchronize completely. To show that also estimated TC

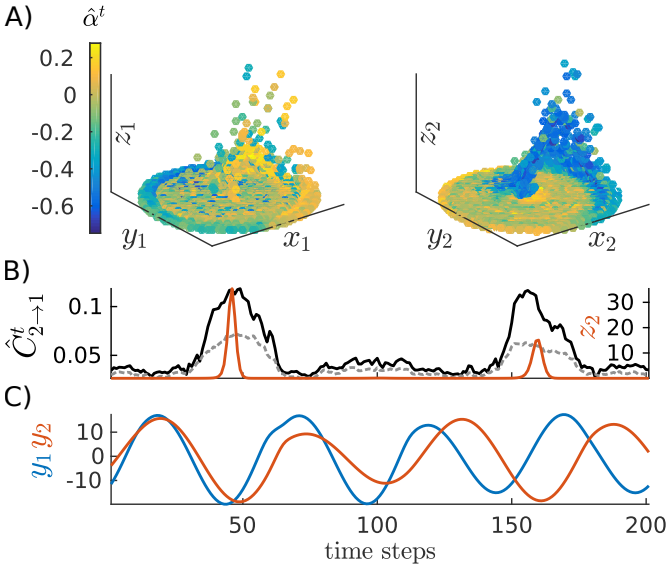

Figure 8.3: Example of time-dependent influences. Two bidirectionally coupled Rössler systems (equations 8.3) with $w_{21} = 0.2; w_{12} = 0.6$ and $\Omega_{ij} = w_{ij} z_j(t)$. The time series of 10^5 data points were embedded with dimension $m = 13$. $\hat{\alpha}^t$ and $\hat{C}^t_{2\to1}$ are shifted to be aligned with the temporal mean $t + 1/2(m-1)\tau$ of the corresponding reconstructions $\boldsymbol{r}^{x_i}(t)$. A) Local asymmetry $\hat{\alpha}^t$ of 10^4 points shown on projections of the attractor to each system. B) $\hat{C}^t_{2\to1}$ (black) for 200 consecutive time steps and the corresponding time series of z_2 (orange). The grey dashed line marks chance level. C) Time series y_1 and y_2 used to estimate $\hat{C}^t_{2\to1}$. Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

reflects this property, consider a system described by equations (8.3) ($n = 3$) with coupling functions $\Omega_{ij} = w_{ij}(x_j(t) - x_i(t))$, where only $w_{13} \neq 0$ and $w_{23} \neq 0$, generating a divergent network motif. Here, $\{f_1, f_2, f_3\}$ were set to $\{0.99, 0.97, 0.98\}$. With increasing coupling from x_3 to x_1 and x_2, the latter two synchronize more strongly, masking the actually absent interaction. Fig 8.4 B) shows that $\hat{C}_{1\to2}$ is nearly independent of the common

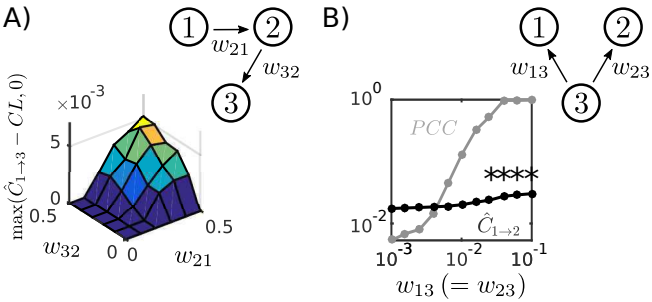

Figure 8.4: Causality properties. **A)** Transitivity. A unidirectionally coupled chain is realized by a system of equations (8.3) ($n = 3$) with only $w_{21} \neq 0$ and $w_{32} \neq 0$ and $\Omega_{ij} = w_{ij}x_j(t)$. $\hat{C}$ was averaged from 10^3 randomly selected data points from time series of 10^5 points with an embedding dimension $m = 13$. A bivariate increase of the excess of $\hat{C}_{1\to3}$ over chance level (CL) is observed, consistent with the theoretical prediction. **B)** Common input investigated in a system of equations (8.3) ($n = 3$) with $\Omega_{ij} = w_{ij}(x_j(t) - x_i(t))$ and only $w_{13} = w_{23} \neq 0$. $\hat{C}_{i\to j}$ is estimated with an embedding dimension $m = 7$. $\hat{C}_{1\to2}$ and $\hat{C}_{2\to1}$ (not shown) depend weakly on the common input and only become significant (marked by $*$) in the presence of high redundancy between 1 and 2, signified by a high Pearson Correlation Coefficient (PCC). Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

drive, and becomes significant only in the presence of substantial correlations.

application to experimental data

To demonstrate the applicability to real experimental time series, EEG data [Nun+99] was analysed for which a predominant information flow from frontal to dorsal channels was identified [Nol+08]. The data is freely accessible at `http://clopinet.com/causality/data/nolte/`. It contains EEG recordings from 10 subjects with 19 electrodes per subject at a sampling rate of 256 Hz, with each recording lasting around 1 minute. The average asymmetry $\hat{\alpha}$ was calculated for every channel pair within each subject. The data was low pass filtered with a stop-band at 3 Hz to remove slow drifts. Embedding parameters optimally represent-

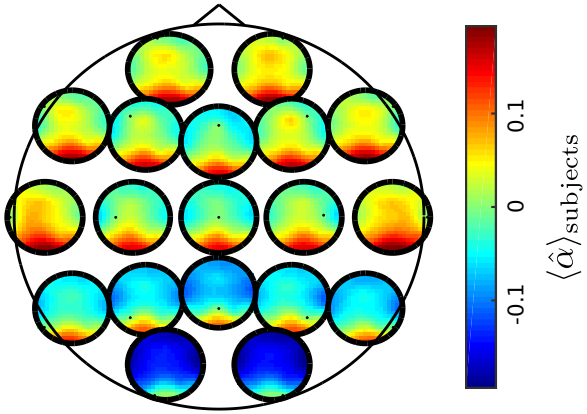

Figure 8.5: TC applied to EEG data. $\hat{\alpha}$ is shown between all possible pairs of 19 channels of the EEG recording, averaged over 10 subjects. The asymmetry from one channel to all others is plotted within the small circles, where the location of the target corresponds to the position of the data point in the small circles. The small circles are positioned at the approximate location of the corresponding electrode on the skull. Plotting routine adapted from [Nol+08]. Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

ing the system [BK15a] were found to be $m = 6; \tau = 6$. Figure 8.5 shows the $\hat{\alpha}$ averaged over subjects. The circles are positioned at the approximate electrode locations on the skull, where the values plotted inside the circles show $\hat{\alpha}$ with respect to the electrode. A predominant influence from frontal to dorsal electrodes is apparent, consistent with previous findings on the same dataset (compare figure 4 in [Nol+08]).

8.5 FURTHER PROPERTIES OF TOPOLOGICAL CAUSALITY

Since TC and the derived measures of effective influence are completely novel, the following section will go deeper into some important properties and considerations for possible applications. First, a connection with information theory will be established

which supports the choice of the expansion as the central measure of TC. Next, the robustness of estimated effective influences is investigated to variations of the embedding parameters and the available length of the time series. The relationship between TC and the related measure of CCM is also discussed, and finally TC, CCM and GC applied to a test linear stochastic system to assess the performance of TC and CCM, tailored to non-linear deterministic dynamical systems, under these circumstances.

8.5.1 *Connection with information theory*

The expansion of the mappings between reconstructions captured by TC is directly related to loss of certainty between states measured with finite precision. More explicitly: let there be two random variables X, Y defined on $\mathcal{X}, \mathcal{Y} \subset [0,1]^N$. Let these random variables be related by the bijective map $\Phi : \mathcal{X} \mapsto \mathcal{Y}$. In this situation information loss only arises between discretizations of X and Y that model measurements with finite precision.

For simplicity let the discretizations X^n, Y^n be uniform, corresponding to hypercubes of side length 2^{-n}. Of interest is now the uncertainty about measurements of Y that are consistent with a measurement of X. A measurement $X^n = x^n$ corresponds to a region (hypercube) of consistent states of X, whose image under Φ gives rise to possible observations of Y^n that form the support of the conditional distribution of Y^n given the realization x^n, $\operatorname{supp} P_{Y^n|x^n} =: \mathcal{Y}^n_{|x^n}$. This leads to

$$\begin{aligned} H(Y^n|X^n) &= \mathbb{E}_{X^n}[H(Y^n|X^n = x^n)] \\ &= -\sum_{x^n \in \mathcal{X}^N} p(x^n) \sum_{y^n \in \mathcal{Y}^n_{|x^n}} p(y^n|x^n) \log p(y^n|x^n) \\ &\leq \sum_{x^n \in \mathcal{X}^N} p(x^n) \log |\mathcal{Y}^n_{|x^n}| = \tilde{H}(Y^n|X^n) \quad , \end{aligned}$$

where $\tilde{H}$ denotes the entropy with respect to $\tilde{p}_{Y^n|x^n} \sim \mathrm{UNI}(\mathcal{Y}^n_{|x^n})$, i.e the consistent states in Y^n are equiprobable. The inequality is

tight. An elementary calculation gives $\log |\mathcal{Y}^n_{|x^n}| \sim e_{x\to y}(x)$ in the limit $n \to \infty$, so that

$$\lim_{n\to\infty} H(Y^n|X^n) \leq \lim_{n\to\infty} \tilde{H}(Y^n|X^n) \approx e_{X\to Y} \,.$$

This result shows that after introduction of a discretization it is *not* the log determinant (used in [Jan+12]) but the log expansion that relates to conditional entropy. The intuition is that one cannot recover information that is lost by the discretization along axes corresponding to singular values smaller than one. This is illustrated in figure 8.6. The log determinant relates to the difference $H(Y^n|X^n) - H(X^n|Y^n)$ and thus only measures the asymmetry of information loss.

8.5.2 *Dependency of TC on embedding parameters and neighbourhood size*

To demonstrate the robustness of estimated TC measures to variations in the embedding parameters m andτ, a Rössler system (equations 8.3, $n = 2$, $w_{12} = 0.2; w_{21} = 0.05$ and $\Omega_{ij} = w_{ij}x_j(t)$) is simulated for 10^5 time steps. $\hat{C}_{1\to2}$ and $\hat{C}_{2\to1}$ are calculated from 1000 randomly selected points with $k = 20$. Whereas the absolute values of $\hat{C}$ can vary, figure 8.7 shows that the correct relation $\hat{C}_{1\to2} > \hat{C}_{2\to1}$ is found even when the embedding parameters m, τ deviate strongly from the optimal parameters $m = 5; \tau = 17$. Crucially, instances of $\hat{C}_{1\to2} < \hat{C}_{2\to1}$ are not found throughout the parameter space. The same holds true when k is varied. Figure 8.8 A) shows that over a wide range of neighbourhood sizes, the relation between $\hat{C}_{1\to2}$ and $\hat{C}_{2\to1}$ is preserved. However, the bigger k, the more the temporal resolution in $\hat{C}^t_{i\to j}$ is lost (trivial, not shown).

It was also investigated how the time series length T influences $\hat{C}_{i\to j}$. The findings are shown in figure 8.8 B). $\hat{C}_{i\to j}$ tends to increase the sparser the time series get, which is reasonable since the bigger the subspace formed by the nearest neighbours around $\boldsymbol{r}^{x_j}(t)$ is in relation to the full manifold reconstruction, the smaller the maximal possible $e_{j\to i}(t)$ will be. However, the relative effective interaction is still qualitatively correctly detected

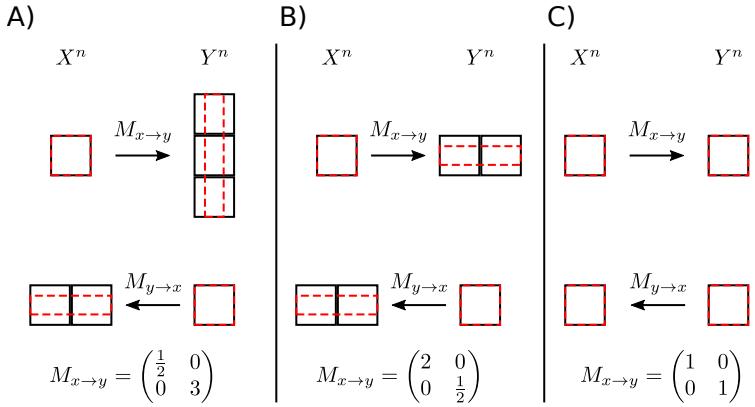

Figure 8.6: Expansion determines information loss. Starting from an observed value in X^n (or Y^n) the region of consistent X (Y) states are shown in dashed red rectangles and are transformed with the indicated mappings. The discretized measurements that can arise from this region are shown in black and the number of possible values is given by the expansion of the mappings. **A)** $M_{x\to y}$ is more expanding in the one direction than shrinking in the other. For one X^n value, there are $e_{x\to y} = 3$ possible values of Y^n, while only $e_{y\to x} = 2$ values of X^n are possible when starting with one value of Y^n. This *asymmetry* of information loss is contained in the determinant $|M_{x\to y}| = \frac{3}{2} > 1$. **B)** and **C)** demonstrate that situations with the same value for the determinant of the Jacobian behave qualitatively different. In **B)** there is the same amount of information lost when projecting from X^n to Y^n and from Y^n to X^n, whereas no information is lost at all in **C)**. Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

even for very short time series. For practical applications figure 8.8 B) suggests that reliable values for the expansions are obtainable when they stop to change with increasing length of the time series. For the present example this would correspond to about $T = 3000$ (Note that T is the length of the actual time series prior to embedding).

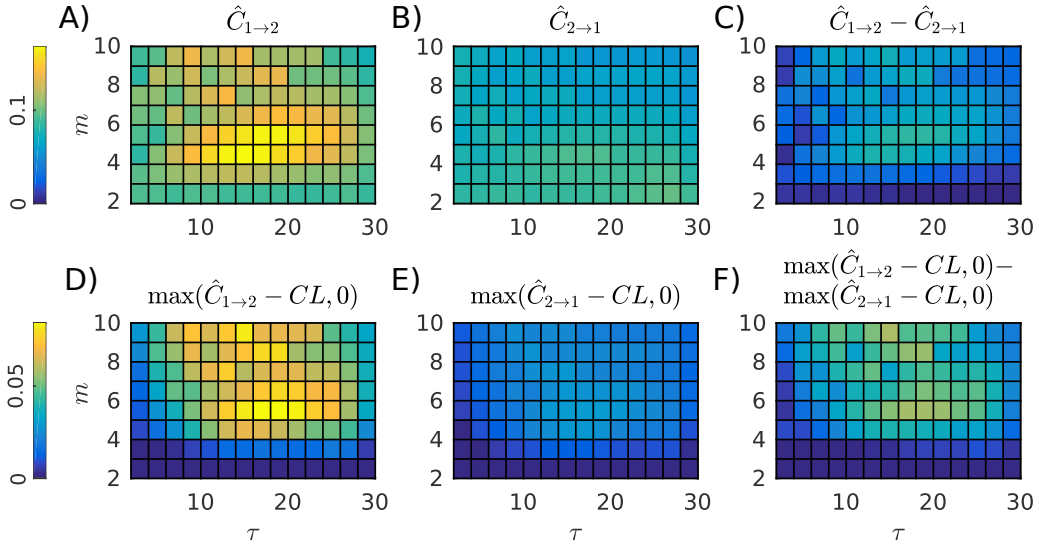

Figure 8.7: Robustness of TC to changes in the embedding dimension and time delay. **A), B), C)** show $\hat{C}_{1\to2}$, $\hat{C}_{2\to1}$, and $\hat{C}_{1\to2} - \hat{C}_{2\to1}$ for a Rössler system given by equations 8.3 ($n = 2$) with $w_{12} = 0.2; w_{21} = 0.05$ and $\Omega_{ij} = w_{ij}x_j(t)$. The system was simulated for 10^5 time steps and $\hat{C}$ was calculated from 1000 randomly selected points with $k = 20$. The colour bar on the left applies to all subplots. **D), E), F)** Same as A) B) C), but for the excess over chance level (CL). Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

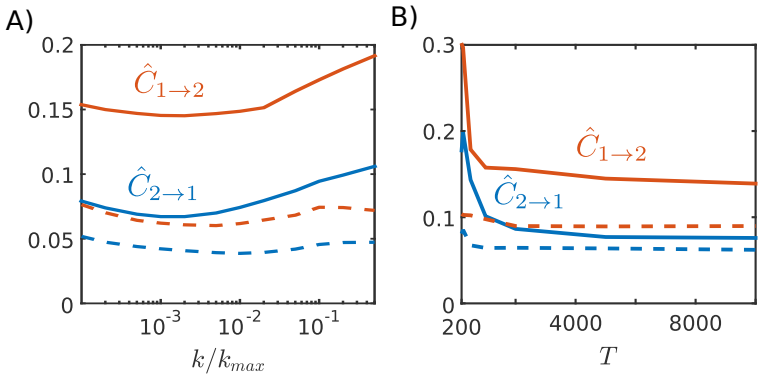

Figure 8.8: Robustness of TC to variations in time series length and neighbourhood size. **A)** $\hat{C}_{1\to2}$ and $\hat{C}_{2\to1}$ averaged over 1000 random reference points for the same Rössler system as in figure 8.7. k is varied while m, τ are set to their optimal values $m = 5; \tau = 17$. The orange (blue) dashed line shows the chance level for $\hat{C}_{1\to2}$ ($\hat{C}_{2\to1}$). k_{max} equals the full length of the embedded time series, i.e. $10^5 - \tau(m-1)$. **B)** $\hat{C}_{1\to2}$ and $\hat{C}_{2\to1}$ for the same Rössler system with increasing time series length T with fixed $m = 5; \tau = 17; k = 10$, averaged over $\min(2000, T - \tau(m-1))$ random reference points. Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

8.5.3 *Relation between TC and CCM*

The recently proposed procedure of convergent cross-mapping (CCM) evaluates the correlation $\rho_{i\to j}$ between a linear prediction of $\boldsymbol{r}^{x_j}(t)$ based on the set of points $\{\boldsymbol{r}^{x_j}(t_1^{x_i}), ..., \boldsymbol{r}^{x_j}(t_k^{x_i})\}$ which correspond to the k nearest neighbours to $\boldsymbol{r}^{x_i}(t)$ on $\boldsymbol{r}^{x_i}$ and the true $\boldsymbol{r}^{x_j}(t)$. If the correlation converges when increasing the number of available data points (called the library length) L, a causal link from $j \to i$ is assumed (see also section 7.3). The procedure is related to TC in that it also draws on Takens' theorem and the existence of topology preserving mappings between reconstructions. The authors of the original paper also observe that stronger coupling seems to imply higher convergence speed [Sug+12].

In relation to TC, it is first noted that there is no clear correlation between prediction error and expansion for a fixed library length L. For this consider the case that $\{\boldsymbol{r}^{x_j}(t_1^{x_i}), ..., \boldsymbol{r}^{x_j}(t_k^{x_i})\}$ lie on a quasi-linear subspace around $\boldsymbol{r}^{x_j}(t)$. It is now possible to shift the points $\{\boldsymbol{r}^{x_j}(t_1^{x_i}), ..., \boldsymbol{r}^{x_j}(t_k^{x_i})\}$ in this quasi-linear subspace such that they change their distance to $\boldsymbol{r}^{x_j}(t)$, which corresponds to changing the expansion of $M^t_{i\to j}$, while retaining the same linear prediction of $\boldsymbol{r}^{x_j}(t)$ (see figure 8.9).

However, I expect a monotonous relationship between the convergence speed of ρ as a function of L and the expansion, since the more expansive a mapping is, the more data points are needed such that $\{\boldsymbol{r}^{x_j}(t_1^{x_i}), ..., \boldsymbol{r}^{x_j}(t_k^{x_i})\}$ lie close enough around $\boldsymbol{r}^{x_j}(t)$ to allow for a good linear prediction. This convergence should be observable even when the embedded system has extended linear subspaces if very small library lengths are considered and the prediction errors are averaged over all observed system states and randomized libraries, which is the standard procedure in CCM.

This is tested in four simple systems of coupled maps for which the analytical expansions are available, namely coupled logistic maps with multiplicative interaction

$$\begin{aligned} x_1(t+1) &= x_1(t)[3.8(1-x_1(t)) - w_{12}x_2(t)] \\ x_2(t+1) &= x_2(t)[3.8(1-x_2(t)) - w_{21}x_1(t)] \quad , \end{aligned} \tag{8.4}$$

coupled logistic maps with additive interaction

$$\begin{aligned} x_1(t+1) &= (1-w_{12})x_1(t)4(1-x_1(t)) + w_{12}x_2(t) \\ x_2(t+1) &= (1-w_{21})x_2(t)4(1-x_2(t)) + w_{21}x_1(t) \quad , \end{aligned} \tag{8.5}$$

coupled shift maps

$$\begin{aligned} x_1(t+1) &= (1-w_{12})\mathrm{mod}(2x_1(t),1) + w_{12}x_2(t) \\ x_2(t+1) &= (1-w_{21})\mathrm{mod}(2x_2(t),1) + w_{21}x_1(t) \quad , \end{aligned} \tag{8.6}$$

and coupled tent maps

$$\begin{aligned} x_1(t+1) &= (1-w_{12})(1-2|x_1(t)-1/2|) + w_{12}x_2(t) \\ x_2(t+1) &= (1-w_{21})(1-2|x_2(t)-1/2|) + w_{21}x_1(t) \quad . \end{aligned} \tag{8.7}$$

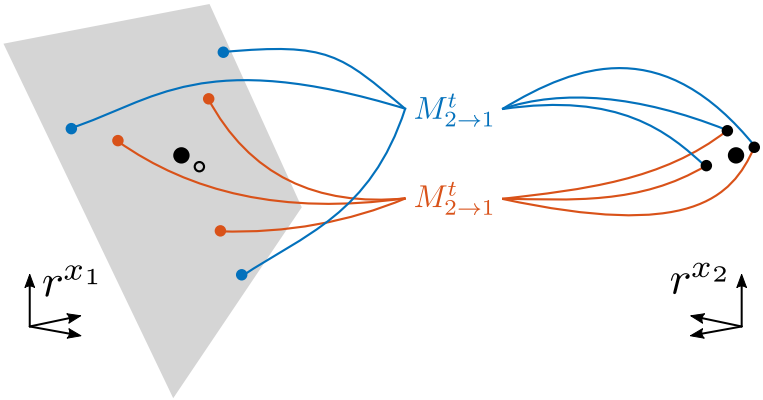

Figure 8.9: Illustration that different expansions $e^t_{i\to j}$ can lead to the same linear prediction of $\boldsymbol{r}^{x_j}(t)$. Assume that the points $\boldsymbol{r}^{x_1}(t)$ and $\boldsymbol{r}^{x_2}(t)$ (large black dots) and the nearest neighbours $\{\boldsymbol{r}^{x_2}(t_1^{x_2}), ..., \boldsymbol{r}^{x_2}(t_k^{x_2})\}$ to $\boldsymbol{r}^{x_2}(t)$ (small black dots) are visited by systems with different expansion. It is clear that different expansions, corresponding to different mappings $M^t_{2\to1}$ (blue: more expansive, orange: less expansive) can lead to the same linear prediction of $\boldsymbol{r}^{x_1}(t)$ (black circle) as long as they lie in the quasi-linear surround of $\boldsymbol{r}^{x_1}(t)$ (grey area). Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

For all maps an embedding dimension of 2 is sufficient, and the analytical solution can be found by simple algebraic manipulation analogous to the procedure sketched in section 8.2. To obtain the convergence curves $\rho_{i\to j}(L)$, the CCM toolbox was used provided by the authors of [Sug+12] available at

`https://cran.r-project.org/web/packages/rEDM/vignettes/rEDM_tutorial.html`.

The convergence curves $\rho_{i\to j}(L)$ were fit to $1 - a\exp(-\theta_{i\to j}L)$ to obtain the convergence time constants $\theta_{i\to j}$. Figure 8.10 shows analytical and estimated expansions and convergence time constants for all four systems while keeping one coupling weight fixed and varying the other. Indeed, higher convergence speed seems to correlate closely with smaller expansion.

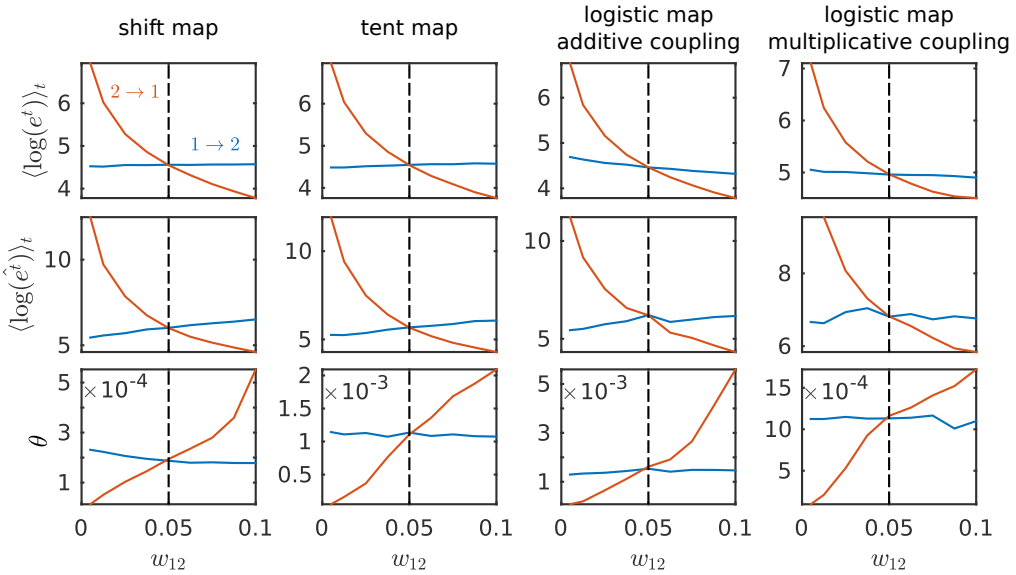

Figure 8.10: Comparison of TC and CCM. Different systems given by equations 8.4, 8.5, 8.6, and 8.7 were simulated for 10^4 time steps with $w_{21} = 0.05$ and varying w_{12}. The first row shows analytical results for the average log expansion, the second row estimated expansions with neighbourhood size $k = 5$, both averaged over all available states, and the third the time constants θ derived from CCM. The convergence curves that underlie the estimation of θ were computed with library lengths ranging between 5 and $10^4 - \tau(m-1)$, where for every length 10 randomized libraries were drawn and averaged. For all embeddings $m = 2$ and $\tau = 1$ was used. Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

To compare an average asymmetry measure derived from the convergence time constants θ, a slightly different definition of α is used than the one stated before:

$$\alpha_{TC} = \frac{\exp(\langle \log(e^t_{i\to j})\rangle_t) - \exp(\langle \log(e^t_{j\to i})\rangle_t)}{\exp(\langle \log(e^t_{i\to j})\rangle_t) + \exp(\langle \log(e^t_{j\to i})\rangle_t)} .$$

This choice was made since it yields good agreement with

$$\alpha_{CCM} = \frac{\theta_{j\to i} - \theta_{i\to j}}{\theta_{j\to i} + \theta_{i\to j}} \tag{8.8}$$

Note that the main quantity of TC is the expansion, such that the exact definitions of the derived measures α, α^t, C, C^t allow for some freedom as long as they are bounded correctly and monotonously depend on changes in the underlying expansions. Here the choice made is sensible since the analytical solutions of the expansion for systems given by equations 8.5, 8.6, and 8.7 can be written in a form

$$e^t_{i\to j} = \frac{1}{w_{ij}} \chi(w_{ij}, w_{ji}, x_i(t), x_j(t))$$

(compare equations 8.2), where χ is symmetric in i and j. It is thus to be expected that $\exp(\langle \log(e^t_{1\to 2})\rangle_t)$ and $\exp(\langle \log(e^t_{2\to 1})\rangle_t)$ are directly proportional to the coupling weights w_{12} and w_{21} with the same proportionality constant that stems for the temporal averaging of χ. This is why, as a side note, these values are also compared to an asymmetry measure derived form the coupling weights as

$$\alpha_w = \frac{w_{ji} - w_{ij}}{w_{ji} + w_{ij}}$$

Note that the expected proportionality of expansions to coupling weights is not the generic case, but due to the linear interaction terms in these test systems. Figure 8.11 shows that all different asymmetry measures correlate strongly. This leads to the conclusion that the relative convergence times based on CCM are actually good estimators of the relative average expansions.

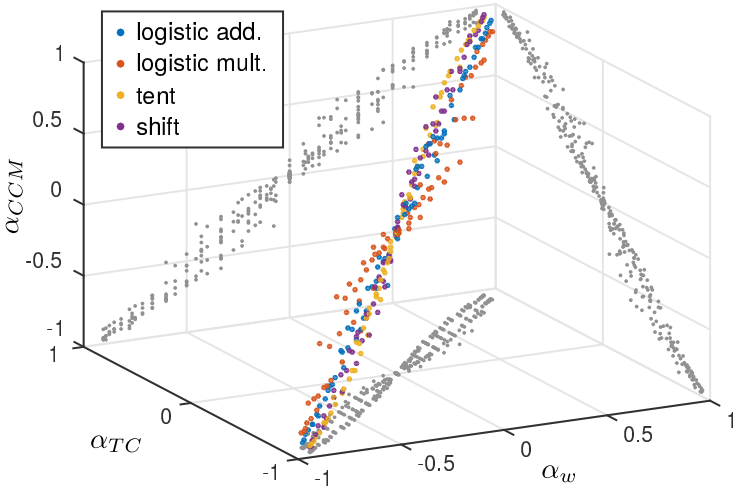

Figure 8.11: Comparison of asymmetry indices based on analytical expansion (α_{TC}), time constants from CCM (α_{CCM}) and coupling weights (α_w) for systems given by equations 8.4, 8.5, 8.6, and 8.7, simulated for 10^4 time steps for various combinations of w_{21} and w_{21}, each ranging between 0.005 and 0.1. The different colours show the results for different systems, where the grey dots are the projections of the whole point cloud to the relative subspaces. Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

8.5.4 *Comparing TC to GC and CCM for linear systems*

After this general comparison, the performance of TC and CCM is also assessed in the context of linear stochastic systems, for which Granger causality (GC) is the theoretically sound method of choice. This is of general importance if measured systems are noisy, are only weakly non-linear, or have piecewise linear regimes. As an example for a linear stochastic dynamical system a vector autoregressive model is considered, described by

$$\begin{aligned} x_1(t+1) &= \delta[(1-w_{12})x_1(t) + w_{12}x_2(t)] + \eta_1(t) \\ x_2(t+1) &= \delta[(1-w_{21})x_2(t) + w_{21}x_1(t)] + \eta_2(t) \quad , \end{aligned} \tag{8.9}$$

where w_{12} and w_{21} are coupling parameters, $\delta < 1$ is a dampening constant to keep the time series stationary, and $\eta_i(t)$ is spatially and temporally uncorrelated Gaussian noise $\mathcal{N}(\mathbf{0}, \sigma_i)$. A measure for asymmetry based on GC is introduced as

$$\alpha_{GC} = \frac{C^G_{2\to1} - C^G_{1\to2}}{C^G_{2\to1} + C^G_{1\to2}} \quad ,$$

with $C^G_{i\to j}$ as defined in section 7.1, equation 7.1. Figure 8.12 shows comparisons of α_{GC} with α_{TC}, α_w and α_{CCM} (section 8.5.3, equation 8.8) for simulations with $\delta = 0.95$, $\sigma_1 = \sigma_2 = 0.1$ and different combinations of coupling weights $0.005 < w_{ij} < 1.0$.

Whereas - as expected - GC captures the coupling asymmetry in the model (figure 8.12 C), so does TC (figure 8.12 A). Analytical results (black dots) are calculated by constructing the perturbation matrix from equations 8.9 under the assumption that the noise terms do not contribute (i.e. using the expected mapping). Estimated results (blue dots) are obtained by using the maximally possible neighbourhood size since the reconstructions are purely linear here.

Figure 8.12 B) shows the results when using the second and third most significant component of $\boldsymbol{P}$ instead of the first two (see section 8.3), which gives results closer to the theoretical prediction, most likely because the first component carries the autocorrelation of the variables. This hints toward possible improvement of the estimation algorithm for the expansions in linear stochastic systems. But, importantly, this example shows that TC does not yield misleading results when the system is stochastic and linear.

CCM fails to detect the causal relations (figure 8.12 D)), which is due to the fact that the $\rho_{i\to j}(L)$ curves in this case often show no clear convergence. This can be understood by considering that, no matter which library size is chosen, the points to predict $\boldsymbol{r}^{x_j}(t)$ from and the true $\boldsymbol{r}^{x_j}(t)$ lie on a (noisy) linear space in $\boldsymbol{r}^{x_j}$ (compare figure 8.9). For fitting $\rho_{i\to j}(L)$ library lengths L between 5 and 100 were used, since $\rho_{i\to j}$ only changes strongly in this range. The curves $\rho_{i\to j}(L)$ were then fit to the model $b - a\exp(-\theta_{i\to j}L)$, with b being the average of $\rho_{i\to j}(L)$ for $50 < L < 100$ to account

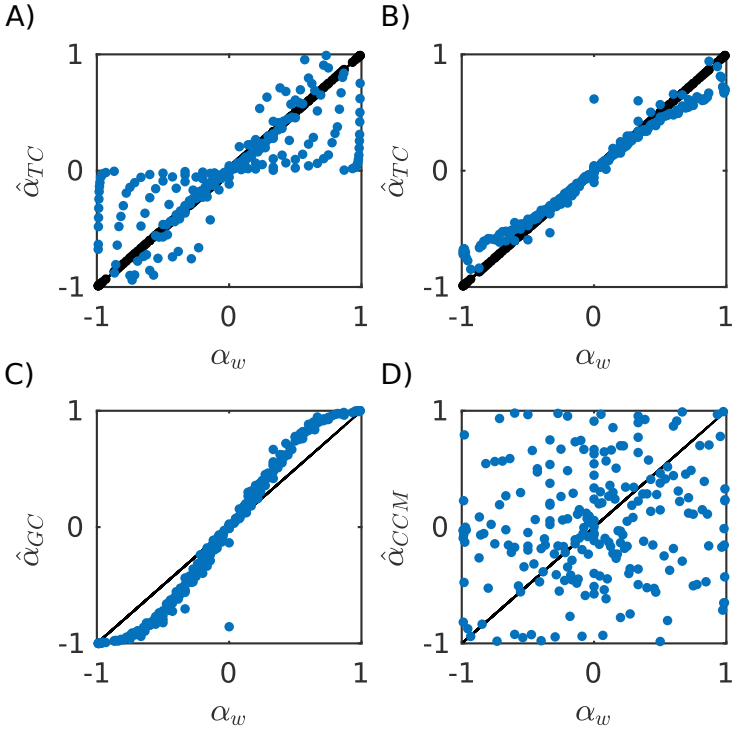

Figure 8.12: Mean asymmetry for three causality measures, TC (A,B), GC (C) and CCM (D), derived for time series from equations 8.9 of length 10^4 with varying coupling parameters w_{12} and w_{21} between 0.005 and 1.0 and $\delta = 0.95$. The time series were transformed to their quantiles before further analysis. The estimated values are plotted against $\alpha_w = \frac{w_{ji}-w_{ij}}{w_{ji}+w_{ij}}$ **A)** For estimation, the maximum neighbourhood size $k = 10^4 - \tau(m-1)$ and embedding parameters $\tau = 1$ and $m = 2$ are used (blue). Black dots show the analytical prediction. **B)** Same as A), but using the second and third most significant component of $\boldsymbol{P}$ instead of the first two. **C)** GC is calculated with $p = 1$. **D)** The same embedding parameters as for TC, $\tau = 1$ and $m = 2$, are used and the convergence curves were computed with library lengths L ranging between 5 and 100. Reprinted figure with permission from D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems". Physical Review Letters 119.9, p.098301 (2017). Copyright 2017 by the American Physical Society.

for the strongly varying asymptotic values of $\rho_{i\to j}$ across the coupling weight combinations in this example.

A comparison between GC and TC for non-linear systems was not conducted since the basic premise of separability that GC theoretically needs would be violated. Hence there is ample evidence of classical GC failing in non-linear systems (e.g. [SP17a; Sug+12]).

8.6 DISCUSSION

TC provides mathematically sound definition of effective influence

In this chapter, the theory of topological causality (TC) was developed, which provides a mathematically well defined framework to measure effective influence in generic deterministic dynamical systems. It is based on the central insight that expansions of mappings between time-delay state space reconstructions from different observables systematically reflect effective, state space dependent influences among parts of non-separable deterministic dynamical systems. Measures of effective influence were presented analytically in simple examples, and showcased numerically for more complex mathematical models and experimental EEG data. In further investigations, a link to information theoretical approaches was discovered, and applicability also to predominantly stochastic systems.

TC works for linear stochastic systems

Being tailored systems which preserve information between coupled components, TC seems complementary to methods for determining causal influences in stochastic systems. Most prominent examples are Granger causality (GC) [Gra69] and transfer entropy (TE) [Scho00], which are conceptually related [BBS09]. Both are based on the reduction of uncertainty in one time series by including past information from the other. However, the approaches of GC/TE and TC are not independent: In stochastic linear systems, the observed dynamics in a time delay coordinate space can be interpreted as samples from a probability density of consecutive sequences of length m (if $1/\tau$ equals the sampling rate) in the observables. And an expansive mapping between probability densities induces information loss. Formally, expansion can be directly related to loss of certainty between states measured with finite precision. In stark contrast to the usual applications of GC/TE, TC exploits the expansion of the "*backward*" mapping from "effect" $\boldsymbol{x}_2$

to "cause" $\boldsymbol{x}_1$ for determining the effective influence from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$. In a preliminary investigation it was observed that TC can indeed detect effective influences in predominantly stochastic systems, raising the intriguing possibility that it is well suited for both deterministic and stochastic systems (figure 8.12). Supporting evidence to analyse the backward mapping in stochastic systems is given by [Hau+12], who apply GC in this way and find similar results compared to the classical GC.

TC encompasses and extends previous measures

To overcome the limitations of GC when dealing with non-separable dynamical systems, several approaches have been based on relations among state-space reconstructions. For example, tests for the existence of directed unique mappings between reconstructed manifolds can be used as an all-or-nothing criterion to infer causal links [LPS91; CA09; MAC14; Sug+12; CGS15; PCH95], to which TC represents an extension since it allows for gradual quantification of the influence. TC is most closely related to the empirical procedure of convergent cross-mapping (CCM) [Sug+12] that yields interesting results in a range of applications, e.g. [Sug+12; Wan+14; Taj+15; Nes+15].

TC vs CCM

The CCM measure relies on errors when predicting one reconstruction manifold from another: the slower the convergence of the prediction error of $\boldsymbol{r}^{x_i}$ from $\boldsymbol{r}^{x_j}$ with increasing time series length, the weaker the causation $\boldsymbol{x}_i$ to $\boldsymbol{x}_j$. I suspect that this effect is a consequence of the expansion: the more expansive the mapping $M_{j\to i}$ locally is, the more its non-linearities will hamper predictions with a given finite number of data points. Something similar was observed in [Jia+16], however without establishing the link to the expansion of the mappings. In other words, I argue that CCM evaluates deviations from the assumption that the mapping $\{\boldsymbol{r}^{x_i}(t_1^{x_i}), ..., \boldsymbol{r}^{x_i}(t_k^{x_i})\}$ to $\{\boldsymbol{r}^{x_j}(t_1^{x_i}), ..., \boldsymbol{r}^{x_j}(t_k^{x_i})\}$ is linear and therefore is an indirect estimate of the underlying effective influence (figure 8.10), which TC measures directly through the expansion. Supporting this, it is found that CCM convergence speeds do not share the ability of TC to detect influences in linear stochastic systems (figure 8.12).

8.7 ACKNOWLEDGEMENTS

This chapter, including figures, was published in similar form in

> D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological Causality in Dynamical Systems," Physical Review Letters **119**, 098301 (2017). Copyright 2017 by the American Physical Society.

and its supplement. The text underwent some restructuring and the connection to part I was added (first two paragraphs). The study was designed by me and Klaus R. Pawelzik. I wrote the manuscript for publication, did most of the calculations, analysed and performed simulation and prepared the figures. Erik Laminski contributed to development and testing of numerical procedures, provided a draft text for the section on numerical methods in the paper (section 8.3) and designed figure 8.3. Maik Schünemann provided the mathematics for invariance under transformations and the connection to information theory as well as draft text for the correspond section in the paper (section 8.5.1) and designed figure 8.6.

9 CONCLUSION AND OUTLOOK

The scientific motivation of this part was to find a measure to quantify effective influence between interacting neuronal populations. It led to the formulation of topological causality (TC), a theory which provides a stringent mathematical basis for effective influence measures in generic dynamical system, which are objects of study in a wide variety of fields such as ecology, climatology, economy and neuroscience. Notably it provides a foundation and expands upon the CCM procedure that gained considerable traction recently in detection of causal links from experimental time series. The contents of chapter 8 were published in similar form [Har+17].

One aspect which was only cursorily touched upon was the applicability of TC in the presence of noise. Whereas numerical investigations suggest that stochasticity does not pose a principal problem, a proper mathematical proof is lacking in this exposition. The issue is actively researched in follow-up projects in the Pawelzik lab, along with improvements of the numerical methods.

A natural next step is to apply TC to neuroscientific data from experiments similar to the ones presented in chapters 3 and 5. This is pursued at the moment in the form of a master's thesis.

Once a robust toolbox is ready for release, I envision this work to contribute significantly by enabling detection of previously concealed interactions. Especially state-dependent tracking of effective influences in a complex system holds the potential to identify time points where the desired effect of an intervention is maximal. This could be used to control systems where some states on the attractor are undesirable, such as epileptic brain states or extinction events in ecosystems.

Closing the circle and coming back to experiments on flexible information routing, the TC approach holds the potential to shed more light onto how the involved populations are communicat-

ing. Possibly, communicating populations join a transient common attractor which should be characterized by a strongly state dependent feed forward effective influence if the CTC hypothesis is true. Attention would then be conceptualized as an intervention which drives the system towards these states. Since, as noted earlier, analysis of communicating populations under attentional conditions were so far almost exclusively analysed using Granger causality, there is room for possibly surprising results.

BIBLIOGRAPHY

[AH82] D. G. Albrecht and D. B. Hamilton. "Striate cortex of monkey and cat: contrast response function." In: *Journal of Neurophysiology* 48.1 (1982), pp. 217–237.

[AK10] T. Akam and D. M. Kullmann. "Oscillations and filtering networks support flexible routing of information." In: *Neuron* 67.2 (2010), pp. 308–320.

[AK90] L. F. Abbott and T. B. Kepler. "Model neurons: from Hodgkin-Huxley to Hopfield." In: *Statistical Mechanics of Neural Networks*. Ed. by L. Garrido. Vol. 368. Lecture Notes in Physics. Berlin Heidelberg: Springer, 1990, pp. 156–165.

[Alb+02] D. G. Albrecht, W. S. Geisler, R. A. Frazor, and A. M. Crane. "Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function." In: *Journal of Neurophysiology* 88.2 (2002), pp. 888–913.

[Alb75] K. Albus. "A auantitative study of the projection area of the central and the paracentral visual field in area 17 of the cat." In: *Experimental Brain Research* 24 (1975), pp. 159–179.

[Alb95] D. G. Albrecht. "Visual cortex neurons in monkey and cat: effect of contrast on the spatial and temporal phase transfer functions." In: *Visual Neuroscience* 12.6 (1995), pp. 1191–1210.

[AP08] N. Ay and D. Polani. "Information flows in causal networks." In: *Advances in Complex Systems* 11 (2008), pp. 17–41.

[AriBCa] Aristotle. *Metaphysics*. 350 BC.

[AriBCb] Aristotle. *On the parts of animals*. 350 BC.

[BA97] A. W. Bowman and A. Azzalini. *Applied smoothing techniques for data analysis: the kernel approach with s-plus illustrations*. Vol. 18. OUP Oxford, 1997.

[BAK12] G. Buzsáki, C. A. Anastassiou, and C. Koch. "The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes." In: *Nature Reviews Neuroscience* 13.6 (2012), pp. 407–420.

[Bar+02] M. Bartos, I. Vida, M. Frotscher, A. Meyer, H. Monyer, J. R. P. Geiger, and P. Jonas. "Fast synaptic inhibition promotes synchronized gamma oscillations in hippocampal interneuron networks." In: *PNAS* 99.20 (2002), pp. 13222–13227.

[Bas+15] A. M. Bastos, J. Vezoli, C. A. Bosman, J.-M. Schoffelen, R. Oostenveld, J. R. Dowdall, P. De Weerd, H. Kennedy, and P. Fries. "Visual areas exert feedforward and feedback influences through distinct frequency channels." In: *Neuron* 85.2 (2015), pp. 390–401.

[BBS09] L. Barnett, A. B. Barrett, and A. K. Seth. "Granger causality and transfer entropy are equivalent for Gaussian variables." In: *Physical Review Letters* 103 (2009), p. 238701.

[BBS17] L. Barnett, A. B. Barrett, and A. K. Seth. "Solved problems and remaining challenges for Granger causality analysis in neuroscience: A response to Stokes and Purdon (2017)." In: *arXiv preprint* (2017).

[BD04] G. Buzsáki and A. Draguhn. "Neuronal oscillations in cortical networks." In: *Science* 304.5679 (2004), pp. 1926–1929.

[Bea+92] C. Beaulieu, Z. Kisvarday, P. Somogyi, M. Cynader, and A. Cowey. "Quantitative distribution of GABA-immunopositive and immunonegative neurons and synapses in the monkey striate cortex (area 17)." In: *Cerebral Cortex* 2 (1992), pp. 295–309.

[BEK08] C. Boergers, S. Epstein, and N. J. Kopell. "Gamma oscillations mediate stimulus competition and attentional selection in a cortical network model." In: *PNAS* 105.46 (2008), pp. 18023–18028.

[BK08] C. Boergers and N. J. Kopell. "Gamma oscillations and stimulus selection." In: *Neural Computation* 20.2 (2008), pp. 383–414.

[BK15a] E. Bradley and H. Kantz. "Nonlinear time-series analysis revisited." In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25.9 (2015), p. 097610.

[BK15b] T. J. Buschman and S. Kastner. "From behavior to neural dynamics: an integrated theory of attention." In: *Neuron* 88.1 (2015), pp. 127–144.

[Bla01] R. Blake. "A primer on binocular rivalry, including current controversies." In: *Brain and Mind* 2 (2001), pp. 5–38.

[BLS95] R. Ben Yishai, R. Lev Bar-Or, and H. Sompolinsky. "Theory of orientation tuning in visual cortex." In: *PNAS* 92 (1995), pp. 3844–3848.

[BM07] T. J. Buschmann and E. K. Miller. "Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices." In: *Science* 315 (2007), pp. 1860–1862.

[BM09] T. J. Buschmann and E. K. Miller. "Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations." In: *Neuron* 63 (2009), pp. 386–396.

[Bon91] A. Bonds. "Temporal dynamics of contrast gain in single cells of the cat striate cortex." In: *Visual Neuroscience* 6.3 (1991), pp. 239–255.

[Bos+12] C. A. Bosman, J.-M. Schoffelen, N. Brunet, R. Oostenveld, A. M. Bastos, T. Womelsdorf, B. Rubehn, T. Stieglitz, P. De Weerd, and P. Fries. "Attentional stimulus selection through selective synchronization between monkey visual areas." In: *Neuron* 75 (2012), pp. 875–888.

[Bos+97] W. H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick. "Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex." In: *The Journal of Neuroscience* 17.6 (1997), pp. 2112–2127.

[BSE14] B. Bobier, T. C. Stewart, and C. Eliasmith. "A unifying mechanistic model of selective attention in spiking neurons." In: *PLoS Computational Biology* 10 (2014), e1003577.

[Buf+10] E. A. Buffalo, P. Fries, R. Landman, H. Liang, and R. Desimone. "A backward progression of attentional effects in the ventral stream." In: *PNAS* 107.1 (2010), pp. 361–365.

[Buz04] G. Buzsáki. "Large-scale recording of neuronal ensembles." In: *Nature Neuroscience* 7.5 (2004), p. 446.

[BW12] G. Buzsáki and X.-J. Wang. "Mechanisms of gamma oscillations." In: *Annual Review of Neuroscience* 35 (2012), pp. 203–225.

[BW76] B. D. Burns and A. C. Webb. "The spontaneous activity of neurones in the cat's visual cortex." In: *Proceedings of the Royal Society B: Biological Sciences* 194 (1976), pp. 211–223.

[BWD12] R. Bakker, T. Wachtler, and M. Diesmann. "CoCoMac 2.0 and the future of tract-tracing databases." In: *Frontiers in Neuroinformatics* 6 (2012), p. 30.

[CA09] D. Chicharro and R. G. Andrzejak. "Reliable detection of directional coupling using rank statistics." In: *Physical Review E* 80 (2009), p. 026217.

[Caj94] S. R. Y. Cajal. "The Croonian lecture: la fine structure des centres nerveux." In: *Proceedings of the Royal Society of London Series I* 55 (1894), pp. 444–468.

[CBM02] J. R. Cavanaugh, W. Bair, and J. A. Movshon. "Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons." In: *Journal of Neurophysiology* 88 (2002), pp. 2530–2546.

[CF04] H. J. Chisum and D. Fitzpatrick. "The contribution of vertical and horizontal connections to the receptive field center and surround in V1." In: *Neural Networks* 17 (2004), pp. 681–693.

[CG98] G. Caputo and S. Guerra. "Attentional selection by distractor suppression." In: *Vision Research* 38.5 (1998), pp. 669–689.

[CGS15] B. Cummins, T. Gedeon, and K. Spendlove. "On the efficacy of state space reconstruction methods in determining causality." In: *SIAM J. Applied Dynamical Systems* 14 (2015), pp. 335–381.

[CH94] M. Carandini and D. J. Heeger. "Summation and division by neurons in primate visual cortex." In: *Science* 264.5163 (1994), pp. 1333–1336.

[Cha+10] M. Chalk, J. L. Herrero, A. A. Gieselmann, L. S. Delicato, S. Gotthardt, and A. Thiele. "Attention reduces stimulus-driven gamma frequency oscillations and spike field coherence in V1." In: *Neuron* 66 (2010), pp. 114–125.

[CLR04] M. Carrasco, S. Ling, and S. Read. "Attention alters appearance." In: *Nature Neuroscience* 7.3 (2004), pp. 308–313.

[Csi+98] J. Csicsvari, H. Hirase, A. Czurkó, and G. Buzsáki. "Reliability and state dependence of pyramidal cell - interneuron synapses in the hippocampus: an ensemble approach in the behaving rat." In: *Neuron* 21 (1998), pp. 179–189.

[Csi+99] J. Csicsvari, H. Hirase, A. Czurkó, A. Mayima, and G. Buzsáki. "Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat." In: *The Journal of Neuroscience* 19 (1999), pp. 274–278.

[CWG02] C. Constantinidis, G. V. Williams, and P. S. Goldman-Rakic. "A role for inhibition in shaping the temporal flow of information in prefrontal cortex." In: *Nature Neuroscience* 5 (2002), pp. 175–180.

[DA01] P. Dayan and L. F. Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Ed. by P. Dayan. MIT Press, 2001.

[DB55] E. D. De Robertis and H. S. Bennett. "Some features of the submicroscopic morphology of synapses in frog and earthworm." In: *The Journal of Cell Biology* 1.1 (1955), pp. 47–58.

[Dow88] C. J. Downing. "Expectancy and visual-spatial attention: effects on perceptual quality." In: *Journal of Experimental Psychology: Human perception and performance* 14.2 (1988), p. 188.

[DR05] G. Deco and E. T. Rolls. "Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons." In: *Journal of Neurophysiology* 94 (2005), pp. 295–313.

[Efr79] B. Efron. "Bootstrap methods: another look at the Jacknife." In: *The Annals of Statistics* 7.1 (1979), pp. 1–26.

[EH72] C. W. Eriksen and J. E. Hoffman. "Temporal and spatial characteristics of selective encoding from visual displays." In: *Attention, Perception, & Psychophysics* 12.2 (1972), pp. 201–204.

[Ein+13] G. T. Einevoll, C. Kayser, N. K. Logothetis, and S. Panzeri. "Modelling and analysis of local field potentials for studying the function of cortical circuits." In: *Nature Reviews Neuroscience* 14.11 (2013), p. 770.

[Fri+01] P. Fries, J. H. Reynolds, A. E. Rorie, and R. Desimone. "Modulation of oscillatory neural synchronization by selective visual attention." In: *Science* 291 (2001), pp. 1560–1563.

[Fri+08] P. Fries, T. Womelsdorf, R. Oostenveld, and R. Desimone. "The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4." In: *The Journal of Neuroscience* 28 (2008), pp. 4823–4835.

[Fri05] P. Fries. "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence." In: *Trends in Cognitive Sciences* 9.10 (2005), pp. 474–480.

[Fri09] P. Fries. "Neuronal gamma-band synchronization as a fundamental process in cortical computation." In: *Annual Review of Neuroscience* 32 (2009), pp. 209–224.

[Fri15] P. Fries. "Rhythms for cognition: communication through coherence." In: *Neuron* 88.1 (2015), pp. 220–235.

[FSM17] L. Faes, S. Stramaglia, and D. Marinazzo. "On the interpretability and computational reliability of frequency-domain Granger causality." In: *arXiv preprint* (2017).

[FV91] D. J. Felleman and D. C. Van Essen. "Distributed hierarchical processing in the primate cerebral cortex." In: *Cerebral Cortex* 1 (1991), pp. 1047–3211.

[GAC07] W. S. Geisler, D. G. Albrecht, and A. M. Crane. "Responses of neurons in primary visual cortex to transient changes in local contrast and luminance." In: *Journal of Neuroscience* 27.19 (2007), pp. 5063–5067.

[Gas+16] J. M. Gaspar, G. J. Christie, D. J. Prime, P. Jolicœur, and J. J. McDonald. "Inability to suppress salient distractors predicts low visual working memory capacity." In: *PNAS* 113.13 (2016), pp. 3693–3698.

[GB09] D. F. M. Goodman and R. Brette. "The Brian simulator." In: *Frontiers in Neuroscience* 3.2 (2009), pp. 192–197.

[Gew82] J. Geweke. "Measurement of linear dependence and feedback between multiple time series." In: *Journal of the American Statistical Association* 77.378 (1982), pp. 304–313.

[GGS81] R. Gattass, C. Gross, and J. Sandell. "Visual topography of V2 in the macaque." In: *Journal of Comparative Neurology* 201.4 (1981), pp. 519–539.

[GPS15] G. G. Gregoriou, S. Paneri, and P. Sapountzis. "Oscillatory synchrony as a mechanism of attentional processing." In: *Brain Research* (2015). in press.

[Gra69] C. W. J. Granger. "Investigating causal relations by econometric models and cross-spectral methods." In: *Econometrica* 37 (1969), pp. 424–438.

[Gre+09] G. G. Gregoriou, S. J. Gotts, H. Zhou, and R. Desimone. "High-frequency, long range coupling between prefrontal and visual cortex during attention." In: *Science* 324.1 (2009), pp. 1207–1210.

[Gro+12] I. Grothe, S. D. Neitzel, S. Mandon, and A. K. Kreiter. "Switching neuronal inputs by differential modulations of gamma-band phase-coherence." In: *The Journal of Neuroscience* 32.46 (2012), pp. 16172–16180.

[Gro+15] I. Grothe, D. Rotermund, S. D. Neitzel, S. Mandon, U. A. Ernst, A. K. Kreiter, and K. R. Pawelzik. "Attention selectively gates afferent signal transmission to area V4." In: *arxiv* (2015).

[Gro95] C. G. Gross. "Aristotle on the brain." In: *The Neuroscientist* 1.4 (1995), pp. 245–250.

[GS89] C. M. Gray and W. Singer. "Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex." In: *PNAS* 86.5 (1989), pp. 1698–1702.

[GSG88] R. Gattass, A. Sousa, and C. Gross. "Visuotopic organization and extent of V3 and V4 of the macaque." In: *Journal of Neuroscience* 8.6 (1988), pp. 1831–1845.

[Gul+14] R. Gulbinaite, A. Johnson, R. de Jong, C. C. Morey, and H. van Rijn. "Dissociable mechanisms underlying individual differences in visual working memory capacity." In: *Neuroimage* 99 (2014), pp. 197–206.

[Har+17] D. Harnack, E. Laminski, M. Schünemann, and K. R. Pawelzik. "Topological causality in dynamical systems." In: *Physical Review Letters* 119.9 (2017), p. 098301.

[Has+17] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick. "Neuroscience-inspired artificial intelligence." In: *Neuron* 95.2 (2017), pp. 245–258.

[Hau+12] S. Haufe, V. V. Nikulin, K.-R. Müller, and G. Nolte. "A critical assessment of connectivity measures for eeg data: A simulation study." In: *Neuroimage* 64 (2012), pp. 120–133.

[HEP15] D. Harnack, U. A. Ernst, and K. R. Pawelzik. "A model for attentional information routing through coherence predicts biased competition and multistable perception." In: *Journal of Neurophysiology* 114.3 (2015), pp. 1593–1605.

[Her+08] J. L. Herrero, M. J. Roberts, L. S. Delicato, M. A. Gieselmann, P. Dayan, and A. Thiele. "Acetylcholine contributes through muscarinic receptors to attentional modulation in V1." In: *Nature* 454 (2008), pp. 1110–1114.

[HG91] J. A. Hirsch and C. D. Gilbert. "Synaptic physiology of horizontal connections in the cat's visual cortex." In: *The Journal of Neuroscience* 11 (1991), pp. 1800–1809.

[HH52] A. L. Hodgkin and A. F. Huxley. "A quantitative description of membrane current and its application to conduction and excitation in nerve." In: *The Journal of Physiology* 117.4 (1952), pp. 500–544.

[HHL13] F. Hutter, H. Hoos, and K. Leyton-Brown. "An evaluation of sequential model-based optimization for expensive blackbox functions." In: *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*. ACM. 2013, pp. 1209–1216.

[HM07] S. Haeusler and W. Maass. "A statiscal analysis of information-processing properties of lamina-specific cortical mircocircuit models." In: *Cerebral Cortex* 17 (2007), pp. 149–162.

[Hof+11] S. B. Hofer, H. Ko, B. Pichler, J. Vogelstein, H. Ros, H. Zeng, E. Lein, N. A. Lesica, and T. D. Mrsic-Flogel. "Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex." In: *Nature Neuroscience* 14 (2011), pp. 1045–1052.

[Hop+06] J.-M. Hopf, C. Boehler, S. Luck, J. Tsotsos, H.-J. Heinze, and M. Schoenfeld. "Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision." In: *PNAS* 103.4 (2006), pp. 1053–1058.

[HSF09] M. Helmstaedter, B. Sakman, and D. Feldmeyer. "Neuronal correlates of local, lateral, and translaminar inhibition with reference to cortical columns." In: *Cerebral Cortex* 19.4 (2009), pp. 926–937.

[Hun07] J. D. Hunter. "Matplotlib: a 2D graphics environment." In: *Computing in Science and Engineering* 9.3 (2007), pp. 90–95.

[HV06] J. Hegde and D. C. Van Essen. "A comparative study of shape representation in macaque visual areas V2 and V4." In: *Cerebral Cortex* 17.5 (2006), pp. 1100–1116.

[IG99] M. Ito and C. D. Gilbert. "Attention modulates contextual influences in the primary visual cortex of alert monkeys." In: *Neuron* 22.3 (1999), pp. 593–604.

[IS11] J. S. Isaacson and M. Scanziani. "How inhibition shapes cortical activity." In: *Neuron* 72 (2011), pp. 231–243.

[Jam90] W. James. *The principles of psychology*. Ed. by W. James. Vol. 1. Henry Holt, New York, 1890.

[Jan+12] D. Janzing, J. Mooij, K. Zhang, J. Lemeire, J. Zscheischler, P. Daniušsis, B. Steudel, and B. Schölkopf. "Information - geometric approach to inferring causal directions." In: *Artificial Intelligence* 182–183 (2012), pp. 1–31.

[Jan08] G. Janzen. "Bennett and Hacker on neural materialism." In: *Acta Analytica* 23.3 (2008), pp. 273–286.

[Jia+16] J.-J. Jiang, Z.-G. Huang, L. Huang, H. Liu, and Y.-C. Lai. "Directed dynamical influence is more detectable with noise." In: *Scientific Reports* 6 (2016).

[Kano0] E. R. Kandel. *Principles of neural science*. Ed. by T. M. J. Eric R. Kandel James H. Schwartz. 4th ed. McGraw-Hill, 2000.

[KE13] J. Kretzberg and U. Ernst. "Vision." In: *Neurosciences. From molecule to behavior: a university textbook*. Ed. by C. G. Galizia and P. M. Lledo. Berlin Heidelberg: Springer, 2013.

[Ker+10] A. M. Kerlin, M. L. Andermann, V. K. Berezovskii, and R. C. Reid. "Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex." In: *Neuron* 67.5 (2010), pp. 858–871.

[KLZ13] R. J. Krauzlis, L. P. Lovejoy, and A. Zénon. "Superior colliculus and visual spatial attention." In: *Annual Review of Neuroscience* 36.1 (2013), pp. 165–182.

[KN11] V. V. Klinshov and V. I. Nekorkin. "Synchronization of time-delay coupled pulse oscillators." In: *Chaos, Solitons and Fractals* 44 (2011), pp. 98–107.

[LaB83] D. LaBerge. "Spatial extent of attention to letters and words." In: *Journal of Experimental Psychology: Human Perception and Performance* 9.3 (1983), pp. 371–379.

[LC06] S. Ling and M. Carrasco. "Sustained and transient covert attention enhance the signal via different contrast response functions." In: *Vision Research* 46.8 (2006), pp. 1210–1220.

[LE16] D. Lisitsyn and U. Ernst. "Model-based inferences into attention and bistability information routing control via precisely-timed perturbations." In: Berlin: BCCN Berlin, 2016.

[Lee+14] A. T. Lee, S. G. Gee, D. Vogt, T. Patel, J. L. Rubinstein, and V. S. Sohal. "Pyramidal neurons in prefrontal cortex receive subtype-specific form of excitation and inhibition." In: *Neuron* 81 (2014), pp. 61–68.

[Li05] Z. Li. "The primary visual cortex creates a bottom-up saliency map." In: *Neurobiology of Attention*. Ed. by L. Itti, G. Rees, and J. K. Tsotsos. Elsevier, 2005.

[Lor63] E. N. Lorenz. "Deterministic nonperiodic flow." In: *Journal of the Atmospheric Sciences* 20 (1963), pp. 130–141.

[LPG04] W. Li, V. Piëch, and C. D. Gilbert. "Perceptual learning and top-down influences in primary visual cortex." In: *Nature Neuroscience* 7.6 (2004), pp. 651–657.

[LPS91] W. Liebert, K. R. Pawelzik, and H. G. Schuster. "Optimal embeddings of chaotic attractors from topological considerations." In: *Europhysics Letters* 14 (1991), pp. 521–526.

[LST15] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum. "Human-level concept learning through probabilistic program induction." In: *Science* 350.6266 (2015), pp. 1332–1338.

[Luc+97] S. J. Luck, L. Chelazzi, S. A. Hillyard, and R. Desimone. "Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex." In: *Journal of Neurophysiology* 77.1 (1997), pp. 24–42.

[Luc95] S. J. Luck. "Multiple mechanisms of visual-spatial attention: recent evidence from human electrophysiology." In: *Behavioural Brain Research* 71.1-2 (1995), pp. 113–123.

[MAC14] H. Ma, K. Aihara, and L. Chen. "Detecting causality from nonlinear dynamics with short-term time series." In: *Scientific Reports* 4 (2014), p. 7464.

[Mal+93] R. Malach, Y. Amir, M. Harel, and A. Grinvald. "Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex." In: *PNAS* 90.22 (1993), pp. 10469–10473.

[Mal81] C. von der Malsburg. "The correlation theory of brain function. MPI Biophysical Chemistry, Internal Report 81-2." In: *Models of neural networks, II (1994)*. Ed. by E. Domany, J. L. van Hemmen, and K. Schulten. Springer, 1981.

[Mas09] N. Masuda. "Selective population rate coding: a possible computational role of gamma oscillations in selective attention." In: *Neural Computation* 21 (2009), pp. 3335–3362.

[McC+06] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon. "A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955." In: *AI magazine* 27.4 (2006), p. 12.

[MCW08] K. McAlonan, J. R. Cavanaugh, and R. H. Wurtz. "Guarding the gateway to cortex with attention in visual thalamus." In: *Nature* 456.1 (2008), pp. 391–395.

[MD85] J. Moran and R. Desimone. "Selective attention gates visual processing in the extrastriate cortex." In: *Science* 229 (1985), pp. 782–784.

[MFS06] J. Mishra, J.-M. Fellous, and T. J. Sejnowski. "Selective attention through phase relationship of excitatory and inhibitory input synchrony in a model cortical neuron." In: *Neural Networks* 19 (2006), pp. 1329–1346.

[MGI93] T. Murakoshi, J.-Z. Guo, and T. Ichinose. "Electropyhsiological identification of horizontal synaptic connections in rat visual cortex in vitro." In: *Neuroscience Letters* 163 (1993), pp. 211–214.

[MI05] R. Marois and J. Ivanoff. "Capacity limits of information processing in the brain." In: *Trends in Cognitive Sciences* 9.6 (2005), pp. 296–305.

[Mic+16] G. Michalareas, J. Vezoli, S. Van Pelt, J.-M. Schoffelen, H. Kennedy, and P. Fries. "Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas." In: *Neuron* 89.2 (2016), pp. 384–397.

[MKW12] J. S. Montijn, P. C. Klink, and R. J. A. van Wezel. "Divisive normalization and neuronal oscillations in a single hierarchical framework of selctive visual attention." In: *Frontiers in Neural Circuits* 6 (2012), p. 22.

[MLF08] C. M. Moore, L. K. Lanagan-Leitzel, and E. M. Fine. "Distinguishing between the precision of attentional localization and attentional resolution." In: *Perception & Psychophysics* 70.4 (2008), pp. 573–582.

[MM99] C. J. McAdams and J. H. Maunsell. "Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4." In: *Journal of Neuroscience* 19.1 (1999), pp. 431–441.

[Mot09] B. C. Motter. "Central V4 receptive fields are scaled by the V1 cortical magnification and correspond to a constant-sized sampling of the V1 surface." In: *The Journal of Neuroscience* 29 (2009), pp. 5749–5757.

[Mot93] B. C. Motter. "Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli." In: *Journal of Neurophysiology* 70.3 (1993), pp. 909–919.

[MSR07] J. F. Mitchell, K. A. Sundberg, and J. H. Reynolds. "Differential attention-dependent response modulation across cell classes in macaque visual area V4." In: *Neuron* 55 (2007), pp. 131–141.

[MT02] J. C. Martinez-Trujillo and S. Treue. "Attentional modulation strength in cortical area MT depends on stimulus contrast." In: *Neuron* 35 (2002), pp. 365–370.

[Nes+15] E. H. van Nes, M. Scheffer, V. Brovkin, T. M. Lenton, H. Ye, E. Deyle, and G. Sugihara. "Causal feedbacks in climate change." In: *Nature Climate Change* 5.5 (2015), pp. 445–448.

[Ni+16] J. Ni, T. Wunderle, C. M. Lewis, R. Desimone, I. Diester, and P. Fries. "Gamma-rhythmic gain modulation." In: *Neuron* 92.1 (2016), pp. 240–251.

[NK95] D. A. Nelson and L. C. Katz. "Emergence of functional circuits in ferret visual cortex visualized by optical imaging." In: *Neuron* 15 (1995), pp. 23–34.

[Nol+08] G. Nolte, A. Ziehe, V. V. Nikulin, A. Schlögl, N. Krämer, T. Brismar, and K.-R. Müller. "Robustly estimating the flow direction of information in complex physical systems." In: *Physical Review Letters* 100.23 (2008), p. 234101.

[Noo+16] M. P. Noonan, N. Adamian, A. Pike, F. Printzlau, B. M. Crittenden, and M. G. Stokes. "Distinct mechanisms for distractor suppression and target facilitation." In: *Journal of Neuroscience* 36.6 (2016), pp. 1797–1807.

[Nun+99] P. L. Nunez, R. B. Silberstein, Z. Shi, M. R. Carpenter, R. Srinivasan, D. M. Tucker, S. M. Doran, P. J. Cadusch, and R. S. Wijesinghe. "EEG coherency II: experimental comparisons of multiple measures." In: *Clinical Neurophysiology* 110.3 (1999), pp. 469–486.

[Oli07] T. E. Oliphant. "Python for scientific computing." In: *Computing in Science and Engineering* 9.3 (2007), pp. 10–20.

[Owe+16] A. Owens, P. Isola, J. McDermott, A. Torralba, E. H. Adelson, and W. T. Freeman. "Visually indicated sounds." In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

[Pac+80] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. "Geometry from a time series." In: *Physical Review Letters* 45 (1980), pp. 712–716.

[Pal+17] A. Palmigiano, T. Geisel, F. Wolf, and D. Battaglia. "Flexible information routing by transient synchrony." In: *Nature Neuroscience* 20.7 (2017), p. 1014.

[PC99] A. Pasupathy and C. E. Connor. "Responses to contour features in macaque area V4." In: *Journal of Neurophysiology* 82.5 (1999), pp. 2490–2502.

[PCH95] L. M. Pecora, T. L. Carroll, and J. F. Heagy. "Statistics for mathematical properties of maps between time series embeddings." In: *Physical Review E* 52.4 (1995), p. 3420.

[PG07] F. Pérez and B. E. Granger. "IPython: a system for interactive scientific computing." In: *Computing in Science and Engineering* 9.3 (2007), pp. 21–29.

[Pos+08] M. Pospischil, M. Toledo-Rodriguez, C. Monier, Z. Piwkowska, T. Bal, Y. Frégnac, H. Markram, and A. Destexhe. "Minimal Hodgkin–Huxley type models for different classes of cortical and thalamic neurons." In: *Biological Cybernetics* 99 (2008), pp. 427–441.

[QNB04] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul. "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering." In: *Neural Computation* 16.8 (2004), pp. 1661–1687.

[RCD99] J. H. Reynolds, L. Chelazzi, and R. Desimone. "Competitive mechanisms subserve attention in macaque areas V2 and V4." In: *The Journal of Neuroscience* 19 (1999), pp. 1736–1753.

[RD03] J. H. Reynolds and R. Desimone. "Interacting roles of attention and visual saliency in V4." In: *Neuron* 37 (2003), pp. 853–863.

[RL83] K. S. Rockland and J. S. Lund. "Intrinsic laminar lattice connections in primate visual cortex." In: *The Journal of Comparative Neurology* 216 (1983), pp. 303–318.

[RM10] S. Ray and J. H. R. Maunsell. "Differences in gamma frequencies across visual cortex restrict their possible use in computation." In: *Neuron* 67 (2010), pp. 885–896.

[RN10] C. E. Rasmussen and H. Nickisch. "Gaussian processes for machine learning (GPML) toolbox." In: *Journal of Machine Learning Research* 11 (2010), pp. 3011–3015.

[Rob+13] M. J. Roberts, E. Lowet, N. M. Brunet, M. Ter Wal, P. Tiesinga, P. Fries, and P. De Weerd. "Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching." In: *Neuron* 78.3 (2013), pp. 523–536.

[Rös76] O. E. Rössler. "An equation for continuous chaos." In: *Physics Letters A* 57.5 (1976), pp. 397–398.

[RPD00] J. H. Reynolds, T. Pasternak, and R. Desimone. "Attention increases sensitivity of V4 neurons." In: *Neuron* 26.3 (2000), pp. 703–714.

[RTN02] M. C. van Rossum, G. G. Turrigiano, and S. B. Nelson. "Fast propagation of firing rates through layered networks of noisy neurons." In: *Journal of Neuroscience* 22.5 (2002), pp. 1956–1966.

[SA14] V. S. Störmer and G. A. Alvarez. "Feature-based attention elicits surround suppression in feature space." In: *Current Biology* 24.17 (2014), pp. 1985–1988.

[Schoo] T. Schreiber. "Measuring information transfer." In: *Physical Review Letters* 85 (2000), pp. 461–464.

[Sha+16] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas. "Taking the human out of the loop: A review of Bayesian optimization." In: *Proceedings of the IEEE* 104.1 (2016), pp. 148–175.

[Sil+17] D. Silver et al. "Mastering the game of Go without human knowledge." In: *Nature* 550 (2017), pp. 354–359.

[Sin99] W. Singer. "Neuronal synchrony: a versatile code for the definition of relations?" In: *Neuron* 24 (1999), pp. 49–65.

[SK93] W. R. Softky and C. Koch. "The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs." In: *The Journal of Neuroscience* 13.1 (1993), pp. 334–350.

[SN98] M. N. Shadlen and W. T. Newsome. "The variable discharge of cortical neurons: implications for connectivity, computation, and information coding." In: *Journal of Neuroscience* 18.10 (1998), pp. 3870–3896.

[SP17a] P. A. Stokes and P. L. Purdon. "A study of problems encountered in Granger causality analysis from a neuroscience perspective." In: *PNAS* 114.34 (2017), E7063–E7072.

[SP17b] P. A. Stokes and P. L. Purdon. "In reply to Faes et al. and Barnett et al. regarding: A study of problems encountered in Granger causality analysis from a neuroscience perspective." In: *arXiv preprint* (2017).

[SP96] P. A. Salin and D. A. Prince. "Electropyhsiological mapping of GABAA receptor-mediated inhibition in adult rat somatosensory cortex." In: *Journal of Neurophysiology* 75 (1996), pp. 1589–1600.

[Sri+09] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger. "Gaussian process optimization in the bandit setting: No regret and experimental design." In: *arXiv preprint* (2009).

[Sug+12] G. Sugihara, R. M. May, H. Ye, C. Hsieh, E. R. Deyle, and M. Fogarty. "Detecting causality in complex ecosystems." In: *Science* 334 (2012), pp. 496–500.

[Syl+08] C. M. Sylvester, A. I. Jack, M. Corbetta, and G. L. Shulman. "Anticipatory suppression of nonattended locations in visual cortex marks target location and predicts perception." In: *Journal of Neuroscience* 28.26 (2008), pp. 6549–6556.

[Tai+14] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. "Deepface: closing the gap to human-level performance in face verification." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1701–1708.

[Taj+15] S. Tajima, T. Yanagawa, N. Fujii, and T. Toyoizumi. "Untangling brain-wide dynamics in consciousness by cross-embedding." In: *PLoS Computational Biology* 1004537 (2015).

[Tak81] F. Takens. "Detecting strange attractors in turbulence." In: *Dynamical systems and turbulence*. Ed. by D. A. Rand and L.-S. Young. Vol. 898. Springer Lecture Notes in Mathematics. Berlin: Springer, 1981.

[Tay+05] K. Taylor, S. Mandon, W. A. Freiwald, and A. K. Kreiter. "Coherent oscillatory activity in monkey area V4 predicts successful allocation of attention." In: *Cerebral Cortex* 15.9 (2005), pp. 1424–1437.

[Thi+09] A. Thiele, A. Pooresmaeili, L. S. Delicato, J. L. Herrero, and P. R. Roelfsema. "Additive effects of attention and stimulus contrast in primary visual cortex." In: *Cerebral Cortex* 19 (2009), pp. 2970–2981.

[TMB06] F. Tong, M. Meng, and R. Blake. "Neural bases of binocular rivalry." In: *Trends in Cognitive Sciences* 10 (2006), pp. 502–511.

[Tor+09] A. B. Tort, R. W. Komorowski, J. R. Manns, N. J. Kopell, and H. Eichenbaum. "Theta–gamma coupling increases during the learning of item–context associations." In: *PNAS* 106.49 (2009), pp. 20942–20947.

[TPR78] R. J. Tusa, L. A. Palmer, and A. C. Rosenquist. "The retinotopic organization of area 17 (striate cortex) in the cat." In: *Journal of Comparative Neurology* 177 (1978), pp. 213–235.

[TS09] P. Tiesinga and T. J. Sejnowski. "Cortical enlightenment: are attentional gamma oscillations driven by ING or PING?" In: *Neuron* 63.6 (2009), pp. 727–732.

[TS10] P. H. Tiesinga and T. J. Sejnowski. "Mechanisms for phase shifting in cortical networks and their role in communication through coherence." In: *Frontiers in Human Neuroscience* 4 (2010), p. 196.

[Tur50] A. M. Turing. "Computing machinery and intelligence." In: *Mind* 59.236 (1950), pp. 433–460.

[VA05] T. P. Vogels and L. F. Abbott. "Signal propagation and logic gating in networks of integrate-and-fire neurons." In: *Journal of Neuroscience* 25.46 (2005), pp. 10786–10795.

[Vin+13] M. Vinck, T. Womelsdorf, E. A. Buffalo, R. Desimone, and P. Fries. "Attentional modulation of cell-class-specific gamma-band synchronization in awake monkey area V4." In: *Neuron* 80.4 (2013), pp. 1077–1089.

[Von67] H. Von Helmholtz. *Handbuch der physiologischen Optik*. Vol. 9. Voss, 1867.

[WAH00] J. M. Wolfe, G. A. Alvarez, and T. S. Horowitz. "Attention is fast but volition is slow." In: *Nature* 406 (2000), p. 691.

[Wan+14] X. Wang et al. "A two-fold increase of carbon cycle sensitivity to tropical temperature variations." In: *Nature* 506 (2014), pp. 212–215.

[WC72] H. R. Wilson and J. D. Cowan. "Excitatory and inhibitory interactions in localized populations of model neurons." In: *Biophysical Journal* 12.1 (1972), pp. 1–24.

[Wom+07] T. Womelsdorf, J.-M. Schoffelen, R. Oostenveld, W. Singer, R. Desimone, A. K. Engel, and P. Fries. "Modulation of neuronal interactions through neuronal synchronization." In: *Science* 316 (2007), pp. 1609–1612.

[Wom+14] T. Womelsdorf, T. A. Valiante, N. T. Sahin, K. J. Miller, and P. Tiesinga. "Dynamic circuit motifs underlying rhythmic gain control, gating and integration." In: *Nature Neuroscience* 17 (2014), pp. 1031–1039.

[WR06] C. K. I. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*. MIT Press, 2006.

[WS12] M. Wildie and M. Shanahan. "Establishing communication between neuronal populations through competitive entrainment." In: *Frontiers in Computational Neuroscience* 5 (2012), p. 62.

[WT17] M. ter Wal and P. H. Tiesinga. "Phase difference between model cortical areas determines level of information transfer." In: *Frontiers in Computational Neuroscience* 11 (2017), p. 6.

[Zem+13] R. Zemankovics, J. M. Veres, I. Oren, and N. Hajos. "Feedforward inhibition underlies the propagation of cholinergically induces gamma oscillations from hippocampal CA3 to CA1." In: *The Journal of Neuroscience* 33 (2013), pp. 12337–12351.

[ZFG08] M. Zeitler, P. Fries, and S. Gielen. "Biased competition through variations in amplitude of γ-oscillations." In: *Journal of Computational Neuroscience* 25.1 (2008), pp. 89–107.

[ZK15] A. Zandvakili and A. Kohn. "Coordinated neuronal activity enhances corticocortical communication." In: *Neuron* 87.4 (2015), pp. 827–839.

ACKNOWLEDGEMENTS

Firstly, I want to thank my supervisor Udo Ernst for giving me the opportunity to pursue science in the beautiful Cognium surrounded by inspiring people, letting me stray a bit from the foreseen path, and giving me the opportunity to be creative.

I also want to thank my committee members Udo A. Ernst, Andreas K. Kreiter, Manfred Radmacher and Michael Mackey for their time and patience.

Without the help of collaborators, with whom I had the pleasure to work over the years, this thesis would not have been possible. They are, in no particular order: Udo A. Ernst, Klaus R. Pawelzik, Eric Drebitz, Dmitriy Lisitsyn, Maik Schünemann, Erik Laminski, Lukas Rausch and Andreas K. Kreiter.

I would also like to thank all my other office mates with whom I shared part of my academic and private endeavours.

FUNDING

My research position was funded by the Bundesministerium für Bildung und Forschung (Bernstein Award Udo Ernst, Grant 01GQ1106).

DECLARATION

Hiermit versichere ich, dass ich

1. die Dissertation ohne unerlaubte Hilfe angefertigt habe,
2. keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet habe,
3. die den zitierten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Bremen, 8. März 2018

Daniel Harnack

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

`https://bitbucket.org/amiede/classicthesis/`

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

`http://postcards.miede.de/`

Thank you very much for your feedback and contribution.

Final Version as of May 18, 2018 (`classicthesis` version 4.4).