Quantitative models of information transfer during mammalian visual processing
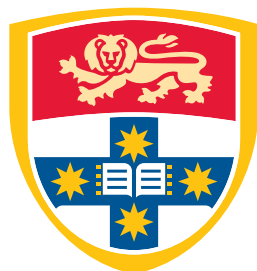
Preethom Pal
SID: 470407190

Supervisor: A/Prof Joseph Lizier

This thesis is submitted in partial fulfillment of
the requirements for the degree of
Bachelor of Science/Advanced Computing (Honours)

School of Computer Science
The University of Sydney
Australia

May 2023

Student Plagiarism: Compliance Statement
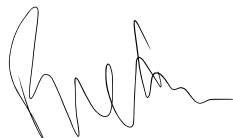
I certify that:

I have read and understood the University of Sydney Student Plagiarism: Coursework Policy and Procedure;

I understand that failure to comply with the Student Plagiarism: Coursework Policy and Procedure can lead to the University commencing proceedings against me for potential student misconduct under Chapter 8 of the University of Sydney By-Law 1999 (as amended);

This Work is substantially my own, and to the extent that any part of this Work is not my own I have indicated that it is not my own by Acknowledging the Source of that part or those parts of the Work.

Name: Preethom Pal

Signature: **Date**:   28 May 2023

Abstract

Information transfer is one piece to the three-part puzzle of explaining information processing which includes information storage, transfer and modification. Quantifying any of these in a natural system like a neuronal population is a significant challenge given the differences in computation by a dynamic complex system, housing highly distributed and parallelised processes, to traditional computers. However, an information theoretic approach allows direct quantification of information processing in a distributed setting, and its measures attain statistical power even in the presence of high-dimensional data. Meanwhile, recent advancements in high density extracellular imaging via silicon probes allowed the capture of a high resolution dataset showing in vivo neural activity in the visual areas of mouse brains. This leads us to a unique opportunity to directly quantify visual information processing in mice. Retinotopy and lamination in the mammalian visual cortex suggests hierarchical information processing of the visual field. Although this can be demonstrated by studying latency of functional activity between layers, the literature does not offer quantification of typical information transfer between these cortical layers, or across higher order association visual areas. This project evaluates two major tools that quantify information transfer; transfer entropy and effective network inference. Transfer entropy is able to measure the information transfer between two neuronal processes, over some duration of activity, while inferring an effective network can capture multivariate information transfers inside a neuronal population. Historically, estimating transfer entropy required binning of signalling events into discrete time bins, however recently an estimator for transfer entropy in continuous time was developed. Our results from using the continuous-time estimator in conjunction with effective network inference add nuance to expectations in the primary visual cortex. Hierarchical processing is not necessarily followed during short time windows. Layer 4 is not always the primary source of information to all layers. Layers 2/3 and 5 do not always reflect intermediary and final information processing respectively. Transfers between cortical areas from disparate areas of the visual field mostly follow expectations but we find unexpected transfers from high order to low order layers. These findings reflect how short-term information transfers need not follow functional analysis. Our evaluation establishes the information theoretic approach as necessary to better understand not only information processing in the brain but natural computation as a whole.

Acknowledgements

CONTENTS

List of Figures

List of Tables

Introduction

This project is at the intersection of three disciplines; neuroscience, complex systems theory and information theory. This introduction describes the project and its motivations in general terms so that they are understandable to the computer scientist. Further necessary details from each discipline will be covered in the literature review.

The aim of this project is to use and evaluate various tools that are able to quantify information transfer inside the visual cortex. The results are presented to improve our understanding of mammalian information processing and to showcase novel tools that can be further applied and refined to understand any natural source of information processing. The project is a step towards describing in detail how mammalian brains are able to compute such dense informational inputs for an impressively generalised set of tasks, and describe where pathologies in information processing emerge, a major goal in neuroscience (Oppenheimer and Kelso, 2015). The project also aids a major goal in computer science by way of providing inspiration for the design of artificial intelligence systems from naturally occurring computation. Biological sources of computation are particularly good at generalising across tasks and at producing precise behaviours, under continuous and widely distributed informational input (Wang et al., 2007). Current artificial intelligence rarely generalises well, and processes discrete informational inputs which limits its performance in physical settings, such as robotics and self-driving cars.

These goals are aided as follows. Given data of the times and locations of signalling events in a neuronal population, we quantify the information transferred between neurons over a specific observation window. Information transfer can be quantified using a pairwise or multivariate analysis (Novelli and Lizier, 2021). The latter leads to a network model that reflects transfers between pairs of neurons but in the context of the whole population's activity. This network model is called an effective network (Shorten et al., 2022b). An effective network could vary from task to task, even when inferred from the same neuronal population with a relatively fixed structural connectivity (Friston, 2011). The effective

network does not reflect structural connections and instead reflects multivariate information transfers that are expected to vary across computational tasks. This model could therefore lead to a deeper understanding of how a population of mammalian neurons are able to process continuous informational input across generalised tasks, by pointing to new multivariate transfers in different tasks. Effective networks could also elucidate signatures of information processing typical to some tasks or areas of the brain, and how these differ inside pathologies. Finally, an effective network and future models in the same approach quantify natural information processing, with no associated details about the system's physics or physiology. In other words, the natural system is reduced to a model of its information processing ability only. Such a model would be helpful in engineering an artificial system that mimics natural information processing, since one could begin identify and build the required fundamental elements that are most relevant (Kaiser, 2007).

The visual cortex was chosen as it is a well understood area of neuroscience, being relatively easy to image, with the ability to easily test the effects of varying input on brain activity. This gives the project the opportunity to discuss its findings from novel methods in the context of a well established literature. The current understanding of mammalian visual processing describes complex hierarchical processing inside the primary visual cortex, with nuanced directional information flow (Thomas Yeo et al., 2011). Information flows are described between at least six cortical layers, as well as to and from visual association areas outside the primary visual cortex. Limitations to the current understanding of visual processing arise due to studies focusing on either physical structure or functional connectivity inferred from pairwise comparison of active brain regions during tasks in vivo (Harris and Shepherd, 2015). Physical connectivity does not represent information processing. While functional connectivity is a better analogue for information processing, it stills fails to capture multivariate information processing, such as the visual encoding of a scene allowed by three or more active areas (Geerligs et al., 2016). These are not easy or unnoticed challenges. The high-dimensional nature of neuronal populations explodes the state space inside their activity, where each state could be encoding some useful information to the animal (Novelli et al., 2019). It is difficult to measure how representations of information could be interacting and transforming with such a large search space of possibly meaningful states.

The information theoretic approach evaluated in this project improves on these limitations by offering tools to quantify information transfer in a neuronal population, even with high-dimensional data. These tools measure how much information is being transferred inside a neuronal population and in which directions. The information transfer measured does not describe what the information may represent or

how it might be transformed for the purposes of solving a particular task. The methods used are novel since they quantify information transfer directly, as opposed to some analogue of changing information, as is the case in typical analysis of functional brain imaging (Bzdok and Yeo, 2017).

The data used in this project comes from in vivo extracellular recordings of neuronal signalling events inside and around the primary visual cortex (V1) of mice, measured by Stringer et al. (2019). The results are discussed in the context of current understanding about the hierarchical information processing done by the six cortical layers in V1 and surrounding areas. The methods are described and evaluated in detail so that they may be applied to further recorded neuronal populations to improve understanding of mammalian information processing. The overall approach may also be applied to any recorded source of natural computation.

1.1 List of project aims

Given recordings of signalling events from visual areas in mice, our global aim is to identify and quantify significant information transfers inside the neuronal population.

(1) Search for pairwise information transfer inside and between local excitatory circuits (otherwise called 'columns') in the primary visual cortex (V1) in mice.

    (a) Search for pairwise transfers from cells in each layer of V1 columns, to cells in every other layer.

    (b) Search for pairwise transfers between V1 columns.

    (c) Search for pairwise transfers between V1 columns and association visual areas. The latter are cortices outside of V1 that still process visually related information.

(2) Infer effective network(s) inside and between local excitatory circuits in V1.

(3) Examine how the findings on quantified information transfer can contribute to current knowledge about visual processing.

    (a) Quantification should add to understanding about the direction and capacity of information transfer.

    (b) Comparison of pairwise analysis and the multivariate analysis offered by effective networks should elucidate how visual areas are able to take advantage of a distributed form of information processing.

(4) Evaluate the extent to which the information theoretic approach is able to model natural information processing occurring in a recorded system and describe future directions.

1.2 Overview of approach

This section briefly summarises the information theoretic approach used to meet the above aims. This overview should help indicate why the definitions and past methods described in the literature review are relevant to the project. Further details about the methodology are found in Chapter 3.

The notion of information transfer is in the context of the three component operations to information processing as a whole, first described by Turing (1937). These are information storage, transfer and modification. Since then, information theory has developed to provide definitions for these operations that can be applied to natural distributed systems (Wibral et al., 2014; Mitchell, 1998; Langton, 1990).

Information theory describes information as the reduction in uncertainty about an unknown variable, given knowledge about another variable. Information storage refers to some stable representation that *later* reduces uncertainty about a new state in the system (Wibral et al., 2014). Information transfer can be thought of as an *immediate* reduction in uncertainty about a new state, given the state of another distant variable in the system (Bossomaier et al., 2016). There is some contention about how information modification should be defined in distributed systems (Lizier et al., 2013). However, it can be thought of as the nontrivial synthesis of stored or transferred information, resulting in changes to these that would be unexpected by uniquely considering either.

This project quantifies information transfer using an information theoretic measure called transfer entropy (Bossomaier et al., 2016). In the following context, a process is some observed variable that transforms its state over time. Transfer entropy is designed to capture the reduction in uncertainty about the next state of some process due to the state of a distinct process. Average transfer entropy as a rate can calculate contributions to information transfer between states in the source process to the next state of the target process over the whole observation window. See section 2.2 for more details.

Standard transfer entropy is a pairwise measure. In order to measure multivariate information transfer, a conditional form of transfer entropy is used (Novelli and Lizier, 2021). This captures the reduction in uncertainty about the next state of some process, due to the current state of a distinct process, in the context of one or more states of other source processes. This form can be applied iteratively in a neuronal

population to build an effective network, which reflects multivariate transfers inside the neuronal population over the observation window. The multivariate transfers are only inferred if they are significant, in the sense that they maximally explain the target process' states with the minimal number of sources required. A full description effective networks and the greedy algorithm used for their inference can be found in section 2.5.

The estimation of transfer entropy from recordings of neuronal activity has historically been limited (Shorten et al., 2021). This is because neuronal signals are in effect instantaneous events that occur in continuous time, while estimators for transfer entropy had only been designed to operate on events in discrete time. Shorten et al. (2021) presented a continuous-time estimator for transfer entropy between events which proved more consistent and less biased than discrete time estimators in the neuronal setting.

Further, the continuous-time estimator removes limitations on effective network inference from neuronal populations (Shorten et al., 2022b). The way in which the estimator encodes events allows it to operate on short and long time scales at once, preventing an explosion in state space to search through when estimating the substantial number of conditional probabilities needed by effective network inference. See sections 2.2 and 2.5 for further details of the advantages of the new estimator over discrete-time estimators.

This project therefore evaluates the continuous-time estimator in conjunction with effective network inference. The high dimensional recordings offered by the dataset should leverage the advantages of the new estimator, offering new insights into information processing in visual areas. Expected hierarchical information processing (see section 2.6 is not always found by our results. This leads to a more nuanced view of visual processing wherein the direction of significant transfers are unpredictable on short enough time scales, under unknown informational input. By being the most principled in quantifying information transfer, the above information theoretic approach is suggested to be critical then in future studies of neural information processing.

Literature Review

The goal of this literature review is to reason why the current approach towards capturing multivariate information transfers in recorded neuronal populations from visual areas in mice is indeed best suited to do so, and to point out any limitations. The project goal is framed inside the pursuit towards understanding mammalian information processing and natural sources of computation as a whole. Any relevant background in neuroscience, complex systems theory, and information theory unfamiliar to computer science graduates are also explained in detail.

2.1 Challenges in modelling neural information processing

By first discussing the challenges in modelling neural information processing (and from similar natural sources) when using traditional models of computation, this section introduces how the application of complex systems theory and information theory can address those limitations. Relevant detail about the physiology of neurons are also described when needed.

2.1.1 Turing machines do not model natural computation

Biological systems that process information are particularly good at processing continuous and widely distributed informational input in a highly parallelised manner (MacLennan, 2004; Ben-Jacob, 2009; Kaiser, 2007). This is distinct to most traditional models of computation that process discrete, serialised informational inputs. Consider how information is usually stored, transferred and modified in biological systems. First, information is stored in states that are continuous with each other, where a state is some pattern distributed in space and time (MacLennan, 2004). Second, information is communicated between local elements of the system in a highly decentralised, parallel and often random manner (Ben-Jacob, 2009). Finally, computation for a particular task appears to emerge in the system without external control (MacLennan, 2004).

For example, ant colonies are able to self-organise to select paths towards food that optimise their time spent travelling (Dorigo et al., 1999; Oettler et al., 2013). The information allowing the computation of fastest path is likely contained in several dimensions such as ant positions, pheromone deposits, environmental features and locations of food. These dimensions have continuous states and are distributed in space and time. Information is communicated between ants as they continuously interact with the environment and search for food. Although these interactions appear random to a local observer, the ant colony as a whole is able to utilise the parallelised, distributed information processing by individuals to compute the fastest path to food.

A traditional model of computation such as the Turing Machine inaccurately represents each step to information processing here (MacLennan, 2004). Information in the Turing Machine is encoded by an infinite tape with discrete cells, a register pointing to one of a finite number of discrete states and a finite table of instructions declaring how to change between states. Information is transferred between cells of the infinite tape and the next state of the machine in a noise-free and serial manner, following deterministic rules that have been exactly defined by the table of instructions. Lastly, the task-specification of a Turing Machine is controlled by the human writing the computer.

Wherever information is encoded by the Turing Machine, it is represented by elements that are discrete and distinguishable. This is radically different to the continuous states encoding information found in nature (Crnkovic, 2011). Secondly, transfer of information between local elements in biological systems occurs under the effects of noise unlike the communication in Turing Machines (Findling and Wyart, 2021). As explored by Findling and Wyart (2021) this is not necessarily a limit on computation but can be adaptive in many natural contexts. Further, the steps of the Turing Machine's communication occur serially and deterministically, rather than in parallel and stochastically (Ben-Jacob, 2009). Finally, in biological systems, the fitting of a system's characteristics to a computational task occurs without any apparent external control. Rather than being designed for specific computational tasks like Turing Machines, the ability to compute seems to emerge in response to selective pressures (MacLennan, 2004).

2.1.2 Complex systems theory can model natural computation

This subsection explores the potential for complex systems theory to provide a more suitable model of natural computation. This will afterwards be developed into why information theory should be used to add quantified measures of information processing that are accurate to the setting.

Complex systems theory can be applied to capture many key features of biological systems found in nature (Simeonov, 2010; Crnkovic, 2012). A complex system can be described as a system of simple elements, each following a relatively small set of local rules, that interact over time and produce surprisingly ordered structure at scale (Siegenfeld and Bar-Yam, 2020). The surprise an observer feels is a key feature. These ordered macroscopic structures are characteristically difficult to predict even with prior knowledge of the local rules. Studies in complex systems have found that structures are sensitive to network structure and initial conditions and so this has often been a point of experimentation (Newman, 2003).

When the appearance of ordered structure at scale occurs without external control, it is called emergence (Goldstein, 2011). 'Self-organisation' may also be used when the ordered structure appears over time. Sayama (2015) describes self-organisation as "a dynamical process by which a system spontaneously forms non-trivial macroscropic structures/behaviours over time."

Network models that display self-organisation seem suited to identify the information processing allowed by natural systems (Simeonov, 2010). This is because the appearance of ordered structure at scale can imply information processing. For example, Ben-Jacob (2009) suggests that bacterial populations optimise the density of cells across a nutrient-poor or hard surface by relying on information processing that is distributed across individual cells' interactions with their environment and local signalling to other cells. Their study identified this form of distributed information processing after modelling the population as a complex system. Nodes were individual cells and ordered patterns of cell distribution on new surfaces emerged as a consequence of local environmental interaction and communication rules.

In neuronal populations, emergent structures that underlie functions like decision-making and long-term memory are more challenging to identify, since informational inputs and their processing far exceed the complexity of systems like bacterial populations. Nevertheless, network models following the complex systems approach have often been successful in describing the appearance of ordered structure and pointing to possible associated information processing (Denève and Machens, 2016; Buzsáki and Draguhn, 2004; Wang, 2002; Compte et al., 2000; Brunel, 2000; Dlesmann et al., 1999; Hopfield, 1982; Amari and Kishimoto, 1978). Before elaborating on some examples of successful network models that imply the information processing capabilities of neuronal populations, this review explains the physiology of neurons relevant to information processing. This should first reason why the complex systems analysis is biologically faithful.

2.1.3 Complex systems models are biologically faithful

By examining the physiological interactions between neurons that give rise to state changes and the appearance of ordered structure within the population, it becomes evident that modeling the population as a complex system can faithfully capture its dynamics and information processing capabilities (Hopfield, 1982).

Neurons are specialised mammalian brain cells that have distinct states which are controlled by interactions with other neurons and are observed to encode information (Gerstner and Kistler, 2008a). The following neuronal physiology is summarised from Bear et al. (2015). A neuron will either be in a firing, or non-firing ('resting') state, as characterised by the potential difference across its celullar membrane ('membrane potential'), which is controlled by varying ion concentrations inside and outside of the cell. When firing, the neuron will undergo physiological changes such that smaller molecules ('neurotransmitters') are released by a projection from the neuron called an axon. The axon terminates in a tree-like structure, where branches ('axon terminals') interface with other neurons. Arriving neurotransmitters at a target neuron interface with another tree-like structure called the dendrites. Neurotransmitters cause cellular changes that affect the receiving neuron's membrane potential. Given enough neurotransmitters received, the membrane potential will cross a threshold and the neuron will enter the firing state. The firing state is characterised by a sharp increase in membrane potential, followed by a sharp decrease. This shape is the reason for the firing event being called a spike.

The interface between an axon terminal of one neuron and the dendrites of another is called a synapse. The two neurons are respectively referred to as pre-synaptic and post-synaptic. Synapses vary in 'efficacy' or 'weight', in that a neuron's membrane potential is more strongly controlled by any pre-synaptic spikes that arrive at synapses with higher weight. Varying synaptic efficacy can be due to the effects of neuromodulators (bio-molecules present in the synaptic cleft) or due to larger release of neurotransmitters by the pre-synaptic neuron.

Note that neurotransmitters received when firing or immediately after firing (during a 'refractory period') do not lead to dramatic changes in membrane potential, such that another spike or some kind of prolonged spike occurs.

In contrast to the behaviour described, around a fifth of neurons in mammalian brains release inhibitory neurotransmitters. The effect of these is to decrease the targeted neuron's membrane potential and reduce the chance of firing. The majority of neurons release excitatory neurotransmitters, as described above.

In summary, neuronal states are governed by local interaction rules inside a directed network. When enough excitatory pre-synaptic neurons fire that have high enough synaptic weights, within a short enough time frame, the post-synaptic neuron fires.

Spikes therefore occur at continuous points in time, across spatially distributed neurons (Gerstner and Kistler, 2008b). Spikes will be propagated across the network in parallel, though their communication is under the effects of noise (Gerstner and Kistler, 2008c). This is because synapses are easily modulated by their environment, such that the effects of neurotransmitters on the post-synaptic neuron's membrane potential are noisy.

If an observer considers neuronal state changes the primary method to encode information, then neural information processing occurs as follows. Information is encoded by the patterns of spiking events that are continuous across space and time ('spatiotemporal dynamics') (Brunel, 2000). Information is communicated in parallel, and stochastically between neurons (Dlesmann et al., 1999). Although inputs can appear random and communication occurs under the effects of noise, highly ordered spatiotemporal dynamics emerge (Lizier, 2013). These self-organised structures then inform various mammalian behaviours that have evolved.

Network models under the complex systems theory approach therefore seem especially suited to finding emergent, ordered spatiotemporal dynamics of neurons and pointing to their information processing capabilities. They remain biologically faithful so long as the above rules of spiking dynamics are followed. A famous example is the working memory model of cortical neurons developed by Compte et al. (2000). They used a columnar network architecture, close to that found in the prefrontal cortex of macaque monkeys. Neurons (with spikes controlled by a model called 'leaky integrate-and-fire' (Tsodyks et al., 2000)) were spatially distributed on a ring, such that each could be thought of as encoding an angle on a circle. This was to represent the macaque monkey's response in the prefrontal cortex to a oculomotor delayed response task. In this task, subjects fixate on a central cue on a screen. A peripheral target appears, then disappears after some delay. On disappearance, the subject's eyes will flick to the target position, therefore relying on working memory (Tsujimoto and Postle, 2012). Neurons in the monkey prefrontal cortex preferentially respond to the angle to the target cue (Funahashi et al., 1989), and so this preferential sensitivity was modelled by Compte et al. (2000). They found that, using a connectivity where nearby neurons were strongly excited and neurons further away were weakly excited, the network displayed a sustained response, preferential to the target cue, consistent with the delayed response to the disappearing target in animal studies. This sustained response (increase in firing rate) in a subset of

neurons is an example of ordered spatiotemporal dynamics in a network model that could be said to be encoding information relevant to a task.

Ordered spatiotemporal dynamics found under the complex systems approach can be more intricate than increase in a neighbourhood's firing rate. Consider the model of olfactory processing neurons in the antennal lobe of locusts developed by Rabinovich et al. (2001). This model utilised attractor dynamics. Attractor dynamics refer to high probability or 'low energy' states that exist inside the network activity. Random input to a network with attractor dynamics will perturb the system, but usually ultimately lead to one of the attractor states or 'points.' Rabinovich et al. (2001) found that constructing a network with unstable attractor dynamics allowed unique inputs to lead to consistent trajectories through saddles between attractor points. Each trajectory can therefore be thought of as encoding a specific smell. This was similar to specific path firing of olfactory processing neurons found in locusts. Their model discovered how a particular network architecture leads to biologically feasible ordered spatiotemporal dynamics.

The above models still do not offer quantitative descriptions of information processing. They discover ordered spatiotemporal dynamics in certain architectures and allude to possibly associated information processing. The next subsection describes why an application of information theory can address this limitation.

2.1.4 Information theory helps quantify neural information processing

An information theoretic approach can provide quantitative descriptions of the information storage, transfer and modification, controlled by the spatiotemporal dynamics found in a recorded neuronal population (Lizier et al., 2012; Prokopenko and Lizier, 2014; Lizier et al., 2010). Further, we can quantify changes to the information processing ('information dynamics'), due to varying input, task, environment, or other biological factors. These descriptions also give us a sense of the computational capacity in varying populations, that is, how much uncertainty might be reduced, how quickly, and to what degree of accuracy (Mceliece et al., 1987).

The appropriateness of information theory to modelling information processing in complex systems stems from the fact that the measurements of information merely rely on observing state changes (MacKay, 2022). This allows information to be quantified even in the unique distributed setting (Lizier, 2013). Since a complex systems approach allows precise modelling of network dynamics, an application

of information theory to these models in turn allows us to observe state changes, and estimate quantities of information storage, transfer and modification by the network.

The suitability also arises from information being defined independently of the physical mechanics driving the system. This means information can be measured without accounting for physiology unrelated to state changes. Information theory's definitions are built on the uncertainty an observer has about a variable, not what the variable may represent or the mechanics of how the variable transforms (Cover and Thomas, 2010). Following this approach gains quantified insights into information processing before solving the problem of identifying meaningful states in such a high-dimensional system and why they are meaningful to an animal.

The effective network inference method evaluated in this project has been built using information theoretic measures combined with an understanding of complex systems. The search for quantified distributed information processing by complex systems is the source of inspiration for the effective network model (Novelli and Lizier, 2021). This is further explored in section 2.5.

The next section of this review describes in detail the relevant fundamentals of information theory, to reason why information theoretic measures can be applied to provide quantitative descriptions of information processing in neuronal populations.

2.2 Modelling neural information processing via information theory

This section formally introduces the tools developed by information theory to measure information and argues their effectiveness in quantitatively modelling information processing by a complex system like a neuronal population.

2.2.1 Information theory precisely quantifies information

Information theory aims to effectively identify and quantify uncertainty and information (Cover and Thomas, 2010). Information and uncertainty are described as 'two sides of the same coin'. Information is the reduction in uncertainty about a variable given knowledge of another variable. Uncertainty is the predictability associated with a variable considering its range of possible values or states, where some are more probable than others. In other words, the variable's total entropy leads to how predictable it is to an observer. Information theory uses this notion of entropy to quantify uncertainty.

The most famous measure of entropy in the above sense is Shannon entropy (Shannon, 1948). Consider a variable X that can have any value x drawn from some domain A_X. Each x has some probability $p(x)$ of occurring. The Shannon entropy of X is defined by

$$H(X) = \sum_{x \in A_X} p(x) \log_2 \frac{1}{p(x)}, \tag{2.1}$$

and is measured in bits. Consider that as the domain A_X increases in size, Shannon entropy increases. This aligns with the previous notion of uncertainty. Another two useful properties are that the total entropy is continuous with respect to $p(x)$ and the joint entropy of two variables that are independent $(X \perp Y)$ is equal to the sum of their entropies $(H(X, Y) = H(X) + H(Y))$.

Joint entropy (regardless of independence) captures the uncertainty about multiple variables together, i.e. a multivariate $\{X, Y\}$. It is written as follows (MacKay, 2022).

$$H(X, Y) = \sum_{x \in A_X} \sum_{y \in A_Y} p(x, y) \log_2 \frac{1}{p(x, y)} \tag{2.2}$$

This is analagous to Shannon entropy only summing over joint probability $p(x, y)$, i.e. the probability of values x and y occurring together.

Another way an observer may consider the entropy of a multivariate $\{X, Y\}$, is by assuming they already have knowledge of one of the variables Y. The observer's question becomes what the average uncertainty remaining over values x of X is, after knowing y of Y.

$$H(X|Y) = \sum_{x \in A_x} \sum_{y \in A_Y} p(x, y) \log_2 \frac{1}{p(x|y)} \tag{2.3}$$

Equation 2.3 is called conditional entropy and applies conditional probability $p(x|y)$ i.e., the probability of x occurring given y (MacKay, 2022). This is useful in settings that require the inference of directed uncertainty reduction.

Now suppose that an observer might now naturally quantify information as follows. They ask what the reduction in uncertainty about a variable X is given what they know about another variable Y. This

question may be written as the conditional entropy of X given Y removed from the Shannon entropy of X alone.

$$I(X;Y) = H(X) - H(X|Y) \tag{2.4}$$

Equation 2.4 arrives at the definition of mutual information (MI) between X and Y (Gelfand and I'Aglom, 1959). Interestingly, MI is equivalent to asking how much more uncertainty there is about each variable X and Y alone, compared to the uncertainty there is about them as an unknown multivariate.

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{2.5}$$

In the form presented in Equation 2.5, the Shannon entropies of each variable are summed and their joint entropy is removed. This means MI is a symmetric notion of information ($I(X;Y) = I(Y;X)$). Further, MI may be considered a measure of correlation between two variables and does not infer a causal relationship. The observer only asks questions about uncertainty reduction.

MI may be written in terms of probabilities as follows.

$$I(X;Y) = \sum_{x \in A_x} \sum_{y \in A_y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \tag{2.6}$$

Some statistical properties emerge in Equation 2.6. $I(X;Y) = 0$ if and only if X and Y are independent ($p(x,y) = p(x)p(y)$). As a measure of correlation MI is non-linear, making it more applicable where models of data with fewer assumptions are required.

In natural systems there are often more than two variables. To capture more effects when measuring reductions in uncertainty, an observer should consider the contributions of other variables. For example, consider three $\{X, Y, Z\}$. As previous, an observer may like to measure an uncertainty reduction about X given knowledge of Y. However they now ask if knowing Z too leads to additional uncertainty reduction about X, due to 'synergistic effects' between Z and Y. That is, knowledge of Z and Y increases information compared to knowledge of either alone. An example of an informational network where a synergistic contribution is likely is given in Figure 2.1. Similarly, the observer may ask if
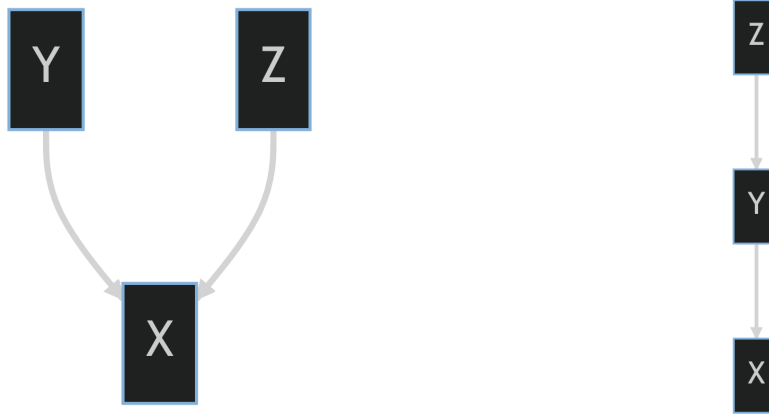
(A) Synergistic contribution between Z and Y to X.   (B) Redundant contribution between Z and Y to X.

FIGURE 2.1. Synergy and redundancy between two informational sources.

knowing Z in addition to Y leads to less uncertainty reduction about X, compared to using only Y, due to 'redundant effects' between Z and Y. A sample drawing where such a redundancy is more likely is also in Figure 2.1. These questions are answered by the conditional mutual information (CMI) measure (Wyner, 1978), defined between two variables X and Y given knowledge of Z as

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z). \tag{2.7}$$

The conditional entropy of X given Y *and* Z is removed from X given Z alone. Should CMI be less than standard MI, that is when $I(X;Y|Z) < I(X;Y)$, this quantifies a redundancy in the information contribution between Y and Z about X. Similarly, when CMI is greater than the standard MI, $I(X;Y|Z) > I(X;Y)$, this quantifies a synergistic contribution between Y and Z about X that cannot be detected by considering either informational source alone. CMI thus provides a building block for measuring interactions between informational sources in a multivariate setting. The next subsection will explain how this can be applied to effectively identify multivariate information transfer in a complex system. For completeness, CMI is written in terms of probabilities as follows.

$$I(X;Y|Z) = \sum_{x \in A_x, y \in A_y, z \in A_z} p(x,y,z) \log_2 \frac{p(x,y|z)}{p(x|z)p(y|z)} \tag{2.8}$$

The entropy measures for both single variables and multivariates, along with the information measures for questions involving two or more variables, are summarised in Table 2.1. Notice that each of these

Measure	Formula
Shannon Entropy	$H(X) = \sum_{x \in A_X} p(x) \log_2 \frac{1}{p(x)}$
Joint Entropy	$H(X, Y) = \sum_{x \in A_X} \sum_{y \in A_Y} p(x, y) \log_2 \frac{1}{p(x,y)}$
Conditional Entropy	$H(X|Y) = \sum_{x \in A_x} \sum_{y \in A_Y} p(x, y) \log_2 \frac{1}{p(x|y)}$
Mutual Information	$I(X; Y) = H(X) - H(X|Y)$
Conditional Mutual Information	$I(X; Y|Z) = H(X|Z) - H(X|Y, Z)$

TABLE 2.1. Fundamental information theoretic measures

fundamental measures leverage the probability distributions of variables to be precise about uncertainty and information. This is advantageous since it allows us to estimate the ground truth distributions from sampled data and in turn estimate uncertainty and information (Cover and Thomas, 2010). Further, there are no assumptions in these measures about the ground truth probability distributions of variables. This helps reduce bias when constructing estimators of information.

2.2.2 Multivariate time series allow quantification of distributed information processing

This section explains how to apply the information theoretic measures to quantify the distributed information processing found in complex systems. Consider that each element in a complex system (in a neuronal population this will be a neuron) varies in some dimension(s) over time. We can represent changes in each dimension for each element over time using a multivariate time series (Novelli and Lizier, 2021).

In information theory, a time series that might encode information is called a 'process' and is denoted by $\mathbf{X}$, where each time index stores a variable X_i (Lizier, 2013). Each process has a 'realisation' for which every variable in the process is sampled such that $X_i = x_i$. A realisation is denoted by $\mathbf{x}$.

In a neuronal population, the state changes over time can be represented by using one process for each neuron. A process may store the changing membrane potential over time or more simply the neuron's changing binary state, where a firing state has instantaneous duration (Shorten et al., 2021). By including some representation of each neuron's spatial location, the resulting multivariate times series captures spatiotemporal dynamics as described in section 2.1.

The question emerges how to effectively identify and quantify information processing within the multivariate time series representation of a neuronal population. In this context of distributed computation, the three key steps to information processing should still be distinguishable; information storage, transfer and modification (Wibral et al., 2014).

For example to identify information storage by a single process, an observer may ask how much information about a process' future state they would have by observing its past history. This can be measured by calculating the MI between all historic variables and the next variable within the process (Lizier et al., 2012). This measure is called active information storage (AIS) and is defined for a process $\mathbf{X}$ with some future variable X_{n+1} and past variables $\mathbf{X}_n^{(k)}$ as

$$A_X(k) = I(\mathbf{X}_n^{(k)}; X_{n+1}), \tag{2.9}$$

where there are n past variables, k of which we would like to observe. AIS quantifies the information afforded to a neuron by its own recent past states. This is information retained by neuronal dynamics, as opposed to more 'passive' forms of information storage, such as information stored via synaptic connection changes. Further, by leveraging MI, the measure takes into account non-linear effects of the past. This allows any form of the neuron's past dynamics to contribute to the stored information and therefore AIS is an unbiased concept of information storage (Lizier, 2013).

Notably, since the measure relies solely on some process X, it has the potential to capture patterns in an input variable that influences X without directly measuring that input variable. AIS measures any information encoded by past dynamics that is relevant to the future state of the process, regardless of the information's original source.

Consider how in contrast an observer might strictly identify the information transfer from a source process to a target process. Their question should seek a directed reduction in uncertainty about a future state of the target process given knowledge of the source process. The observer therefore asks how much information about a future variable in a target process $\mathbf{X}$ can be found from a variable in a source process $\mathbf{Y}$, in the context of the target's past. This can be calculated using the conditional MI between the source process' variable Y_n and the target process' future variable X_{n+1}, conditioned on the target's past $\mathbf{X}_n^{(k)}$. This measure is called transfer entropy (TE) (Bossomaier et al., 2016) and is defined from $\mathbf{Y}$ to $\mathbf{X}$ as

$$T_{Y \to X}(k) = I(Y_n; X_{n+1}|\mathbf{X}_n^{(k)}),\tag{2.10}$$

where there are n past variables in the target process, k of which can be observed.

By conditioning on the target process' past, TE quantifies information after accounting for both redundancy and synergy between the source variable and target past. An observer therefore particularly captures information about state transitions $x_n \to x_{n+1}$ provided by the source variable *only*. Arguably this makes TE an effective measure of 'transfer' from the source variable in particular.

The above definition of information transfer can be extended to account for the effects of redundancies or synergies between the source variable and other processes' variables in the system. This is done by conditioning on the other processes' variables too. This form is called conditional transfer entropy and is defined from $\mathbf{Y}$ to $\mathbf{X}$ as

$$T_{Y \to X}(k) = I(Y_n; X_{n+1}|\mathbf{X}_n^{(k)}, Z_n),\tag{2.11}$$

where Z_n is a variable in another distant process $\mathbf{Z}$ (Bossomaier et al., 2016). Note that any number of variables from distant processes may be conditioned on. By accounting for the new redundancies and synergies, this form proves valuable in identifying information transfer in a multivariate system. For example, consider an informational pathway $\mathbf{Y} \to \mathbf{Z} \to \mathbf{X}$. Conditional TE from $\mathbf{Y}$ to $\mathbf{X}$ would accurately quantify less information transfer given knowledge of a variable in $\mathbf{Z}$, compared to the standard pairwise TE from $\mathbf{Y}$ to $\mathbf{X}$ alone. Indeed, in a complex system with many processes conditional TE would in theory be able to accurately quantify information transfer from each source to each target in the context of every other distant variable. This property of conditional TE forms the foundation of the effective network inference algorithm which evaluates multivariate information transfer in a complex system (Novelli et al., 2019). Further details on how conditional TE is used in effective network inference are in section 2.5.

The final key step to information processing, information modification, has proven difficult to identify in a complex system using the present information theoretic measures and is an area of ongoing research (Lizier et al., 2013). Current work suggests that an observer's question here should be about how information sources are combined leading to unexpected changes to stored and/or transferred information. That is, information modification should be identifiable by some nontrivial synthesis of stored and/or

transferred information that is juxtaposed to the typical reduction in uncertainty due to a 'simple' look up of stored information or a trivial collision-like event of a source with a target. However defining what constitutes a nontrivial synthesis is difficult. Lizier et al. (2013) suggested measuring specifically any synergistic contribution of informational sources to a target variable (that is, excluding sources' unique or redundant effects). While intuitively well principled, this leads to a notion of information modification that is not mutually exclusive with information transfer, since measuring the latter has so far included synergistic effects within it. There is no clear consensus on whether information modification may truly be considered mutually exclusive with transfer or not (Flecker et al., 2011).

In contrast, the explained definitions of active information storage and information transfer in the context of distributed information processing are well established. AIS was estimated from voltage-sensitive dye imaging data of neuronal populations in the cat and could identify neuronal preference for certain stimuli, neuronal surprise upon stimulus change, and neuronal encoding of an ongoing abstract stimulus in spite of local random changes to the stimulus (Wibral et al., 2014). Meanwhile, TE identifies expected information transfer in modelled spiking neural networks with known causal sources (Shorten et al., 2021). Moreover, TE was recently used to quantify information transfer in developing biological neural cell cultures and results reflected informational specialisation of neurons over time (Shorten et al., 2022a). Section 2.4 explains how transfer entropy was able to be estimated in these studies from recordings of spatiotemporal dynamics in continuous-time.

Measure	Formula
Active Information Storage	$A_X(k) = I(\mathbf{X}_n^{(k)}; X_{n+1})$
Transfer Entropy	$T_{Y \to X}(k) = I(Y_n; X_{n+1} | \mathbf{X}_n^{(k)})$
Conditional Transfer Entropy	$T_{Y \to X}(k) = I(Y_n; X_{n+1} | \mathbf{X}_n^{(k)}, Z_n)$

TABLE 2.2. Information theoretic measures in distributed information processing

The established measures of information processing that can be applied to complex systems are summarised in Table 2.2. A limitation to the present measures is that it is sometimes unclear how to apply them to emergent macroscopic structures that have been observed to be relevant to information processing. For example, the information stored by a heteroclinic cycle (Rabinovich et al., 2001) or the information transferred from a bump attractor to some target neuron(s) (Wimmer et al., 2014). These emergent structures can form a coherence beyond which neurons make up those structures (Keane and

Gong, 2015), and so uncertainty reductions due to the structures themselves cannot be estimated by considering the state changes of specific neurons alone. Instead, estimating information here would require an understanding of the possible state configurations of the macroscropic structures, and therefore specific tailoring to each context.

The following section will address how other approaches to understanding neural information processing may aid the development of information theoretic measures.

2.3 Complementary methods in studies of neural information processing

This section outlines various approaches to understanding how neuronal populations process information and discusses how an information theoretic approach, as described in section 2.2, complements these.

2.3.1 Generative modelling of spatiotemporal dynamics

Using techniques borrowed from statistical physics, spatiotemporal dynamics in the brain can be generatively modeled. For example, O'Keefe and Recce (1993) famously designed a network model that was able to generate spatiotemporal dynamics similar to hippocampus place cells in rats. They reasoned that the phase position of a place cell's spike with respect to background oscillation in the nearby population encodes the rat's location in real space. This kind of approach gives us a mechanistic understanding of neural systems, while writing analogies for how computation might be occurring by identifying ordered, emergent spatiotemporal dynamics that are controlled by parameters in the generative models. Further, when the generative models are successfully physically faithful to the underlying system, model parameters in living neural systems can be inferred from neuroimaging, and predictions can be made about what spatiotemporal dynamics might emerge in unique conditions (Miller, 2016).

Describing computation that might be allowed by generated or found spatiotemporal dynamics often remains speculative although theories can be strengthened by behavioural studies, as was the case in (O'Keefe and Recce, 1993). Importantly, there is a distinction between explaining what spatiotemporal dynamics are present in a neuronal population and explaining what information dynamics are present. An information theoretic approach allows the latter to be quantitatively modelled. It follows naturally that we should search for information theoretic measures that can capture information storage, transfer and modification by the various emergent structures found via generative modelling of spatiotemporal dynamics. Well-established emergent structures include:

(1) Oscillators (Strogatz, 2000)

(2) Attractors, characterised as fixed-point, cyclic, Milnor measure or strange (Milnor, 1985)

(3) Localised bumps in mean-field models (Wimmer et al., 2014)

(4) Phase transitions between states in balanced networks or networks poised at criticality (Brunel, 2000; Fontenele et al., 2019; Langton, 1990)

An information theoretic approach to quantifying information processing by these structures requires an understanding of their possible state-configurations. Estimators of information storage, transfer or modification therefore need to be tailored to each context.

The advantage of adding quantified measures of information to generated models of spatiotemporal dynamics was highlighted by Beggs and Plenz (2003). Their study showed that the branching parameter in feedforward spiking neural networks could be poised at criticality to optimise the mutual information between stimulus and neuronal response in an output layer, following an avalanche. By finding a quantifiable measure of the utility of criticality Beggs and Plenz (2003) appreciably added to our understanding of neural information processing. Future work under this approach should similarly be complemented by information theory, while the design of information theoretic estimators should follow found emergent structures.

2.3.2 Statistical inference of brain function

Statistical inference provides a generic framework for descriptive analysis, hypothesis testing, parameter prediction and data-driven discovery and generalisation. It is therefore readily applied to functional neuroimaging to test hypotheses and build models about brain function (Bzdok and Yeo, 2017). While brain function usually involves computation of some informational inputs for a task, statistical inferences can be made about the relationship between spatiotemporal dynamics and brain function without quantifying information at all. For example, Neuper and Pfurtscheller (2001) identified how varying powers of frequency bands recorded by electroencephalography are correlated to effects on attention during a variety of cognitive and motor tasks. By measuring behavioural components, brain function is modelled without quantification of information processing. Statistical inference is also suited to testing the effects of neural components beyond spatiotemporal dynamics on brain function. For example, varying neuromodulator levels, drug treatments and genomes (Bzdok and Yeo, 2017). Historically, identifying statistical significance has been limited to pairwise comparisons despite neuronal populations being a complex system that leverages multivariate effects (Geerligs et al., 2016). Developments are being made

however towards more nuanced multivariate inference (Chen et al., 2021), similar to the development of information theoretic inference outlined in this review.

Statistical inference tools that are often applied to functional neuroimaging include:

(1) Generalised linear models to predict neural response to experimental conditions (Stapleton et al., 2006).

(2) Granger causality for the estimation of causal influences on spatiotemporal dynamics (Bressler and Seth, 2011).

(3) Pearson correlation and variants to test for functional connectivity between regions of interest (Geerligs et al., 2016).

(4) Bayesian methods for statistical inference beyond null hypothesis significance testing. These are wide-ranging but can be applied to identifying multivariate functional connectivity, parameter estimation in models of spatiotemporal dynamics, stimulus-response prediction and stimulus reconstruction (Chen, 2013).

The information theoretic approach is an entirely separate framework that is principled by defining and quantifying information. It appreciably adds to traditional statistical inference whenever precise changes in information processing should be identified, which is at least the case in developing a theory of natural computation (Crnkovic, 2011).

One study by Sharpee et al. (2006) took advantage of information theory's ability to quantify information to improve prediction of neural response which is the average spatiotemporal activity as a function of time, given some set of stimuli. They found that a linear filter of stimuli that maximised mutual information between neural response and the output of the filter could predict how neurons in the visual cortex adapted to stimulus. Quantifying the amount of information encoded by high order sensory neurons allowed a nuanced statistical model of their spatiotemporal dynamics, highlighting the complementary nature of these approaches.

This project capitalises on the information theoretic framework to first estimate information transfer inside neuronal populations recorded in continuous time and then infer a network model of multivariate information transfer. The estimation of transfer entropy in continuous time was only recently developed by Shorten et al. (2021) and is explained in the next section, followed by section 2.5 on effective network inference.

2.4 Transfer entropy estimation between spikes in continuous time

This section presents the latest developments in estimating transfer entropy between processes that store events in continuous-time, such as neural spike trains.

As argued by Shorten et al. (2021), the problem in directly using the formulation of TE in Equation 2.10 from section 2.2 is that it forces representations of spikes to be at discrete time indices which requires some form of binning. Usually, this is done by choosing a time bin width and detecting whether or not a spike occurred in each bin. This loses a representation of more than one spike occurring within a bin's time-frame and of the exact time stamp of each spike.

In the 'discrete-time' form, a measure for the average TE as a rate from variables in $\mathbf{Y}$ to a variable X in $\mathbf{X}$ can be designed as follows (Spinney et al., 2017).

$$\dot{T}_{Y \to X} = \frac{1}{\Delta t} I(X_t; \mathbf{Y}_{<t} | \mathbf{X}_{<t}) = \frac{1}{\tau} \sum_{t=1}^{N_T} \ln \frac{p(x_t | \mathbf{x}_{<t}, \mathbf{y}_{<t})}{p(x_t | \mathbf{x}_{<t})} \tag{2.12}$$

In Equation 2.12, X_t is the current variable of the target while $\mathbf{X}_{<t}$ and $\mathbf{Y}_{<t}$ are the target and source histories. x_t, $\mathbf{x}_{<t}$ and $\mathbf{y}_{<t}$ are the respective realisations, i.e. observed samples from each process. Δt is the length of time between samples, τ is length of the time series and N_T is the number of time samples.

In the discrete-time form, the rate converges to true TE with infinitely small bin width when the underlying process is discrete. However, when the underlying process is continuous, the estimated value is not guaranteed to converge to the true value in the presence of infinite data (Shorten et al., 2021). Since spikes can occur at any point, neural spike trains do indeed form continuous processes.

Shorten et al. (2021) further highlighted that an estimator in discrete time is not able to estimate TE in small time scales and large time scales at once. Choosing a small bin size means that a smaller embedding length, i.e. the number of historic values we would like to observe in each process, is required to avoid the risk of undersampling. Conversely, choosing a large embedding length can only be done with large bin size. Being unable to observe small and large time scales at once is particularly an issue when the tool being used to observe spikes has a high sampling rate. Taking advantage of a high sampling rate would prevent estimating over larger time scales. Worse, Shorten et al. (2021) empirically showed that choice of bin size leads to inconsistent estimates.

This motivates an estimator for transfer entropy that does not require binning and is instead able to rely on observations of spikes in continuous time. Shorten et al. (2021) presented the following 'continuous-time' formulation for the average transfer entropy rate from variables in $\mathbf{Y}$ to a variable in $\mathbf{X}$.

$$\dot{T}_{Y \to X} = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{i=1}^{N_X} \ln \frac{\lambda_{x|\mathbf{x}_{<t}, \mathbf{y}_{<t}} \left[\mathbf{x}_{<x_i}, \mathbf{y}_{<x_i} \right]}{\lambda_{x|\mathbf{x}_{<t}} \left[\mathbf{x}_{<x_i} \right]} \tag{2.13}$$

In Equation 2.13, τ is the time series length and N_X is the number of events in the target process. In the denominator, $\lambda_{x|\mathbf{x}_{<t}}$ is the instantaneous firing rate of the target process at sample x when conditioned on this realisation's history. The $\left[\mathbf{x}_{<x_i} \right]$ indicates to only sample history at the time points x_i of the events in the target. Equivalently in the numerator, the λ expression is the instantaneous firing rate of the target process at sample x when conditioned both on target and source history. The observer only samples the history of target and source at the time points x_i of events in the target.

(Shorten et al., 2021) formally showed that the continuous-time estimator converges to the true rate of transfer entropy in the presence of infinite data when the underlying system is continuous. Convergence also appears faster than in the discrete-time estimator and with less bias that was inherent due to binning.

Further, a computational advantage arises. This is because the continuous-time formalisation of the average rate of TE in Equation 2.13 allows encoding spike event times in inter-spike intervals, such that an observer might jointly estimate in small and large time scales. Finally, computation is also improved by only needing to sum over spiking events rather than over every time step.

Promising results have been found using the continuous-time estimator. Shorten et al. (2022a) was able to infer a functional network, i.e. a directed graph containing significant pairwise information transfer between nodes, from in vitro imaging of developing neural populations. The functional network of information flow characterised nodes according to ratios of in-degree and out-degree. Shorten et al. (2022a) classified these nodes as information 'sinks', 'mediators', and 'receivers', and found that the specialised roles were maintained from a surprisingly early point in development. Furthermore, the inferred information flow network was aligned with expectations from a theory of spike-timing dependent plasticity (Caporale and Y, 2008) and a generative model using Izhikevich neurons (Izhikevich, 2003).

The continuous-time estimator has so far been able to quantify information processing in neuronal populations in vitro, delivering results that align with the expectation that early developing cell cultures

should improve their capacity for information transfer over three to four weeks. This project now lever-ages the high sampling rate and fine spatial scale offered by modern extracellular imaging to apply the continuous-time estimator to neuronal populations in vivo. Pairwise information transfer between spike trains will be estimated following the approach used by Shorten et al. (2021).

The estimator will also be used in combination with effective network inference which finds a network model of multivariate information transfer. The new estimator overcomes limitations in inference from spike trains, not only by operating in continuous time, but also by reducing dimensionality so that TE can be estimated even when conditioning on a substantial number of source processes. The following section addresses details of the algorithm that highlight the significance of conditioning on a large number of source processes.

2.5 Effective network inference

This section formally describes what effective networks are and their usefulness as a tool for evaluating multivariate information transfer.

An effective network represents the multivariate information transfers that occurred inside a complex system during some window of observation (Novelli et al., 2019). The network model is a directed graph, where each node is a process inside the complex system. During observation, each node's parent set formed a significant rate of multivariate information transfer to that node. The notion of significance here is twofold (Shorten et al., 2022b). Firstly, the observed activity of the parent set for each node maximally reduced uncertainty about target states. Any removal of source variables underlying a parent set would increase uncertainty about their target. Secondly, each parent set is minimal in the sense that addition of further source variables would not reduce uncertainty any further.

Since an effective network reflects information transfer during observation, it may be considerably dif-ferent to the underlying structural network. That is, the same neuronal population with relatively fixed connectivity will yield varying effective networks under different informational inputs (Friston, 2011). Effective networks therefore offer insight into how a network structure may process information differ-ently under varying tasks or conditions.

The alternative, naive approach to inferring information flow in a neuronal population would be to measure the pairwise TE between all neuronal processes (Bossomaier et al., 2016). For example an

observer may determine the significance of each pairwise transfer, i.e. probability of obtaining this value from a null distribution where correlation between source and target are destroyed, then add directed links for low enough p-values. However, a pairwise approach (sometimes called 'functional network inference') does not measure synergies or redundancies between sources. The latter are quantified by conditional TE, as explained in section 2.2.

Conditional TE could be leveraged by testing for significant transfer between each source-target pair, conditioned on all other distant variables (Novelli et al., 2019). However, even with the computational advantages of the recent estimator developed by Shorten et al. (2021), conditioning on all other distant variables at once quickly increases dimensionality and the risk of under-sampling.

This arrives at the current approach to find the minimal parent set for each node that maximally explains their target's state transitions. Importantly, the minimal parent set is well principled to identify the 'effective coupling' between source and target variables in a multivariate setting, as was described by Friston (2011). Significance testing under this approach requires measuring TE for each source-target pair conditioned only on other variables underlying the target's parent set. The challenge is to first find the parent set to condition on. A greedy approach adding source variables to a conditioning set, explained in the next subsection, has proven able to to converge to the true parent set for each target with infinite data and stop choosing parents when the statistical power of the data is exhausted (Lizier and Rubinov, 2012).

The continuous-time estimator for TE is well suited to the greedy approach (Shorten et al., 2022b). It is able to operate on small and large time-scales at once without the need to embed process histories with both small and large bin widths. Multi-bin width embedding leads to an explosion of the state space needed to estimate conditional probability distributions for each target process, as the number of variables to condition on increases. The need for all time scales is particularly true in spike trains where correlation between events extend over hundreds of milliseconds (Rudelt et al., 2021). Simultaneously, correlations to neural function can be inferred at the sub-millisecond level.

This project applies the continuous-time estimator in conjunction with the below greedy algorithm for effective network inference (Shorten et al., 2022b). In accordance with the new estimator, process histories are embedded using inter-spike intervals.

2.5.1 Greedy algorithm

Given a set of spike trains $\mathcal{R}$, this algorithm finds a minimal parent set of source processes $\mathcal{S}_X$ for each target $\mathbf{X}$ in $\mathcal{R}$ that can maximally explain target state. Inference of parent set for each target is independent and may be completed in parallel.

For each target $\mathbf{X}$, first determine the set of target history intervals h_X that will be used to condition on when calculating multivariate information transfer. This can be thought of as finding the target embedding that would maximise its active information storage.

(1) Initialise h_x to store the most recent target interval.

(2) Estimate the average MI a between target state and the interval previous to the earliest interval in h_x, conditioned on intervals already in h_X.

(3) If a is significant (tested using a surrogate generation method described in subsubsection 2.5.1.1), then add the tested interval to h_X and go back to step 2.

(4) Otherwise, stop adding to the number of target history intervals.

Once the embedding for the target history has been chosen, we can begin determining intervals from all other source processes that offer significant multivariate transfer to this target.

(1) Initialise a conditioning set c_X which for each process other than the target will store the source intervals that offer a significant rate of multivariate transfer.

(2) Iterate over all candidate source processes. For candidate source $\mathbf{Y}$, estimate the average rate of TE between target state i and the most recent interval from $\mathbf{Y}$ not already in c_X, conditioned on all intervals already in c_X. Note that the target history is embedded using h_X.

(3) Select the candidate source interval which offered the maximum conditional TE. Test its significance using the maximum statistic test described in subsubsection 2.5.1.2.

(4) If significant, add this source interval to c_X. Go back to step 2.

(5) Otherwise, stop adding source intervals to the conditioning set.

Once we have a conditioning set for this target, we can prune from the added source intervals. We prune until we are only left with source intervals with significant multivariate transfer to the target, when conditioning on all intervals in the conditioning set.

(1) Iterate over all sources that had an interval in c_X. For source $\mathbf{Y}$ with an interval in c_X, estimate the average rate of TE between target state i and the last added interval from $\mathbf{Y}$ in c_X. Test the significance of this estimate using the surrogate generation method explained in subsubsection 2.5.1.1. If insignificant, remove this source interval from c_X.

(2) As long as there was at least one insignificant rate of multivariate transfer found between a source interval in c_X and target state, go back to step 1.

(3) Otherwise, stop pruning source intervals from the conditioning set.

The parent set $\mathcal{S}_X$ is finally determined by including any processes that had an interval left in c_X.

2.5.1.1 Testing for non-zero TE using surrogates

To test the significance of an obtained TE value with respect to the state space of realisations for source, target and conditional processes, a surrogate generation method may be used. The following approach is summarised from Shorten et al. (2021).

Generate $N_{surrogates}$ surrogate processes that conform to the null hypothesis of zero information transfer, then estimate TE on each of the generated surrogates. The p-value used for significance testing is then the proportion of surrogate estimates that are greater than or equal to the original obtained TE value. The p-value therefore estimates the probability that we would observe a value of TE greater than or equal to the original estimation, under the null hypothesis of observing zero TE.

2.5.1.2 Maximum statistic test

The standard surrogate generation approach to test for non-zero TE fails when adding source intervals to the conditioning set. This is because finding conditional TE of all candidate source intervals, before determining the significance of the maximum estimate leads to a multiple comparisons problem (Shorten et al., 2022b).

In order to compensate for this, for each candidate source interval, first generate $N_{surrogates}$ surrogate processes as normal. Then, for $i \in \{1, 2 \ldots N_{surrogates}\}$ find the maximum estimate for each i across all candidate source intervals. The null distribution is therefore made up of $N_{surrogates}$ maximum values, and the p-value calculation continues as normal. This essentially replicates the selection of maximum candidate source, inside the generation step.

This concludes this review's explanation of effective network inference. In this project, this approach will be used to search for multivariate transfer that explains effective coupling between layers in the visual areas of the mouse brain. By being principled in the distributed information processing found in dynamic complex systems, this approach should appreciably increase understanding of visual processing by adding a quantified measure of transfers. The next section presents where the expectations about visual information processing that are evaluated in this project come from.

2.6 Information processing in visual areas

This section summarises the consensus on visual information processing, which has been arrived at using statistical inference of brain function for the most part and sometimes generative modelling of spatiotemporal dynamics (Marr, 1982; Adelson and Bergen, 1985). As outlined in section 2.3 these general approaches are not principled in quantifying information. Statistical inference of functional and effective connectivity can offer evidence of information processing (Friston, 2011) but without quantifying information storage, modification or transfer. Furthermore, in deeper brain areas where the spatial resolution of functional neuroimaging has been limited, theories of visual processing might turn to structural imaging (Bzdok and Yeo, 2017). Structural imaging is poorly principled to reflect information processing since the same structure can give rise to varying spatiotemporal dynamics depending on perturbations (Friston, 2011).

The primary visual cortex (V1), like all areas in the neocortex, is organised into six functional layers of densely connected neurons (Harris and Shepherd, 2015). These layers can be thought of as horizontal sheets. The neocortex is further organised, by vertical columns of excitatory neurons interconnected across layers, called local excitatory circuits. Functional connectivity inside local excitatory circuits is well-understood but how inhibition is controlled is less easy to infer from functional imaging (Logothetis and Wandell, 2004).

The visual informational pathway originates at the retina and projects to V1 via a subcortical structure called the thalamus (Bear et al., 2015). The dominant projection from the thalamus to V1 is to layer 4 and broadly maintains a topographic organisation from the retina. This phenomenon is called 'retinotopy' and is thought to lead to feedforward convolution-like processing of the visual field (Lamme and Roelfsema, 2000). For example in cats, neurons from the lateral geniculate nucleus of the thalamus (LGN) have receptive fields that are spatially aligned and circular; they fire when a stimulus enters or

leaves some location in the visual field. The layer 4 cells in turn respond to orientations of stimuli by integrating LGN inputs (Clay Reid and Alonso, 1995).

Thalamic inputs to V1 fall under two broad classes, originating from core-type and matrix-type neurons (Jones, 2001). Core-type are thought to contain 'rapid' sensory information that leads to quick feedforward processing and target layer 4. Matrix-type instead target layers 1 and 5, though the quality of information being processed is poorly understood (Clascá et al., 2012). Visual input is thought to be the only sensory modality being processed by layer 4, since there is no variability that would be expected due to modulation by other modalities (Hansen et al., 2012).

Layer 4, in turn, asymmetrically projects heavily to layers 2, 3 and 5 (Harris and Shepherd, 2015). Harris and Shepherd (2015) reviews how cells in these layers are specialised in terms of morphology, physiology and average spike rate, which leads to the assumption of their information processing specificity. Layer 1 contains inhibitory neurons and layers 2 - 6 contain intratelencephalic (IT) neurons. Further, pyramidal tract (PT) neurons are found in layer 5B and corticothalmic (CT) neurons are contained by layer 6.

Within a local excitatory circuit, each class of neurons forms recurrent connections with themselves (Thomson and Lamy, 2007). Between classes, connections are asymmetrical. IT neurons consistently form 'exits' in a column, projecting to other cortical areas.

IT neurons in layers 2 and 3 are thought to code sparsely in temporally and spatially specific bursts (Petersen and Crochet, 2013). This is in spite of being heavily innervated. Petersen and Crochet (2013) suggests this is due to a high stimulus selectivity and that a reduction in state space due to sparse coding is useful for associative learning.

After being innervated by layer 4, layers 2 and 3 in turn have a major descending axonal projection to layers 5A and 5B. PT neurons in Layer 5B are thought to exhibit dense coding, since they are highly recurrent and have high firing rates (Peters and Kara, 1985). Harris and Mrsic-Flogel (2013) suggested that dense coding allows encoding information using variation of firing rates.

Layer 5 in turn projects to subcerebral regions, back to the thalamus, as well as to other ipsilateral cortical areas (Harris and Shepherd, 2015). There are few return projections to the local excitatory circuit, though some feedforward to layer 6. Feldmeyer (2012) suggests that layer 5 can be thought of as the most downstream part of the circuit, integrating results of processing by the local excitatory circuit

with thalamic input and broadcasting results, mainly to subcerebral structures and also other cortical areas.

Layer 6 is somewhat innervated by layer 4, layer 5, and even the thalamus but is dominated by long-distance inputs from across the neocortex and claustrum (the 'wall' of the neocortex) (Stepanyants et al., 2009). IT neurons here have horizontal projections that exit their column and even V1. CT neurons here project back to thalamic nuclei though are thought to be modulatory rather than drivers (Guillery and Sherman, 2002). Rather than closing the thalamocortical loop, CT neurons in layer 6 are thought to integrate long-range signals to modulate thalamocortical activity.

Within the local excitatory circuit however, specialised function of layer 6 CT neurons is difficult to identify. They have sparse, locally ascending projections to layers 5A and 4 (Fitzpatrick, 1996). They are remarkably silent during various tasks, and have significant delays in their projections (up to 30 ms). Vélez-Fort et al. (2014) found that in mice these neurons predominantly inhibit all other layers via inhibitory neurons called parvalbumin-positive interneurons.

For completeness, note layer 1 is sparsely populated by neurons and largely contains inhibitory neurons (Cruikshank et al., 2012). It is not heavily innervated by any particular layer but receives inputs from thalamic matrix-type neurons and nearby layer ITs. Interestingly, layer 5 ITs have thin-tufted apical dendrites that extend to layer 1, allowing layer 1 to innervate layer 5. There may be sparse projections from layer 1, to all other layers, in each column.

In summary, functional connectivity within local excitatory circuits suggest hierarchical information processing, first from layer 4 to 2/3 and 5, then from 2/3 to 5. Broadcasting out of the column probably occurs in layers 5 and 6. Integration probably occurs in layer 6. Latency of functional activity between layers suggests that feedforward passes of information through this hierarchy occurs rapidly in response to visual stimulus, while recurrent transfers occur slowly (Lamme and Roelfsema, 2000). The rapid 'sweep' leverages tuning by successive neurons in the hierarchy, for example to orientation, direction and motion. The slow feedback is thought to be required for dynamic updates to tuning allowing selective visual attention, for example to separate objects from background scenery.

This review of visual areas informs the hypotheses about visual information processing that are presented in Chapter 3. This will followed by Chapter 4 which tests expectations before evaluating the review's

overarching argument. This is that the information theoretic approach guided by complex systems theory appreciably adds to understanding of neural information processing beyond functional analysis, by quantifying information.

Experiments

This chapter describes what the hypotheses about visual information processing were and how they were tested within the available data.

3.1 Hypotheses

There are three major hypotheses (H1-3) that arise from our survey of the literature on information processing in the visual areas of the mammalian brain. These are listed below.

Note that layers 2 and 3 inside the mammalian cortex are anatomically similar and are labelled together in our dataset. They will be referred to as layer 2/3.

H1. Inside local excitatory circuits, otherwise called 'columns' of the primary visual cortex, layer 4 is a major information source to other layers. This would be supported by:

  (1) High proportion of significant pairwise transfer entropy from layer 4 to all layers.

  (2) Low ratio in layer 4 of incoming to outgoing edges in an effective network.

H2. Layer 2/3 forms a sparse-coding integration station, primarily driven by layer 4. Specialised neurons in layer 2/3 are thought to code sparsely in that they fire less often but in spatiotemporally specific bursts. Sparse coding might correlate to a high information transfer per firing event. This would be supported by:

  (1) High proportion of significant pairwise transfer entropy from layer 4 to layer 2/3.

  (2) High average transfer entropy per firing event for connections originating in layer 2/3.

  (3) Higher proportion of significant edges from layer 4 to 2/3 than from other layers to 2/3 in an effective network.

  (4) High proportion of significant transfers within layer 2/3.

H3. Layer 5 is a dense-coding integration station and information sink primarily driven by layer 4 and somewhat driven by layers 2/3. Specialised neurons in layer 5 are thought to code densely, that is firing often in less specific spatiotemporal locations. These neurons offer few return projections to other layers, are highly recurrent and have the highest firing rate of all excitatory cell classes. This would be supported by

  (1) High proportion of significant pairwise transfer entropy from layer 4 to 5, and from layer 2/3 to 5.

  (2) High average rate of transfer entropy within layer 5, compared to within other layers.

  (3) Highest proportion of edges from layer 4 to 5, with some edges from layer 2 & 3 to 5 in an effective network.

Some minor hypotheses (H4-7) that were either less novel or expected to be challenging to evaluate within our dataset are listed below.

H4. Each layer in a local excitatory circuit of the primary visual cortex forms a computationally specific module. This is suggested by most layers containing physiologically specialised cells with recurrent connections. This would be supported by

  (1) High rate of average transfer entropy from each layer to itself.

H5. Layer 5 broadcasts outside of the local excitatory circuits in the primary visual cortex to visual association areas. Layer 5 pyramidal tract neurons are known to physically target ipsilateral cortical areas. This would be supported by:

  (1) Some proportion of significant pairwise or multivariate transfer entropy from layer 5 in a V1 probe to neurons sampled by a probe through a visual association area.

H6. Layer 6 broadcasts information from the column to horizontally distant targets. This is suggested by intractectal neurons in layer 6 being driven by intratectal neurons in layers 5 and 6. Layer 6 is then observed to target other columns horizontally as well as visual association areas outside the primary visual cortex. This would be supported by:

  (1) High proportion of significant pairwise transfer entropy from layers 5/6 to 6 inside V1 columns.

  (2) High proportion of significant edges from layers 5/6 to 6 inside effective networks of V1 columns.

  (3) Some significant pairwise or multivariate transfer entropy from layer 6 in one column to layers 5 and 6 in other columns of V1.

(4) Some significant pairwise or multivariate transfer entropy from layer 6 in V1 to neurons sampled by a probe through a visual association area.

H7. Layer 6 in the primary visual cortex integrates information from distant sources. The reasoning for this hypothesis is that corticothalamic neurons in layer 6 are primarily innervated by neurons in higher order ipsilateral cortical areas and not neurons inside the same column. However, corticothalamic neurons have often been observed to be quiet in activity for entire tasks and could be considered an 'optional' integration step. This hypothesis would be supported by:

(1) A high proportion of pairwise or multivariate transfer entropy from a visual association area to layer 6 in the primary visual cortex,

(2) Possibly relatively low rate of transfer entropy from layer 6 to itself, suggesting corticothalamic neurons were quiet during observation.

3.2 Data

The dataset used in this project was first published by Stringer et al. (2019). The original dataset contains the spike times of several hundred cortical neurons that were recorded via high density extracellular imaging using silicon probes inserted into three mice, in vivo. These probes are a somewhat recently developed technology called 'Neuropixels' and offer 384 electrode sites on a 70 x 100 x $20\mu m$ shank (Jun et al., 2017). Stringer et al. (2019) inserted eight Neuropixels probes in each mouse to target various cortical areas. During recording, the mice were in a light-isolated enclosure, head-fixed by an apparatus, with front-paws free to move on a rolling ball. They were presented with three computer screens though Stringer et al. (2019) only published Neuropixels recordings during spontaneous activity. This is when the screens were black.

Spike times in the dataset are already spike-sorted, meaning that they have been labelled as belonging to different neurons. Each labelled neuron is already estimated to be located inside an anatomical brain area. Note that the mammalian cortex has a layered structure, where most areas of the cortical plate have six identifiable layers within them. The high resolution of Neuropixels probes allows each neuron to be estimated as being located inside a particular layer of each cortical area. Layers 2 and 3 are similar in anatomy and have been labelled together as layer 2/3.

For this project, we selected probes that recorded neurons from the visual areas of the mouse brains. This led to two mice being used (named Waksman and Krebs in the original dataset) each with four probes selected. The locations of these probes are listed in table 3.1. Recorded layers are also noted in table 3.1 on the right.

In summary, the dataset used is made up of arrays (n=446) containing timestamps of spike events, where each array is belongs to some neuron that is located inside a layer of some cortical area related to visual processing. The number of neurons recorded by each probe are reported in table 3.1. Note that probe 7 in Waksman goes through two areas and so is intentionally listed twice.

Mouse	Probe	Brain structure	Allen Brain Atlas Acronym
Waksman	Probe 3 (n=104)	Left Primary Visual Cortex	VISp[2/3,4,5,6]
	Probe 6 (n=6)	Right Anterior Visual Area	VISa[4,5]
	Probe 7 (n=66)	Left Rostrolateral Visual Area	VISrl[4,5,6]
	Probe 7 (n=77)	Dorsal LGN of Thalamus	LGd[-co,-sh]
	Probe 8 (n=60)	Right Primary Visual Cortex	VISp[2/3,4,5,6]
Krebs	Probe 3 (n=52)	Right Primary Visual Cortex	VISp[2/3,4,5,6]
	Probe 4 (n=44)	Left Primary Visual Cortex	VISp[5,6]
	Probe 7 (n=10)	Right Primary Visual Cortex	VISp[5,6]
	Probe 8 (n=27)	Left Primary Visual Cortex	VISp[4,5,6]

TABLE 3.1. Locations of Neuropixels probes used in each mouse.

3.3 Methods

3.3.1 Pairwise transfer entropy estimation in continuous time

To investigate the hypotheses about information transfers inside columns of the primary visual cortex, first pairwise transfer entropy was estimated between neurons inside each probe going through this area. The estimator for transfer entropy between events in continuous time as first reported by (Shorten et al.,

2021) was used. It is important to note that the continuous-time estimator returns a rate of transfer entropy between two spike trains. It estimates the contribution of each source-target spike pair to the total information transfer and divides this by the length of observation. The rate of transfer entropy reported therefore captures information transfer over the whole observation window.

To standardise the statistical power drawn from each pair of neurons inside a probe, exactly $n_{target\ spikes}$ spikes were sampled from each target neuron. $n_{target\ spikes}$ for each probe are listed in table 3.2 and were chosen by considering the minimum number of spikes recorded across most neurons in each probe. Source spikes were sampled such that they occurred only within the window of target spikes. If the number of source spikes fell below one hundred for any pair of neurons, the pair was skipped.

Mouse	Probe	$n_{target\ spikes}$
Waksman	Probe 3 VISp	1500
	Probe 6 VISa	800
	Probe 7 VISrl	1000
	Probe 7 LGd	3000
	Probe 8 VISp	1000
Krebs	Probe 3 VISp	995
	Probe 4 VISp	700
	Probe 7 VISp	700
	Probe 8 VISp	800

TABLE 3.2. Number of target spikes used during pairwise TE estimation in each probe.

The number of surrogate processes to generate to test the null hypothesis of no transfer between a source and a target, $n_{surrogates}$, was chosen to be 100 for all source-target pairs.

In order to indicate the relative degree of information transfer between layers, the proportion of significant transfers to the total number of tested pairs of cells across layers was calculated. The average rate of transfer between layers was also calculated after zeroing source-target pairs with non-significant transfers.

In order to test the sparse vs dense-coding nature of layers, the rate of transfer per source spike was also calculated. This was done by dividing the total contribution of each source spike with each target spike by the number of source spikes. Source-target pairs were then averaged across layers again, after zeroing non-significant transfers.

To investigate the hypotheses about information transfers between columns inside the primary visual cortex, the above methods for pairwise transfer entropy estimation was repeated, only using pairs of cells between probes that were in V1 of the same mouse.

In addition, pairwise transfer entropy was estimated between primary visual cortices and association areas, wherever probes were in the same hemisphere of the same mouse. Since we had spiking data from the thalamus, this was included here, to test for any drivers to or targets from neurons recorded in visual areas. $n_{target\ spikes}$ for probes through these areas outside the primary visual cortex are included in table 3.2.

3.3.2 Effective network inference

In order to test the hypotheses about information processing by columns inside the primary visual cortex under a multivariate approach, we next inferred effective networks of neurons in probes through this area.

The algorithm presented in section 2.5 was used in conjunction with the continuous-time estimator, following the same approach as Shorten et al. (2022b). $N_{surrogates}$ was chosen to be 100 for all source-target pairs. The number of nearest neighbours to use in each conditional TE estimation was 10. The number of nearest neighbours to consider when using the local permutation method to create surrogates was 20 (Shorten et al., 2021).

Due to limited neurons in some probes, we inferred effective networks in each of Waksman's left and right V1, and Krebs' right V1 (via probe 3). These were the only probes through the primary visual cortex with cells in all layers 2/3 - 6.

Since effective network inference only infers significant transfers where statistical power can still be drawn from source-target pairs, any number of spikes for each target can be used during estimation. Where possible we used all target spikes, however to limit computation time a max number of target spikes was set to 3000. That is, all target spikes beyond the first 3000 were excluded. Source spikes

were sampled such that they fell inside the window of target spikes. If any source had less than ten spikes available, they were excluded from inference for that target.

Inference was completed in parallel for each target. Where inference for a target exceeded 60 to 80 hours, computation was sometimes completed by limiting the number of intervals that could be added to the conditioning set before continuing to pruning. Any inference of sources for targets that were limited in this way (n=42) are listed in Table A.1 in the appendix.

In order to test hypotheses about information transfers between columns inside the primary visual cortex, the above methods for effective network inference were repeated only using all cells for probes going through V1 of the same mouse.

3.3.3 Code availability

All information theoretic estimators and inference algorithms used are from the Java Information Dynamics Toolkit (JIDT) (Lizier, 2014). Experimental scripts are available at github.com/preqon/spikes-information-transfer. JIDT is available at github.com/jlizier/jidt.

CHAPTER 4

Results

This chapter presents the results and discusses them in relation to expectations from the literature on visual information processing. The overall information theoretic approach is then evaluated in relation to increasing understanding of neural information processing. Future directions are included at the end of the chapter.

4.1 Intra-column information transfer in V1

4.1.1 Layer 4 is not always the primary source of information to other layers

To test whether or not layer 4 is a primary source of information to other layers (**H1**), we calculated the proportion of significant transfers to total possible transfers from layer 4 to other layers. Proportions for the pairwise analysis are presented in figure 4.1.

Inside Waksman's right primary visual cortex, layer 4 indeed shows higher proportion of significant pairwise transfers to all layers. However this is by no means a consistent story in all recordings of the primary visual cortex in the two mice. In the opposite hemisphere of Waksman itself, none of the 22 sampled cells in layer 4 offered any transfer to cells in other layers. Instead, both layer 5 and 6 were the primary source of information to all other layers, including themselves. The pairwise analysis suggests that layer 4 is not always the primary driver to a column in V1 and that higher order layers can instead become primary sources to lower order layers.

A biologically plausible situation where layer 4 may not be driving other cortical layers would be when layer 4 itself is receiving less thalamic input that maps from one area of the visual field (Lamme and Roelfsema, 2000; Harris and Shepherd, 2015). In other words, where layer 4 was not a primary source, it is possible that no stimulus may have been present in the area of the mouse's visual field corresponding to the recorded column. To add credence to this suggestion without being able to test it directly, we
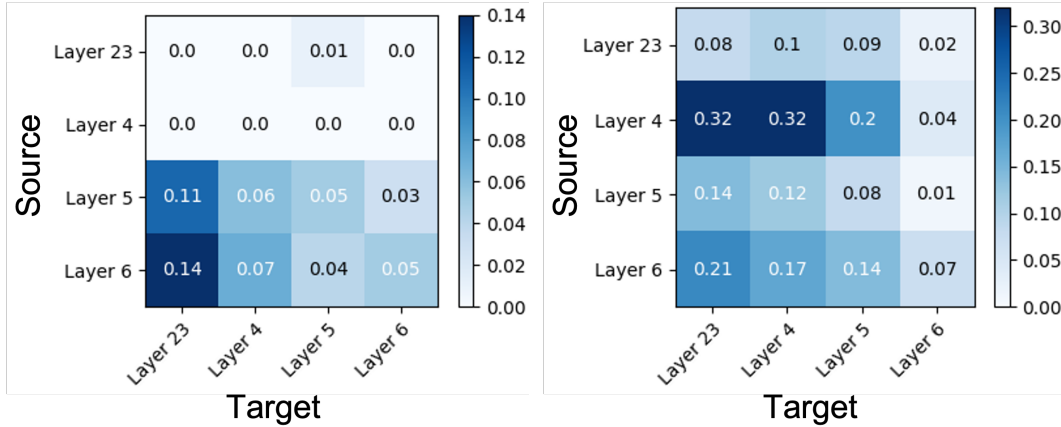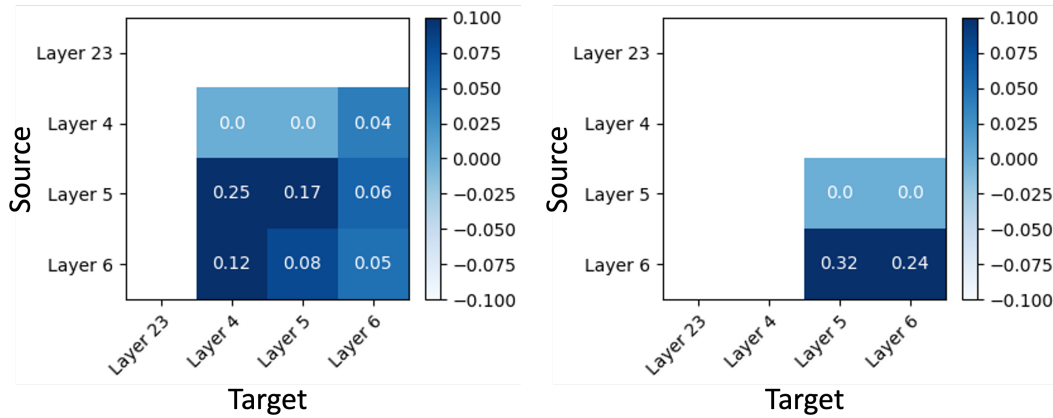
(A) Waksman left primary visual cortex.

(B) Waksman right primary visual cortex.
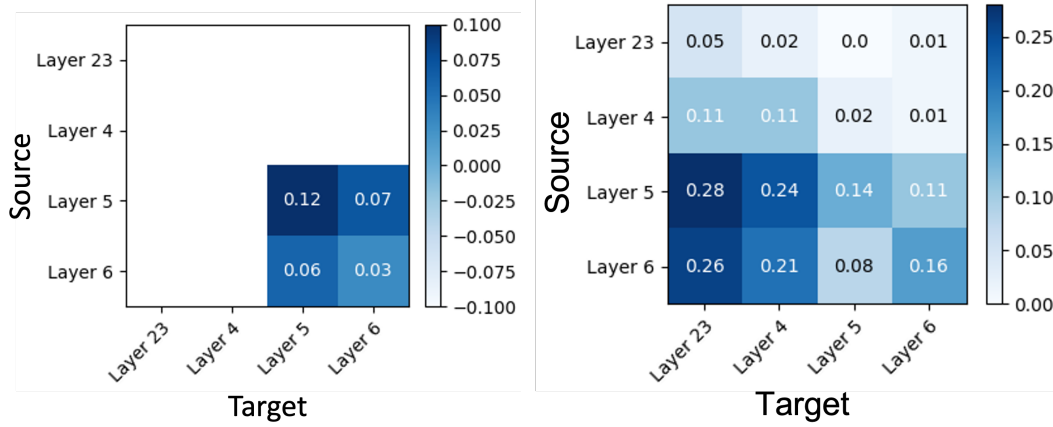
(C) Krebs left primary visual cortex (a)

(D) Krebs right primary visual cortex (a)

(E) Krebs left primary visual cortex (b)

(F) Krebs right primary visual cortex (b)

FIGURE 4.1. Proportion of significant pairwise information transfers to total possible links between layers inside the primary visual cortex.

considered transfers from thalamic cells to the primary visual cortex. The dataset offered one recording of thalamic cells, inside Waksman's left dorsal geniculate nucleus. Indeed, none of the 77 thalamic core and shell cells recorded here offered significant pairwise transfer to layer 4 cells of Waksman's left V1, where in turn layer 4 is not a primary source. The absence of transfer from the thalamus may be because there was no stimulus in the relevant area of the visual field. Although, the thalamic cells from the same hemisphere of the same mouse, were recorded by a different probe to the probe through V1. With this distance, there may be enough hidden processes between the recorded cells to remove possibility of transfer.

Meanwhile in the other mouse Krebs, the proportion of significant pairwise transfers are consistent with the case in Waksman's left V1 where layer 4 is not a primary source. In one probe through Krebs' right V1, layer 4 does offer information transfer to all other layers, however the primary sources of information are layers 5 and 6. The dataset did not offer recordings from thalamic cells inside Krebs to test whether or not a plausible reason for this was lower thalamic drive.

Note that blanks in Figure 4.1 reflect no sampled pairs. While three probes in Krebs do not sample cells from all layers, they still begin to suggest transfers where layers 5 and 6 were dominant sources.

H1 can of course also be evaluated by considering multivariate transfer. A multivariate analysis may be more likely to identify hierarchical processing if this were to exist, since a target of transfer can now be due to the synergy of multiple processes within one layer. Conversely, a source of transfer is less likely to be present if that source is redundant in the context of processes between the source and target in the hierarchy.

Effective network inference was therefore completed inside the three V1 probes that sampled cells from all layers. The extent to which each layer may be considered a dominant source can be first gleaned by calculating the ratio of incoming to outgoing edges in each layer (Figure 4.2). Note a whole effective network is drawn as an example in Appendix A.

Since a ratio less than 1 reflects a layer being more of a source of multivariate transfer than a target, we observe that layer 4 is indeed a source in Waksman's left and right V1s. However, layer 4 is not more dominant as a source compared to other layers in either case. Meanwhile in Krebs' right V1 layer 4 is more of a target than a source. In fact, all layers are heavily targeted with the exception being layer 6. Since layer 6 is a known driver of inhibition within V1 columns, a plausible explanation for this might be suppression for perceptual selection during this particular period of observation (Wang et al., 2020).
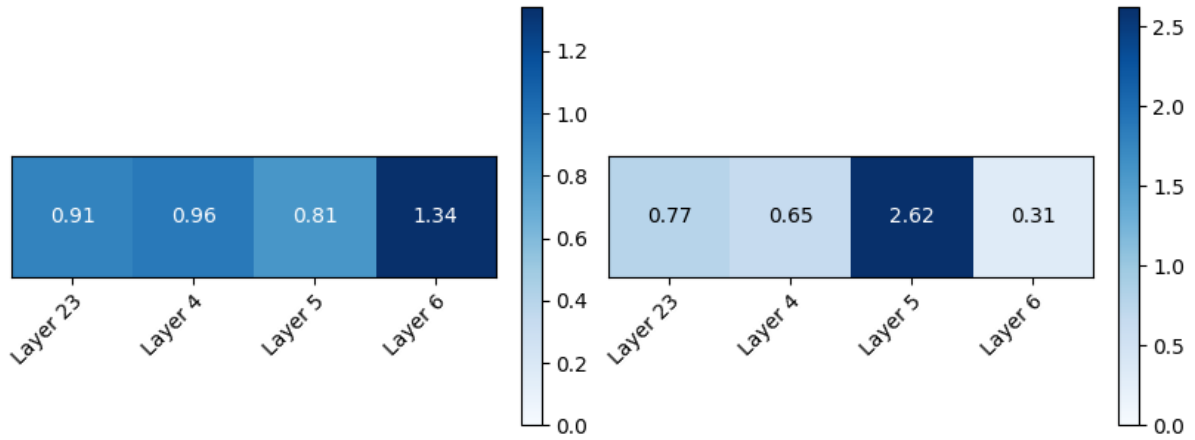
(A) Waksman left primary visual cortex.

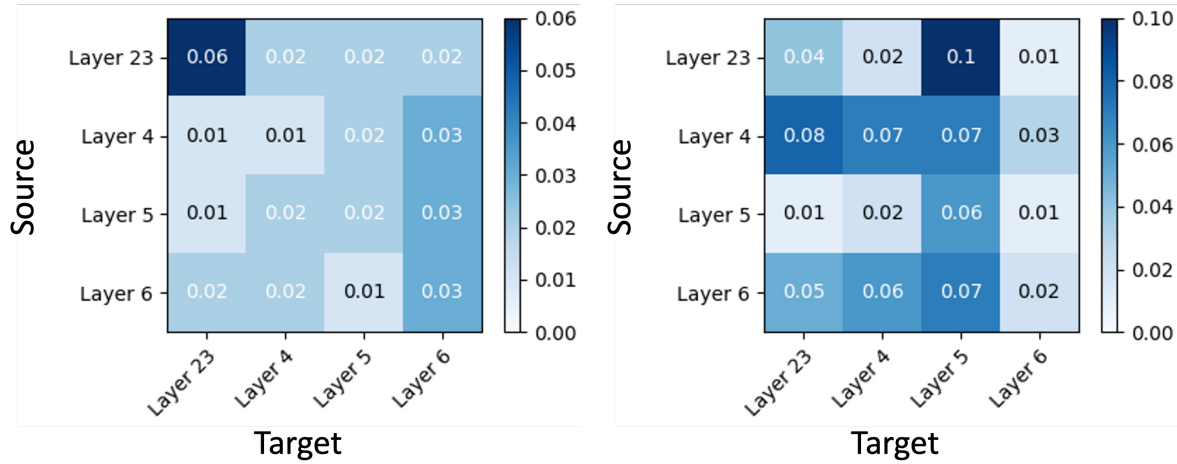(B) Waksman right primary visual cortex.
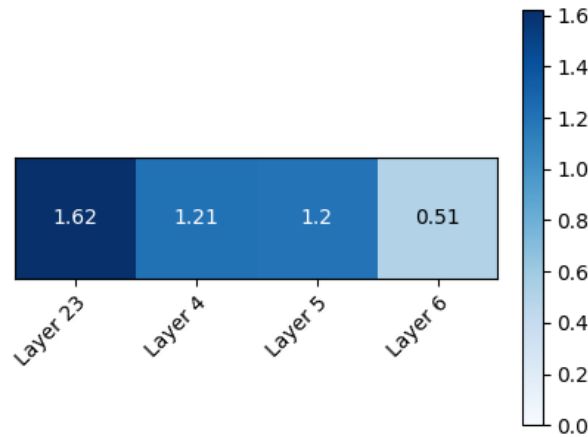
(C) Krebs right primary visual cortex.

FIGURE 4.2. Ratio of incoming to outgoing edges in each layer from effective networks inferred from the primary visual cortex.
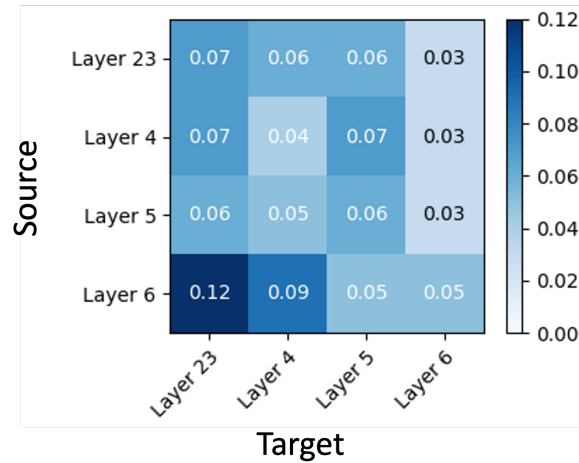
The proportion of significant edges to total possible links between layers in the effective networks are shown in Figure 4.3. Note that even a proportion of 0.01 or higher can be considered a very high proportion of significant transfer. This is because of the maximum statistic test during inference which accounts for multiple comparisons between candidate source intervals (see section 2.5). Cut-off for significant p-values is in effect reduced by a factor of however many source processes are being considered. For these three probes, this leads to the probability of an edge being inferred due to random chance being between 0.0005 and 0.001.

Similar to the pairwise analysis, Waksman's right V1 displayed multivariate information transfer that was most aligned with **H1** that layer 4 would be the primary source of information to other layers. However this was again inconsistent with other probes. The other two probes, from Waksman's left V1

(A) Waksman left primary visual cortex.

(B) Waksman right primary visual cortex.

(C) Krebs right primary visual cortex.

FIGURE 4.3. Proportion of significant edges to total possible links between layers after effective network inference inside the primary visual cortex.

and Krebs right V1 showed a largely homogeneous proportion of multivariate information transfer from each layer. While layer 4 did always show a high proportion of information transfer to other layers in every effective network, it was not always the primary source.

The homogeneous multivariate transfer challenges our expectation of hierarchical information processing, driven primarily by layer 4. Even if lower thalamic-stimulus turns out to be the reason behind layer 4 not emerging as a primary source, these results find quantified evidence of information processing in each column during observation. This means that the same laminar structure of the visual cortex that

has been suggested to best facilitate hierarchical processing (Thomas Yeo et al., 2011) is quite capable of processing information in a non-hierarchical manner, as reflected by the homogeneous transfers.

The mixed results for **H4** highlight how information processing over small windows of observation need not follow expectations from functional connectivity inferred on average. The above results are observed in average windows of length 13 min. In a small time frame like this, it is possible for a subset of drivers in a network to be active. In this case the network would in turn explore only a subset of the state space of spatiotemporal dynamics. Further, hidden variables in the neuronal population, i.e. activity of unsampled neurons, may introduce redundancies and synergies that would lead to a different effective network. However note that, as shown by (Shorten et al., 2022b) the greedy approach is conservative, in the sense that it infers very few false positives. Nevertheless, for the total state space of spatiotemporal dynamics offered by a neuronal population to be observed, larger sample size over longer periods of time are required. Conversely if the goal is to infer information processing on short time scales, observers should not expect information processing to follow functional connectivity inferred over longer time scales.

4.1.2 Layer 2/3 is not always driven by layer 4. Sparse coding was not found by transfer entropy per source spike.

In order to test whether layer 2/3 is primarily driven by layer 4 (**H2**), we can consider the proportion of significant pairwise transfers from layer 4 (Figure 4.6, and also the proportion of significant edges reflecting multivariate transfer in an effective network (Figure 4.3). Inside Waksman's right V1, both the proportions of significant pairwise transfers and multivariate transfers reflect that layer 4 was indeed the primary source to layer 2/3.

However this was again inconsistent with other probes. In the pairwise analysis, Waksman's left V1 showed no transfer whatsoever from layer 4 to layer 2/3. Krebs' right V1 showed some pairwise transfer but layer 4 was not the primary source to layer 2/3; instead layers 5 and 6 were. In the multivariate analysis, some edges do emerge from layer 4 to layer 2/3 in Waksman's left V1, however these transfers homeogeneously emerge across the network. In Krebs' right V1, layer 6 is notably the primary source of multivariate transfer to 2/3, while layer 4 is not more of a source than 2/3 and 5.

In Krebs' right V1, Layer 6 appears as the primary source to 2/3 in the multivariate analysis, while the pairwise analysis includes both layer 5 and 6 as the primary sources. This is aligned with the earlier

suggestion that layer 6 may be controlling informational feedback for perceptual selection, in Krebs' right V1 (Wang et al., 2020). Since feedback to layer 2/3 may be via layer 5 (Fitzpatrick, 1996), this explains the pairwise analysis failing to identify this redundancy.
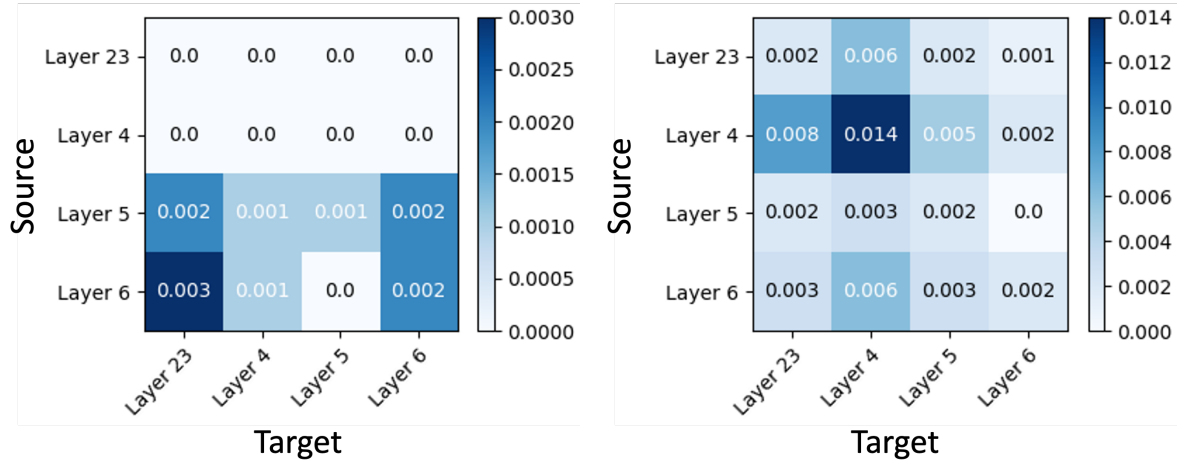
(A) Waksman left primary visual cortex.

(B) Waksman right primary visual cortex.
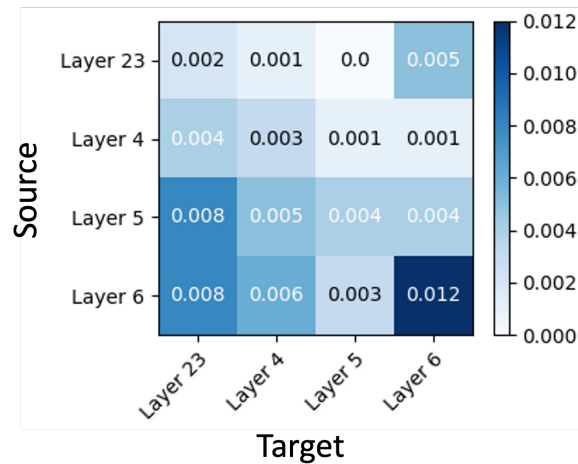
(C) Krebs right primary visual cortex.

FIGURE 4.4. Average pairwise transfer entropy per source spike between layers inside the primary visual cortex.

In order to test whether layer 2/3 offers specialised sparse coding (**H2**), we calculated the average pairwise transfer entropy per source spike between layers. As shown in Figure 4.4, layer 2/3 was not observed to offer a higher transfer entropy per source spike compared to transfers originating in other layers. This is not evidence against sparse coding, nor can the validity of calculating average TE per source spike be tested with this sample size. It is possible however that average TE per source spike may not be the most principled estimation of sparse coding. An example of an alternative approach would

be to estimate the contribution of each burst event to information transfer instead of each spike event (Shorten et al., 2021).

H2 also proposed that layer 2/3 was an intermediary integration station for information and that this might be observed in a higher proportion of significant transfer in layer 2/3 to itself. In the pairwise analysis, layer 2/3 to itself does not stand out compared to other layers to themselves. This is the same in the multivariate analysis, apart from Waksman's left V1. Here, multivariate transfer was homogeneous except for layer 2/3 to itself, which displayed a very high proportion of significant transfer. The ability for layer 2/3 to integrate informational inputs from the local excitatory circuit is observed here, but not as an intermediary stage between layers 4 and 5.

4.1.3 Layer 5 is not always driven by layer 4 and dense-coding is not captured by average rate of transfer entropy

In order to test whether layer 5 is primarily driven by layer 4 and then in part by layer 2/3 (**H3**) we can consider both the proportion of significant pairwise transfers (Figure 4.1 and of multivariate transfers (Figure 4.3 targeting layer 5.

Waksman's right V1 is the most aligned with **H3** in both pairwise and multivariate information transfer. In the pairwise analysis, layer 4 is a clear dominant source, with the highest proportion of significant transfers to layer 5. Layer 2/3 shows some significant transfers, though not more than layer 6 and 5 itself. Interestingly, effective network inference in Waksman's right V1 showed that layer 2/3 was the primary source of multivariate transfer to layer 5. Layer 4 was not more of a source than layer 6 and 5 itself, thus reversing the pairwise inference.

Effective network inference should in principle be better suited to testing a hypothesis about one layer being a primary source of information. This is because the presence of synergy within that layer and the presence of redundancy with other layers is tested. We therefore suggest that layer 2/3 was the most primary source of information to layer 5 in Waksman's right V1 column during observation and not layer 4.

In other probes, the pairwise analysis reflects that layers 5 and 6 were the more primary sources to layer 5. The multivariate analysis in other probes, as outlined previously shows homogeneous from all layers to layer 5. Most recordings in V1 therefore do not reflect **H3** that layer 4 and in part 2/3 would be the primary sources to layer 5.

Waksman's right V1 has so far shown evidence for all major hypotheses (**H1-3**) while few of these hypotheses are supported by other probes. Although substantial claims cannot be made about V1 columns in general with only 8 probes to consider (only three of which sample cells in all layers), this suggests information processing by Waksman's right V1 that was somehow typical to expectations. Without being able to test reasons for this, we suggest one possibility is that a stimulus was present in the area of the visual field relevant to this column, leading to feedforward drivers in the hierarchy becoming active (Lamme and Roelfsema, 2000).

Next, in order to test whether layer 5 codes densely (**H3**), we can consider the average rate of pairwise transfer entropy inside layer 5 compared to inside other layers (Figure 4.5). The average rate of transfer within layer 5 only stands out in Krebs' left primary visual cortex, when compared to other rates of transfer within each layer. All other probes do not show the average rate of transfer within layer 5 to be significantly higher. Similar to **H2**, this is not evidence against dense-coding in layer 5 and it is difficult to conclude whether or not average rate of transfer entropy is well suited to quantify dense coding. If dense coding leads to a macroscopic structure in the spatiotemporal dynamics encoding information, such as changing firing rate (Harris and Mrsic-Flogel, 2013), then a more principled estimation might be found by considering a process to contain population rate rather than spike events.

In order to test whether layer 5 is an information sink (**H3**), we can consider the ratio of incoming to outgoing edges in each layer, following effective network inference (Figure 4.2). Inside Waksman's right V1, layer 5 was indeed an information sink, reflected by a very high ratio of incoming edges compared to outgoing edges.

In Waksman's opposite hemisphere however, layer 5 does not have more incoming than outgoing edges. Here, instead layer 6 is the only layer that is predominantly a sink. In the other mouse Krebs, layer 5 does have more incoming edges than outgoing edges but the difference is not pronounced compared to other layers. Here, layers 2/3 and 4 have a similar ratio of incoming to outgoing edges, with layer 6 the only layer that is predominantly a source.

A consistent story about layer 5 being an information sink was overall not observed. Again, Waksman's right V1 was the most aligned to this hypothesis, suggesting information processing by this column more typical to hierarchical processing, perhaps due to visual stimulus.
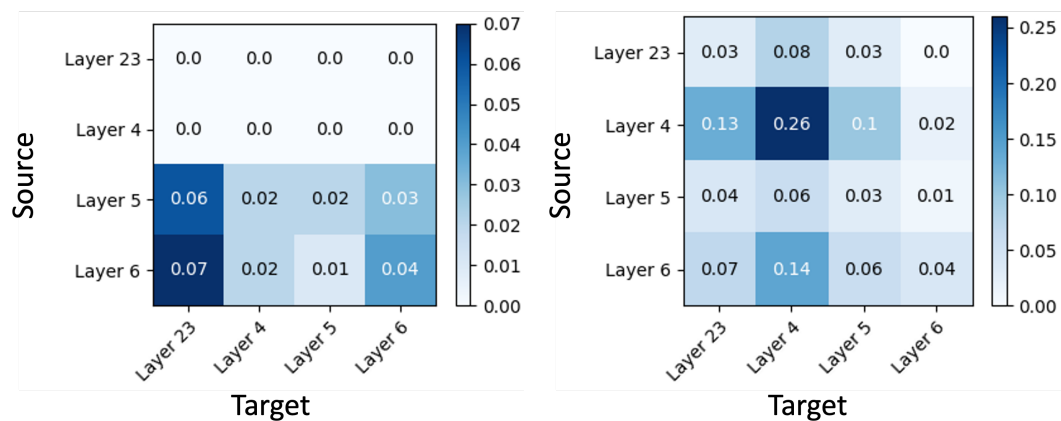
(A) Waksman's left V1.

(B) Waksman's right V1.
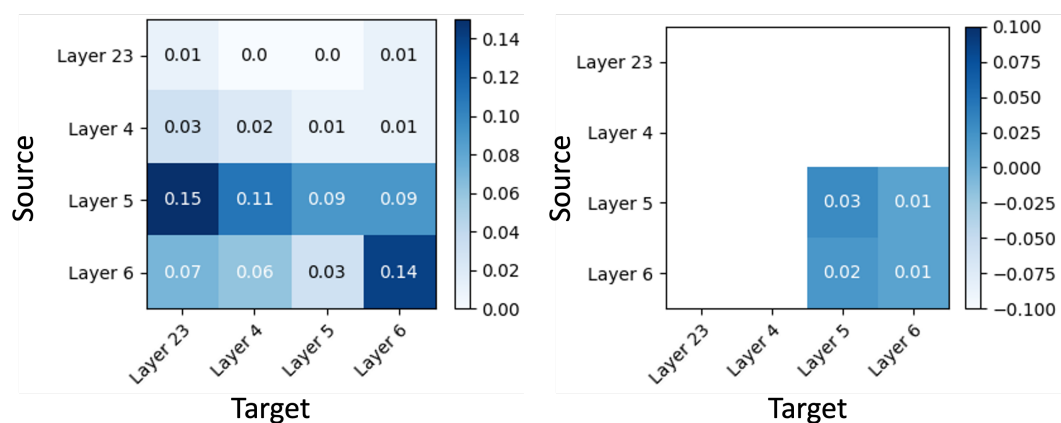
(C) Krebs' left V1 (a).

(D) Krebs' right V1 (a).

(E) Krebs' left V1 (a).

(F) Krebs' right V1 (b).

FIGURE 4.5. Average rate of pairwise information transfer between layers inside the primary visual cortex.

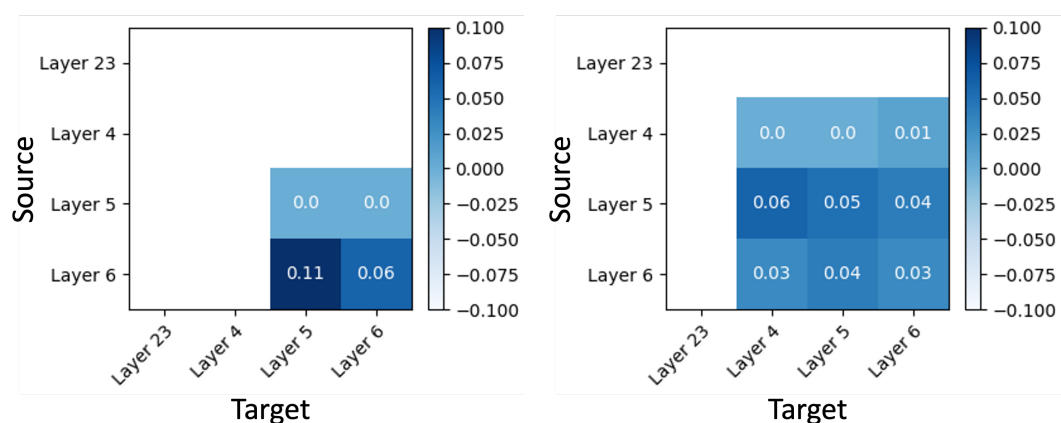Finally, evidence for **H4** that each layer forms a computationally specific module was not found which was surprising even in the short time scale and small number of probes analysed. The analysis of average pairwise transfer entropy rate in Figure 4.5 almost never reflects that each layer has a high rate to itself in comparison to targeting other layers. This hypothesis was expected to be supported given the physiological and behavioural specification of each layer (Harris and Shepherd, 2015). However, computational specificity can be implied by things other than unique amounts of information transfer, such as task-relevance of the information. The quantity of information is not associated with its computational meaning.

4.2 Extra-column information transfer in V1 and association areas

In order to test which layers offer information transfer between columns, we considered the proportion of significant pairwise transfers between probes in the same hemisphere of Krebs' primary visual cortex (Figure 4.6). Note that in this figure source layers and target layers are from different columns.

H6 considered that layer 6 would be a source of information transfer to layers in other columns. This was supported, both in Krebs' left and right primary visual cortices. Although not hypothesised, layer 5 showed a comparable proportion of transfers to layers in disparate columns.

Further, probe (a) in Krebs' right V1 sampled layers 2/3 and 4. These show less broadcasting to disparate columns than layers 5 and 6, which aligns with **H6**. Unexpectedly, layer 6 shows a high proportion of transfer to layer 2/3 , indicating transfers are possible to non-horizontal targets in the disparate column.

Interestingly, there is asymmetry in both hemispheres. In Krebs' left V1, probe (a)'s layer 5 shows high transfer to probe (b)'s layer 4, though none is returned in the other direction. Conversely probe (a)'s layer 5 to probe (b)'s layer 5 is much lower than the other direction. Meanwhile, in Krebs' right V1, there is even stronger asymmetry. Almost all layers in probe (a) show transfer to layer 5 in probe (b), but none is returned. Further, almost all layers in probe (a) show little to no transfer to layer 6 in probe (b) but a large proportion of transfer is found in the other direction. Overall, the asymmetry here suggests that columns as a whole offer some computational specification. This supports the view from Harris and Shepherd (2015) that even though local excitatory circuits are homologous, their retinotopy leads to varied function. That is, columns process different areas of the visual field.

(A) Probe (a) to (b) inside Krebs' left V1

(B) Probe (b) to (a) inside Krebs' left V1.

(C) Probe (a) to (b) inside Krebs' right V1

(D) Probe (b) to (a) inside Krebs' right V1
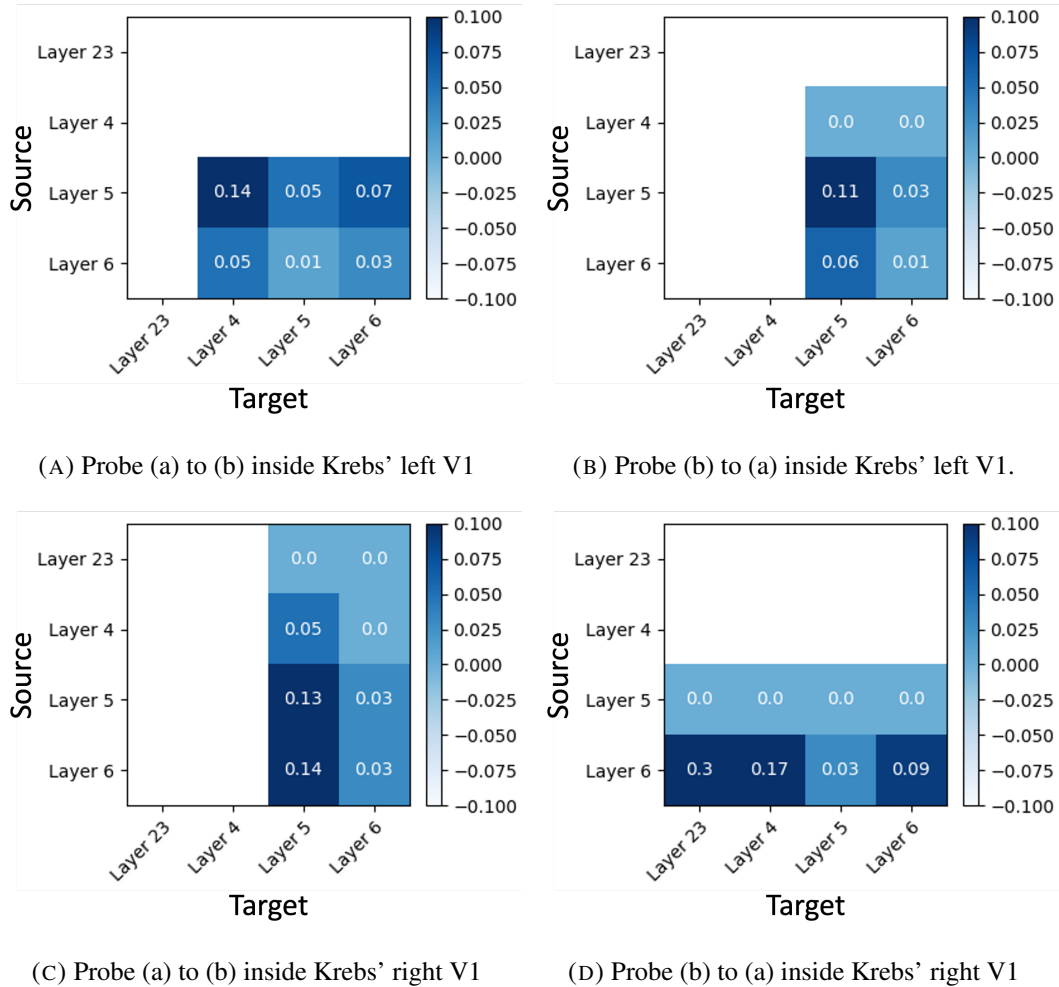
FIGURE 4.6. Proportion of significant pairwise information transfers to total possible links between layers in different columns inside the primary visual cortex.

Finally, results for transfers from V1 columns to association areas did not support either of **H5** or **H7**. These were that layer 5 and layer 6 would respectively broadcast and integrate long distance information from association areas. In our pairwise analysis, no transfers were found whatsoever between Waksman's right V1 and anterior visual area. Similarly, virtually no transfer was found between Waksman's left V1 and rostrolateral visual area (VISrl). Of the transfers that were found, layer 5 did not broadcast to cells in VISrl (**H5**). Instead, a small proportion of significant transfer was found from layer 6 to cells in VISrl. While layer 6 did receive a small proportion of significant transfer from cells in VISrl (**H7**), this was comparable to other layers. These were all sampled association areas and so association was not investigated further.

4.3 Evaluation of approach and future directions

This project adopted an information theoretic approach to quantify neural information processing. The continuous-time estimator for transfer entropy and algorithm for effective network inference were leveraged for their ability to quantify information transfers even in the high-dimensional, distributed setting offered by neuronal populations. We first emphasise how this approach in recordings of visual areas of mouse brains led to findings that exemplify how information processing need not follow structural organisation. The primary visual cortex (V1) is structurally organised such that each local excitatory circuit forms a hierarchy of layers within it. Information transfers in V1 recorded in our dataset did not adhere to this hierarchical organisation, instead often being homogeneous between layers. Network structure does not offer insight into its dynamical state space and therefore the information processing allowed by its activity.

However, our survey of the literature showed that the hierarchy in V1 emerges not only in structure, but also from the inference of functional connectivity and even the multivariate form of coupling of activity (Guillery and Sherman, 2002; Wang et al., 2020). This leads us to a simple yet profound nuance about distributed information processing. It is true that functional connectivity especially in a multivariate form, should correlate to information processing, despite not directly quantifying it (Friston, 2011). However when statistically inferred over long time windows, functional connectivity does not necessarily imply the information processing that is still present during short time windows of observation. Both our pairwise and multivariate quantification of information transfers showed cases where functional connectivity was not adhered to, over short observations of 10 - 15 min.

It is possible that given longer time periods, the information theoretic approach would yield quantified transfers aligned to expectations from functional connectivity. An alternative improvement to our approach is to test task or condition-related information processing, which could still be appreciably quantified in short time windows. The dataset used by this project offered spontaneous recordings but the same estimator and inference algorithm can be applied in recordings during task. One limitation the information theoretic tools may encounter in this direction is that it may not be obvious how to estimate the amount of information transferred to the task output/behaviour itself or the amount of information stored related to stimulus. This would require tailoring the tools to the test condition after considering the relevant state spaces of behaviour or stimulus.

A limitation of the approach in our dataset was that macroscopic structures in the spatiotemporal dynamics offered by some layers were unable to be tested for their transfer capacity. Layer 2/3 was suggested to code sparsely in temporally and spatially specific bursts but this phenomenon was not quantified. A future direction would be in the adaptation of the continuous-time transfer entropy estimator to be able to first identify burst events before quantifying related transfers. This might be to other burst events or to spike events. The latter would be more principled by the expected neural coding between layer 2/3 and other layers (Petersen and Crochet, 2013). Similarly, layer 5 was suggested to code densely, allowing varying firing rate to encode information. Quantifying information transfer between a change in firing rate and other signalling events would require another adjustment to the estimator. Once these notions of transfers from macroscopic structures can be estimated, their contributions to significant multivariate transfer can then be tested using effective network inference. This would yield a model more principled by the complete distributed information processing allowed by a neuronal population.

Overall, this evaluation in visual areas finds the information theoretic approach to still be necessary in quantifying information. Statistical inferences of brain function offer a distinct framework. The continuous-time estimator for transfer entropy can be tailored to new contexts, where it can still be applied to effective network inference allowing further insight into neural information processing.

CHAPTER 5

Conclusion

This project leveraged recent advancements in the estimation of transfer entropy from spike trains to take advantage of high dimensional recordings in the visual areas of mouse brains. Quantified information transfers were first under a pairwise analysis between source and target cells. A multivariate analysis where significance reflected maximal explanation of target state by the minimal set of sources was also applied. Our evaluation suggests that the latter under effective network inference is better suited to quantifying information transfers in a complex system containing highly distributed and parallelised processes.

By quantifying information transfer in short time windows, the findings added nuance to expectations from functional connectivity inside visual areas. In particular, local excitatory circuits in the primary visual cortex show a range of information processing not limited to hierarchical organisation. Transfers between layers can be homogeneous instead of hierarchical and layers other than layer 4 can be the primary source inside a column. This demonstrates how the information theoretic approach, which is the only framework that quantifies information, appreciably adds to models of neural information processing.

Future work should be towards tailoring the continuous-time estimator to task or stimulus driven neural information processing, as well as towards being able to identify transfers between macroscopic structures in spatiotemporal dynamics. Quantitative models of neural information processing, principled by the total state space offered by dynamic complex systems, can then increase understanding of brain function and natural computation as a whole.

Bibliography

Edward H. Adelson and James R. Bergen. 1985. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 2(2):284 – 299. Cited by: 2613; All Open Access, Green Open Access.

Shun-ichi Amari and Kazuo Kishimoto. 1978. Dynamics of excitation patterns in lateral inhibitory neural fields. *Electronics & communications in Japan*, 61(7):1 – 8. Cited by: 1409.

M. F. Bear, B. W. Conners, and M. A. Paradiso. 2015. *Neuroscience: Exploring the Brain.* Williams & Wilkins, Baltimore.

John M. Beggs and Dietmar Plenz. 2003. Neuronal avalanches in neocortical circuits. *Journal of Neuroscience*, 23(35):11167 – 11177. Cited by: 1407; All Open Access, Green Open Access, Hybrid Gold Open Access.

Eshel Ben-Jacob. 2009. Learning from bacteria about natural information processing. *Annals of the New York Academy of Sciences*, 1178(1):78–90.

Terry Bossomaier, Lionel Barnett, Michael Harré, and Joseph T. Lizier. 2016. *An introduction to transfer entropy: Information flow in complex systems*. Springer Cham. Cited by: 166; All Open Access, Green Open Access.

Steven L. Bressler and Anil K. Seth. 2011. Wiener-granger causality: A well established methodology. *NeuroImage*, 58(2):323 – 329. Cited by: 537.

Nicolas Brunel. 2000. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 8(3):183 – 208. Cited by: 1081.

György Buzsáki and Andreas Draguhn. 2004. Neuronal oscillations in cortical networks. *Science*, 304(5679):1926 – 1929. Cited by: 4166; All Open Access, Green Open Access.

Danilo Bzdok and B.T. Thomas Yeo. 2017. Inference in the age of big data: Future perspectives on neuroscience. *NeuroImage*, 155:549 – 564. Cited by: 115; All Open Access, Green Open Access, Hybrid Gold Open Access.

N Caporale and Dan Y. 2008. Spike timing–dependent plasticity: a hebbian learning rule. *Annual Review of Neuroscience*, 31:25–46.

Gang Chen, Paul A. Taylor, Joel Stoddard, Robert W. Cox, Peter A. Bandettini, and Luiz Pessoa. 2021. Sources of information waste in neuroimaging: mishandling structures, thinking dichotomously, and over-reducing data. *bioRxiv*.

Zhe Chen. 2013. An overview of bayesian methods for neural spike train analysis. *Computational Intelligence and Neuroscience*, 2013. Cited by: 39; All Open Access, Gold Open Access, Green Open Access.

Francisco Clascá, Pablo Rubio-Garrido, and Denis Jabaudon. 2012. Unveiling the diversity of thalamo-cortical neuron subtypes. *European Journal of Neuroscience*, 35(10):1524 – 1532. Cited by: 105.

R. Clay Reid and Jose-Manuel Alonso. 1995. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378(6554):281 – 284. Cited by: 469.

Albert Compte, Nicolas Brunel, Patricia S. Goldman-Rakic, and Xiao-Jing Wang. 2000. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10(9):910 – 923. Cited by: 761; All Open Access, Bronze Open Access, Green Open Access.

T. M. Cover and Joy A. Thomas. 2010. *Elements of information theory*. Wiley-India.

Gordana Dodig Crnkovic. 2011. Dynamics of information as natural computation. *Information (Switzerland)*, 2(3):460 – 477. Cited by: 26; All Open Access, Gold Open Access, Green Open Access.

Gordana Dodig Crnkovic. 2012. Physical computation as dynamics of form that glues everything together. *Information (Switzerland)*, 3(2):204 – 218. Cited by: 22; All Open Access, Gold Open Access, Green Open Access.

Scott J. Cruikshank, Omar J. Ahmed, Tanya R. Stevens, Saundra L. Patrick, Amalia N. Gonzalez, Margot Elmaleh, and Barry W. Connors. 2012. Thalamic control of layer 1 circuits in prefrontal cortex. *Journal of Neuroscience*, 32(49):17813 – 17823. Cited by: 150; All Open Access, Green Open Access, Hybrid Gold Open Access.

Sophie Denève and Christian K. Machens. 2016. Efficient codes and balanced networks. *Nature Neuroscience*, 19(3):375 – 382. Cited by: 237.

Markus Dlesmann, Marc-Oliver Gewaltig, and Ad Aertsen. 1999. Stable propagation of synchronous spiking in cortical neural networks. *Nature*, 402(6761):529 – 533. Cited by: 702.

Marco Dorigo, Gianni Di Caro, and Luca M. Gambardella. 1999. Ant algorithms for discrete optimization. *Artificial Life*, 5(2):137 – 172. Cited by: 2165.

Dirk Feldmeyer. 2012. Excitatory neuronal connectivity in the barrel cortex. *Frontiers in Neuroanatomy*. Cited by: 166; All Open Access, Gold Open Access, Green Open Access.

Charles Findling and Valentin Wyart. 2021. Computation noise in human learning and decision-making: origin, impact, function. *Current Opinion in Behavioral Sciences*, 38:124–132. Computational cognitive neuroscience.

David Fitzpatrick. 1996. The functional organization of local circuits in visual cortex: Insights from the study of tree shrew striate cortex. *Cerebral Cortex*, 6(3):329 – 341. Cited by: 166.

Benjamin Flecker, Wesley Alford, John M. Beggs, Paul L. Williams, and Randall D. Beer. 2011. Partial information decomposition as a spatiotemporal filter. *Chaos*, 21(3). Cited by: 20.

Antonio J. Fontenele, Nivaldo A. P. De Vasconcelos, Thaís Feliciano, Leandro A. A. Aguiar, Carina Soares-Cunha, Bárbara Coimbra, Leonardo Dalla Porta, Sidarta Ribeiro, Ana João Rodrigues, Nuno Sousa, Pedro V. Carelli, and Mauro Copelli. 2019. Criticality between cortical states. *Physical Review Letters*, 122(20). Cited by: 104; All Open Access, Green Open Access.

Karl J. Friston. 2011. Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13 – 36. Cited by: 2089; All Open Access, Green Open Access.

S. Funahashi, C.J. Bruce, and P.S. Goldman-Rakic. 1989. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331 – 349. Cited by: 1974.

Linda Geerligs, Richard N. Henson, and Cam-CAN. 2016. Functional connectivity and structural covariance between regions of interest can be measured more accurately using multivariate distance correlation. *NeuroImage*, 135:16 – 31. Cited by: 72; All Open Access, Green Open Access, Hybrid Gold Open Access.

I.M. Gelfand and A.M. I'Aglom. 1959. *Calculation of the amount of information about a random function contained in another such function*. American Mathematical Society.

Wulfram Gerstner and Werner M. Kistler. 2008a. *Elements of Neuronal Systems*, chapter 1.1. Cambridge University Press.

Wulfram Gerstner and Werner M. Kistler. 2008b. *Formal Spiking Neuron Models*, chapter 4. Cambridge University Press.

Wulfram Gerstner and Werner M. Kistler. 2008c. *Noise in Spiking Neuron Models*, chapter 5. Cambridge University Press.

Jeffrey Goldstein. 2011. Emergence in complex systems. *The Sage Handbook of Complexity and Management*, pages 65–78.

R.W. Guillery and S.Murray Sherman. 2002. Thalamic relay functions and their role in corticocortical communication: Generalizations from the visual system. *Neuron*, 33(2):163 – 175. Cited by: 467; All Open Access, Bronze Open Access.

Bryan J. Hansen, Mircea I. Chelaru, and Valentin Dragoi. 2012. Correlated variability in laminar cortical circuits. *Neuron*, 76(3):590 – 602. Cited by: 99; All Open Access, Bronze Open Access, Green Open Access.

Kenneth D. Harris and Thomas D. Mrsic-Flogel. 2013. Cortical connectivity and sensory coding. *Nature*, 503(7474):51 – 58. Cited by: 353.

Kenneth D Harris and Gordon M G Shepherd. 2015. The neocortical circuit: themes and variations. *Nature Neuroscience*, 18(2):170–181.

J.J. Hopfield. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8):2554 – 2558. Cited by: 12185; All Open Access, Green Open Access.

E.M. Izhikevich. 2003. Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6):1569–1572.

Edward G Jones. 2001. The thalamic matrix and thalamocortical synchrony. *Trends in Neurosciences*, 24(10):595 – 601. Cited by: 451.

James J. Jun, Nicholas A. Steinmetz, Joshua H. Siegle, Daniel J. Denman, Marius Bauza, Brian Barbarits, Albert K. Lee, Costas A. Anastassiou, Alexandru Andrei, Çağatay Aydin, Mladen Barbic, Timothy J. Blanche, Vincent Bonin, João Couto, Barundeb Dutta, Sergey L. Gratiy, Diego A. Gutnisky, Michael Häusser, Bill Karsh, Peter Ledochowitsch, Carolina Mora Lopez, Catalin Mitelut,

Silke Musa, Michael Okun, Marius Pachitariu, Jan Putzeys, P. Dylan Rich, Cyrille Rossant, Wei-Lung Sun, Karel Svoboda, Matteo Carandini, Kenneth D. Harris, Christof Koch, John O'Keefe, and Timothy D. Harris. 2017. Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232 – 236. Cited by: 895; All Open Access, Green Open Access.

Marcus Kaiser. 2007. Brain architecture: A design for natural computation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1861):3033 – 3045. Cited by: 56; All Open Access, Green Open Access.

Adam Keane and Pulin Gong. 2015. Propagating waves can explain irregular neural dynamics. *Journal of Neuroscience*, 35(4):1591 – 1605. Cited by: 42; All Open Access, Green Open Access, Hybrid Gold Open Access.

Victor A.F. Lamme and Pieter R. Roelfsema. 2000. The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11):571 – 579. Cited by: 1599.

Chris G. Langton. 1990. Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D: Nonlinear Phenomena*, 42(1-3):12 – 37. Cited by: 1004; All Open Access, Green Open Access.

Joseph T. Lizier. 2013. *Computation in Complex Systems*, pages 13–52. Springer Berlin Heidelberg, Berlin, Heidelberg.

Joseph T. Lizier. 2014. Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI*, 1.

Joseph T. Lizier, Benjamin Flecker, and Paul L. Williams. 2013. Towards a synergy-based approach to measuring information modification. In *2013 IEEE Symposium on Artificial Life (ALife)*, January, page 43 – 51. Cited by: 38; All Open Access, Green Open Access.

Joseph T. Lizier, Mikhail Prokopenko, and Albert Y. Zomaya. 2010. Information modification and particle collisions in distributed computation. *Chaos*, 20(3). Cited by: 89.

Joseph T. Lizier, Mikhail Prokopenko, and Albert Y. Zomaya. 2012. Local measures of information storage in complex distributed computation. *Information Sciences*, 208:39 – 54. Cited by: 107.

Joseph T. Lizier and Mikail Rubinov. 2012. Multivariate construction of effective computational networks from observational data. In *Max Planck Institute for Mathematics in the Sciences Preprint*.

Nikos K. Logothetis and Brian A. Wandell. 2004. Interpreting the bold signal. *Annual Review of Physiology*, 66:735 – 769. Cited by: 1122.

David J.C. MacKay. 2022. *Information theory, Inference, and learning algorithms*. Cambridge University Press.

Bruce J. MacLennan. 2004. Natural computation and non-turing models of computation. *Theoretical Computer Science*, 317(1-3):115 – 145. Cited by: 55; All Open Access, Bronze Open Access.

David Marr. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc., New York, NY, USA.

Robert J. Mceliece, Edward C. Posner, Eugene R. Rodemich, and Santosh S. Venkatesh. 1987. The capacity of the hopfield associative memory. *IEEE Transactions on Information Theory*, 33(4):461 – 482. Cited by: 616; All Open Access, Green Open Access.

Paul Miller. 2016. Dynamical systems, attractors, and neural circuits [version 1; referees: 3 approved]. *F1000Research*, 5. Cited by: 24; All Open Access, Gold Open Access, Green Open Access.

John Milnor. 1985. On the concept of attractor. *Communications in Mathematical Physics*, 99(2):177 – 195. Cited by: 685.

Melanie Mitchell. 1998. *Computation in Cellular Automata: A Selected Review*, chapter 4, pages 95–140. John Wiley & Sons, Ltd.

C. Neuper and G. Pfurtscheller. 2001. Event-related dynamics of cortical rhythms: Frequency-specific features and functional correlates. *International Journal of Psychophysiology*, 43(1):41 – 58. Cited by: 581.

M.E.J. Newman. 2003. The structure and function of complex networks. *SIAM Review*, 45(2):167 – 256. Cited by: 13361; All Open Access, Green Open Access.

Leonardo Novelli and Joseph T. Lizier. 2021. Inferring network properties from time series using transfer entropy and mutual information: Validation of multivariate versus bivariate approaches. *Network Neuroscience*, 5(2):373 – 404. Cited by: 8; All Open Access, Gold Open Access, Green Open Access.

Leonardo Novelli, Patricia Wollstadt, Pedro Mediano, Michael Wibral, and Joseph T. Lizier. 2019. Large-scale directed network inference with multivariate transfer entropy and hierarchical statistical testing. *Network Neuroscience*, 3(3):827 – 847. Cited by: 45; All Open Access, Gold Open Access, Green Open Access.

Jan Oettler, Volker S. Schmid, Niko Zankl, Olivier Rey, Andreas Dress, and Jürgen Heinze. 2013. Fermat's principle of least time predicts refraction of ant trails at substrate borders. *PLOS ONE*, 8(3):1–7.

John O'Keefe and Michael L. Recce. 1993. Phase relationship between hippocampal place units and the eeg theta rhythm. *Hippocampus*, 3(3):317–330.

Daniel M. Oppenheimer and Evan Kelso. 2015. Information processing as a paradigm for decision making. *Annual Review of Psychology*, 66(1):277–294. PMID: 25559114.

Alan Peters and Daniel A. Kara. 1985. The neuronal composition of area 17 of rat visual cortex. i. the pyramidal cells. *Journal of Comparative Neurology*, 234(2):218 – 241. Cited by: 137.

Carl C.H. Petersen and Sylvain Crochet. 2013. Synaptic computation and sensory processing in neocortical layer 2/3. *Neuron*, 78(1):28 – 48. Cited by: 164; All Open Access, Bronze Open Access.

Mikhail Prokopenko and Joseph T. Lizier. 2014. Transfer entropy and transient limits of computation. *Scientific Reports*, 4. Cited by: 46; All Open Access, Gold Open Access, Green Open Access.

M. Rabinovich, A. Volkovskii, P. Lecanda, R. Huerta, H.D.I. Abarbanel, and G. Laurent. 2001. Dynamical encoding by networks of competing neuron groups: Winnerless competition. *Physical Review Letters*, 87(6):681021 – 681024. Cited by: 314; All Open Access, Green Open Access.

Lucas Rudelt, Daniel González Marx, Michael Wibral, and Viola Priesemann. 2021. Embedding optimization reveals long-lasting history dependence in neural spiking activity. *PLoS Computational Biology*, 17(6). Cited by: 2; All Open Access, Gold Open Access, Green Open Access.

Hiroki Sayama. 2015. *Introduction to the modeling and analysis of Complex Systems*. Open SUNY Textbooks, Milne Library.

C.E. Shannon. 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379 – 423. Cited by: 27557.

Tatyana O. Sharpee, Hiroki Sugihara, Andrei V. Kurgansky, Sergei P. Rebrik, Michael P. Stryker, and Kenneth D. Miller. 2006. Adaptive filtering enhances information transmission in visual cortex. *Nature*, 439(7079):936 – 942. Cited by: 239; All Open Access, Green Open Access.

David Shorten, Viola Priesseman, Michael Wibral, and Joseph Lizier. 2022a. Early lock- in of structured and specialised information flows during neural development. *eLife*, 11.

David Shorten, Richard Spinney, and Joseph Lizier. 2021. Estimating Transfer Entropy in Continuous Time Between Neural Spike Trains or Other Event-Based Data. *PLoS Computational Biology*, 17(4).

David P. Shorten, Viola Priesemann, Michael Wibral, and Joseph T. Lizier. 2022b. Inferring effective networks of spiking neurons using a continuous-time estimator of1 transfer entropy. In *Centre for Complex Systems, Faculty of Engineering, The University of Sydney Preprint.*

Alexander F. Siegenfeld and Yaneer Bar-Yam. 2020. An introduction to complex systems science and its applications. *Complexity*, 2020. Cited by: 64; All Open Access, Gold Open Access, Green Open Access.

Plamen L. Simeonov. 2010. Integral biomathics: A post-newtonian view into the logos of bios. *Progress in Biophysics and Molecular Biology*, 102(2-3):85 – 121. Cited by: 25; All Open Access, Green Open Access.

Richard E. Spinney, Mikhail Prokopenko, and Joseph T. Lizier. 2017. Transfer entropy in continuous time, with applications to jump and neural spiking processes. *Physical Review E*, 95(3). Cited by: 38; All Open Access, Green Open Access.

Jennifer R. Stapleton, Michael L. Lavine, Robert L. Wolpert, Miguel A. L. Nicolelis, and Sidney A. Simon. 2006. Rapid taste responses in the gustatory cortex during licking. *Journal of Neuroscience*, 26(15):4126 – 4138. Cited by: 109; All Open Access, Green Open Access, Hybrid Gold Open Access.

Armen Stepanyants, Luis M. Martinez, Alex S. Ferecskó, and Zoltán F. Kisvárday. 2009. The fractions of short- and long-range connections in the visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3555 – 3560. Cited by: 139; All Open Access, Bronze Open Access, Green Open Access.

Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D. Harris. 2019. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437). Cited by: 401; All Open Access, Green Open Access.

Steven H. Strogatz. 2000. From kuramoto to crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1-4):1 – 20. Cited by: 2106; All Open Access, Green Open Access.

B.T. Thomas Yeo, Fenna M. Krienen, Jorge Sepulcre, Mert R. Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L. Roffman, Jordan W. Smoller, Lilla Zöllei, Jonathan R. Polimeni, Bruce Fisch, Hesheng Liu, and Randy L. Buckner. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125 – 1165. Cited by: 4338; All Open Access, Green Open Access.

Alex M. Thomson and Christophe Lamy. 2007. Functional maps of neocortical local circuitry. *Frontiers in Neuroscience*, 1(1):19 – 42. Cited by: 303; All Open Access, Gold Open Access, Green Open Access.

M. Tsodyks, A. Uziel, and H. Markram. 2000. Synchrony generation in recurrent networks with frequency-dependent synapses. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20(1):RC50. Cited by: 225; All Open Access, Bronze Open Access, Green Open Access.

Satoshi Tsujimoto and Bradley R. Postle. 2012. The prefrontal cortex and oculomotor delayed response: A reconsideration of the mnemonic scotoma. *Journal of Cognitive Neuroscience*, 24(3):627 – 635. Cited by: 30; All Open Access, Green Open Access.

A. M. Turing. 1937. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1):230–265.

Mateo Vélez-Fort, Charly V. Rousseau, Christian J. Niedworok, Ian R. Wickersham, Ede A. Rancz, Alexander P.Y. Brown, Molly Strom, and Troy W. Margrie. 2014. The stimulus selectivity and connectivity of layer six principal cells reveals cortical microcircuits underlying visual processing. *Neuron*, 83(6):1431 – 1443. Cited by: 107; All Open Access, Green Open Access, Hybrid Gold Open Access.

Lei Wang, Qi Kang, and Qi di Wu. 2007. Nature-inspired computation — effective realization of artificial intelligence. *Systems Engineering - Theory & Practice*, 27(5):126–134.

Tian Wang, Yang Li, Guanzhong Yang, Weifeng Dai, Yi Yang, Chuanliang Han, Xingyun Wang, Yange Zhang, and Dajun Xing. 2020. Laminar subnetworks of response suppression in macaque primary visual cortex. *Journal of Neuroscience*, 40(39):7436 – 7450. Cited by: 15; All Open Access, Bronze Open Access, Green Open Access.

Xiao-Jing Wang. 2002. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36(5):955 – 968. Cited by: 795; All Open Access, Bronze Open Access.

Michael Wibral, Joseph T. Lizier, Sebastian Vögler, Viola Priesemann, and Ralf Galuske. 2014. Local active information storage as a tool to understand distributed neural information processing. *Frontiers in Neuroinformatics*, 8(JAN). Cited by: 66; All Open Access, Gold Open Access, Green Open Access.

Klaus Wimmer, Duane Q Nykamp, Christos Constantinidis, and Albert Compte. 2014. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, 17(3):431 – 439. Cited by: 217.

A.D. Wyner. 1978. A definition of conditional mutual information for arbitrary ensembles. *Information and Control*, 38(1):51 – 59. Cited by: 86; All Open Access, Bronze Open Access.

Drawn effective network and limited computations during inference

For ease of visualisation, an example effective network from our results is drawn in Figure A.1.

As explained in Chapter 3 Experiments, effective network inference was sometimes manually limited to speed up computation. Targets for which inference was limited are listed in Table A.1.

Mouse	Probe	Target	Added intervals
Waksman	Probe 8	2	12
		8	13
		13	13
		17	10
		19	13
		22	22
		43	21
		44	21
		49	11
		51	11
		52	14
		54	15
		55	12
Waksman	Probe 3	2	15
		3	19
		8	14
		11	13
		27	13
		34	15
		37	15
		39	15
		41	15
		49	13
		51	13
		54	11
		69	14
Krebs	Probe 3	1	21
		13	23
		14	18
		27	19
		28	20
		31	22
		38	22
		47	19
Krebs	Probes 3, 7	3	20
		34	20
		38	20
		48	20
Krebs	Probes 4, 8	5	20
		14	20
		31	20
		64	20

TABLE A.1. Targets where inference was manually limited, showing the number of intervals added to the conditioning set before continuing to pruning.
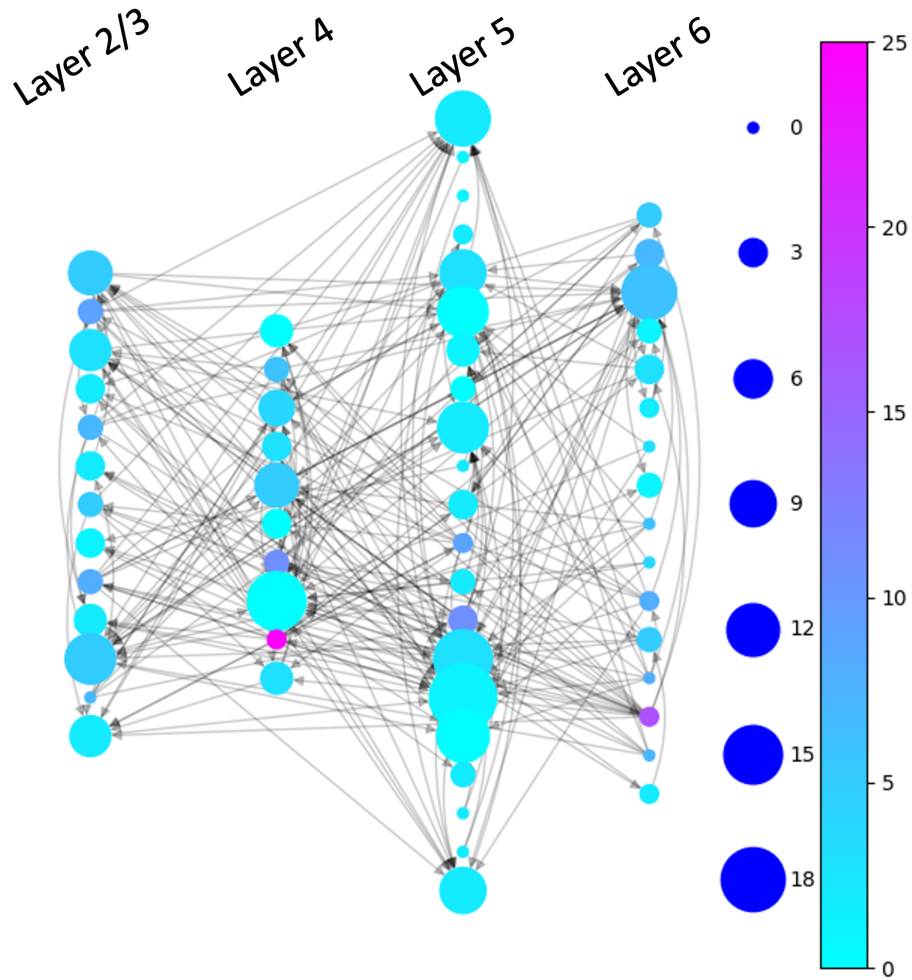
FIGURE A.1. Drawn effective network of sampled neurons in Waksman's right primary visual cortex. Node size indicates in-degree, colour indicates out-degree.