



journal homepage: www.elsevier.com/locate/csbj



Inferring neural information flow from spiking data

Adrià Tauste Campo

Centre for Brain and Cognition, Universitat Pompeu Fabra, Ramon Trias Fargas 25, 08018 Barcelona, Spain



ARTICLE INFO

Article history:

Received 3 July 2020
 Received in revised form 5 September 2020
 Accepted 7 September 2020
 Available online 20 September 2020

Keywords:

Single neuron
 Spike train
 Simultaneous recordings
 Granger causality
 Information flow

ABSTRACT

The brain can be regarded as an information processing system in which neurons store and propagate information about external stimuli and internal processes. Therefore, estimating interactions between neural activity at the cellular scale has significant implications in understanding how neuronal circuits encode and communicate information across brain areas to generate behavior. While the number of simultaneously recorded neurons is growing exponentially, current methods relying only on pairwise statistical dependencies still suffer from a number of conceptual and technical challenges that preclude experimental breakthroughs describing neural information flows. In this review, we examine the evolution of the field over the years, starting from descriptive statistics to model-based and model-free approaches. Then, we discuss in detail the Granger Causality framework, which includes many popular state-of-the-art methods and we highlight some of its limitations from a conceptual and practical estimation perspective. Finally, we discuss directions for future research, including the development of theoretical information flow models and the use of dimensionality reduction techniques to extract relevant interactions from large-scale recording datasets.

© 2020 The Author. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. Introduction	2700
2. From cross-correlations to model-based approaches.	2700
3. The Granger causality framework: main concept and model-free generalizations	2701
3.1. Main concept.	2701
3.2. Model-free generalizations: directed information.	2701
3.3. Model-free generalizations: transfer entropy	2702
3.4. Estimation remarks	2702
3.5. Application to neuroscience studies.	2702
4. GC limitations: estimation and interpretation	2703
4.1. Estimation challenges	2703
4.2. Interpretation issues	2703
5. New directions	2704
5.1. Inferring multivariate interactions via dimensionality reduction	2704
5.2. Introducing the message variable in information flow models	2704
6. Summary and outlook	2706
Declaration of competing interest	2707
Acknowledgements	2707
References	2707

E-mail address: adria.tauste@gmail.com

<https://doi.org/10.1016/j.csbj.2020.09.007>

2001-0370/© 2020 The Author. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

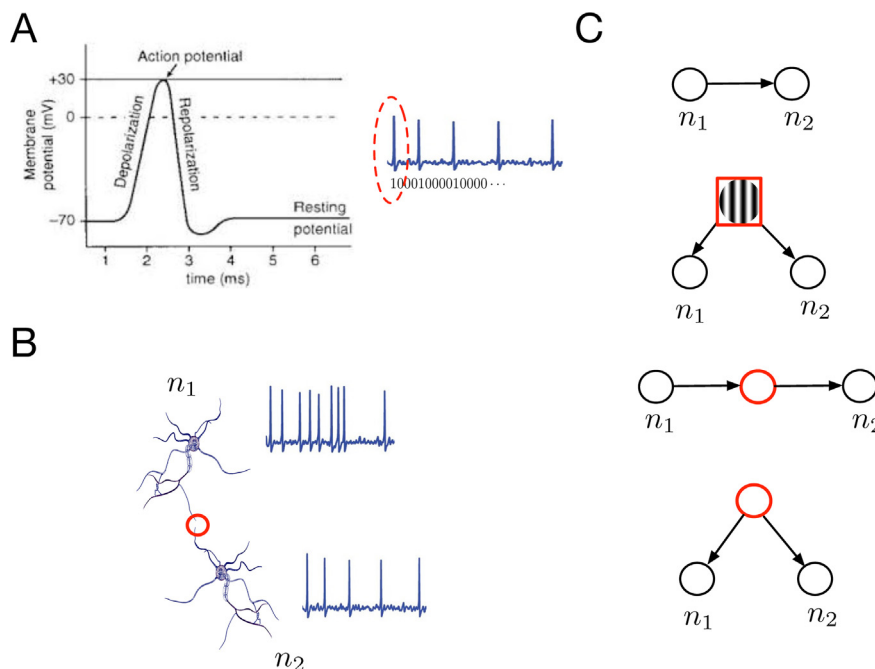


Fig. 1. Inferring single-neuron interactions from spiking data. (A) On the left caption, the time course of an action potential (or spike) showing the rise (“Depolarization”) and fall (“Repolarization”) of the membrane potential with respect to a background level (“Resting potential”). On the right caption, a depiction of a spike train where the first action potential is highlighted in red. Below the spike train, its usual modelization as a binary sequence of 0s and 1-s (spikes) is correspondingly displayed. (B) A schematic depiction of two neurons, n_1 and n_2 , with their respective spike trains, displaying a synaptic connection (in red) between n_1 axon’s terminal and n_2 ’s dendrites. (C) Four model configurations that can explain an estimated pairwise statistical dependence between n_1 and n_2 . On the top caption, a model showing both neurons that are directly connected by a synapse. On the middle-top caption, a model showing a visual stimulus (highlighted in red) exerting a simultaneous effect on both neurons. On the middle-bottom and bottom captions, both neurons being mediated by a third neuron (highlighted in red). The three later examples can be referred to as n_1 and n_2 being independent conditioned to or, equivalently, d-separated by either a stimulus (middle-top) or other neurons’ activity (middle-bottom and bottom) [122]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1. Introduction

A central question in neuroscience research is how the interaction of multiple neurons in the central nervous system leads to cognition. Over years, biology has provided a detailed description of how neurons interact via synapses in terms of electrochemical processes [1]. This interaction is mainly produced by the propagation of action potentials. An action potential (commonly known as a spike) is generated by the abrupt increase and fall of a neuron’s membrane potential. This change of polarization usually occurs in the soma of the neuron and travels down the neuron’s axon towards its terminal to produce electro-chemical signals that are transmitted to the dendrites of synaptically connected neurons, which in turn generate new action potentials (see Fig. 1A and B). Spike propagation is the main means of cell-to-cell communication in the nervous system. Consequently, spikes are analyzed as the main unit of information conveyed by neurons while their temporal sequence of occurrences, known as a spike train, is conceived as the stream of information that travels through the nerves [2]. The usual mathematical symbolization of spike trains is via a binary sequence of 0s and 1s, where the neuron’s time-binned activity is mapped to 1 for spike occurrences and to 0, otherwise ¹ (Fig. 1A). In practice, spikes are measured via extracellular recordings. This type of recordings captures the electrical field generated by the difference in potential between two locations in the extracellular medium [3]. In particular, when these recordings are performed at a very fine scale, spike trains from different neurons can be discriminated by sequentially applying high-

¹ For instance, time bin lengths of 1 ms may be selected so that a maximum of one spike falls per bin.

frequency filtering, spike detection and spike-sorting algorithms on the recorded signals [4].

Nowadays, current technological advances in neural recording systems have allowed to record the electrical activity of an ever-growing number of simultaneous neurons across many species including humans [5]. With these data, one can formulate the general question: Given a subset of simultaneous spike train recordings from different brain areas, how can we reconstruct to a certain precision interactions between the observed neurons to uncover functionally relevant information flows? Despite the interest on the topic, computational approaches to this question are still limited. Indeed, they are diverse in nature, suffer from technical and conceptual shortcomings and can lead to ambiguous biological interpretations. In this paper, we review the main contributions to the topic as well as discuss new promising directions for further development.

2. From cross-correlations to model-based approaches

Since the early birth of simultaneous single-neuron recordings in the mid 1960s [6], neurophysiologists have attempted to jointly analyze and interpret spike trains to provide experimental information about synaptic connections and other potential sources of functional interaction among the detected neurons [7–9]. The initial tools were based on descriptive pairwise statistics such as cross-correlations between neuron’s spike counts, namely the number of spikes over the entire spike train, [7] and bivariate histograms of spike times [8], which were both computed across experimental repetitions (commonly known as *trials*) from the same pair of neurons. Yet, already in 1967, Perkel et al. admitted some of the principal limitations of interpreting neural interactions

via cross-correlation, which also apply to a large ensemble of methods used today [7]. The first limitation is that a pairwise correlation in a neuron pair can be equally explained by a synaptic connection, a third-neuron mediation or by a shared input like stimulus information [7] (see Fig. 1C). The second limitation is that the sequence of trials used for cross-correlation and histogram estimation cannot be in general assumed independent and identically distributed [10] and hence, estimation from multiple trials needs to be performed cautiously. This assumption might be questionable when trials from different days are pooled together or when external and uncontrolled variables (e.g., level of arousal, motivation) have a time-varying effect on subject's behavioral variables (e.g., task performance) across trials.

By the of the 20th century, several works started to address the above-mentioned concerns. On one hand, the authors in [10] were able to develop a robust method to isolate the residual component of the cross-correlation that only accounted for the effect of the stimulus on each neuron's activity. On the other, it was showed that cross-correlations could confound distinct sources of potential covariation that included genuine time synchrony as well as externally-driven covariations of independent neural responses [11]. In this situation, heuristic rules [11] and quantitative methods [12] were proposed to help resolve potential ambiguities and improve the interpretation of experimental outcomes.

While some cross-correlation limitations may be tackled by ad hoc methods [10,12,13], a more general framework is necessary to simultaneously account for all the confounding sources of covariability [14]. In this context, by the end of the last century, several works started to regard spike train sequences in the frequency domain and made use of Fourier methods and spectral measures of association (e.g. coherence) to characterize distinct sources of influence in single-neuron interactions [15–17]. For instance, this approach led to identifying common inputs in pairwise interactions [18] and to the development of partial directed coherence [19,20], a measure of interaction that incorporates directionality and controls for the effect of other observed neurons. In parallel, in the early 2000s, statistical models emerged as a powerful tool to model the influence of covariates such as the stimulus, and the own or other neurons' previous spiking history [21]. Specifically, model-based approaches are grounded on minimal generative assumptions, i.e., how the observed variables are generated, and typically fit the model parameters using maximum-likelihood estimation [22], i.e., choosing those values that maximize the conditional probability of the observed variables given the parameters. A well-known example in neuroscience is the generalized linear model (GLM), which statistically describes spike trains as an inhomogeneous Poisson point processes whose time-varying intensity (also known as Poisson rate) results from a non-linear function of filters, each processing a different variable influencing the neuron's activity, such as the stimulus, the spike train own past activity and the spike trains from other neurons [21]. The use of GLMs has been widely applied to study neural interactions in a number of simultaneous studies [23–26]. For instance, it was showed how retinal cell interactions were more prominent between neighboring cells and how these interactions improved the decoding of visual stimuli [23].

The GLM detailed description of the neurons' spiking activity comes at the expense of a potentially large number of model parameters. This usually can produce poor performance in generalising the results across different experimental sessions. Several studies have overcome this issue by introducing prior knowledge about the observed data. This includes invoking analytical assumptions on the coupling time-varying functions [23] or modelling interaction sparsity using Bayesian inference [27]. Yet, the application of Bayesian inference in this context has its own limitations. Indeed, modelling neural interactions as Bayesian networks with-

out adequate constraints [28] may be computationally unfeasible. Another critical issue is the fact that GLM assumes parameter invariance during repeated experimental trials as we will discuss in the next section.

3. The Granger causality framework: main concept and model-free generalizations

In most applications, a GLM assumes that the underlying Poisson process is stationary within and across trials [21]. Hence, it fits a single coupling filter for each neuron pair across repeated experimental trials, which might obscure the functional relevance of trial-dependent interactions. In particular, trial-to-trial fluctuations can occasionally alter the number of spikes of some driver neurons, producing a larger effect on target neurons during specific trials [29]. We will devote this section to a framework that allows to infer single-trial causal dependencies. We will first state its main concept, we will then define and discuss its generalized information-theoretic formulas and we will conclude by reviewing some applications in neuroscience.

3.1. Main concept

In order to analyze single-trial dependencies, an established approach is to model spike trains as binary time series² and resort to the Granger causality³ (GC) framework [30,31]. Granger's causality is a concept that originated in econometrics in the 1950s whose core idea is the following: a time series X causes⁴ Y if X contains information that helps predict the future of Y better than any information already present in the past of Y , and, if available, in the past of other observed variables Z [31].

3.2. Model-free generalizations: directed information

In its original form, GC was conceived to be applied to multivariate linear autoregressive Gaussian models (MVAR) in both temporal [31] and frequency domains [33,34] but the basic idea can be generalized to arbitrary joint probability distributions governing the observed variables. When an estimation method uniquely relies on the joint probability distribution of the observed variables, it is usually referred to as “model-free” in the neuroscience literature as opposed to “model-based” approaches relying on a predefined statistical model. Crucially, because spike trains are naturally modeled as Poisson and not Gaussian processes [21], model-free methods are more suitably than the GC-MVAR to capture the specificities of spiking activity. In fact, a model-free generalization of the GC concept can be found in the information-theoretic concept of directed information. The directed information (DI) is a functional that was originally developed in [35–37] to study the maximum achievable transmission rates in communication channels with feedback but can also be used to measure causal statistical dependencies between sequences of random variables. Formally, the DI can be defined as the sum of conditional mutual information terms [38], which makes it applicable to arbitrary statistical models and to both discrete and continuous variables.

In the following, we will provide the mathematical definition of DI. Let X, Y and Z be three arbitrary variables. The conditional mutual information between X and Y conditioned on Z is defined as

² Note that a binary time series can be alternatively defined as a sequences of binary variables.

³ It is also known as Granger-Wiener causality to give credit to the earlier related work by N. Wiener [30].

⁴ The causality terminology of the GC framework is interpreted here in a fully statistical sense in contrast to the notion defined by J. Pearl' [32] that involves intervening the observed system.

$$I(X; Y|Z) = E_{XYZ} \left[\log \frac{P_{Y|XZ}}{P_{Y|Z}} \right], \quad (1)$$

where E_{XYZ} denotes the expectation over the joint probability distribution P_{XYZ} . Let us now be more specific and assume that the arbitrary variables above are sequences of random variables. In particular, let us consider two T -length sequences X^T and Y^T defined as $X^T = (X_1, \dots, X_T)$ and $Y^T = (Y_1, \dots, Y_T)$. To introduce the DI, we will make use of the mutual information formula for sequences of variables [38]. The mutual information between X^T and Y^T can be decomposed via the chain rule in the sum of T conditional mutual information terms:

$$I(X^T; Y^T) = \sum_{t=1}^T I(Y_t; X^T | Y^{t-1}), \quad (2)$$

where the notation A_τ^t stands for the sequence $A_\tau^t = (A_\tau, \dots, A_\tau)$, $\tau \geq t$, for which the subscript is dropped when $\tau = 1$ [38]. In contrast to (2), the DI between X^T and Y^T is defined as

$$I(X^T \rightarrow Y^T) = \sum_{t=1}^T I(Y_t; X^t | Y^{t-1}), \quad (3)$$

where, in each summand of (3), the X^T appearing in the second argument of (2) has been replaced by X^t , thus accounting only for the dependency of each Y_t on up to the t th element of X^T . While the mutual information is symmetric, the DI is not, and hence the later yields in general a different value when computed in reverse direction, i.e., from Y^T to X^T , $I(Y^T \rightarrow X^T) \neq I(X^T \rightarrow Y^T)$. Although (3) holds for general statistical models, under certain conditions of stationarity and ergodicity it is more convenient to recall its temporal normalized version, known as the DI rate:

$$\frac{1}{T} I(X^T \rightarrow Y^T). \quad (4)$$

In addition to causal inference, the DI has an operational meaning in different information-theoretic and statistical domains ranging from data compression to channel coding or hypothesis testing [39]. Importantly, the DI is the fundamental limit of communication (that is, the maximum achievable transmission rate) over a certain type of noisy channels when noiseless feedback is present at the transmitter [40,41]. Hence, the DI is not only a convenient measure of causal dependence between data sequences but it is also the theoretical *answer* to problems involving communication models.

Over the last decade, a number of consistent DI rate estimators have appeared in the literature [29,42,43]. For instance, in [29] the authors defined an estimator to infer causal relationships in neural spike trains by assuming a Poisson statistical model and fitting its parameters with GLM over long single trials. Then, the required conditional probabilities of X^T and Y^T were obtained from the model to be plugged into the DI rate formula (4). In the most general case, however, no information about the underlying model is presumed and the joint probability distribution of X^T and Y^T needs to be estimated in a non-parametric form. Under this condition, novel DI rate estimators were defined in [42], where the estimator relied on a sequential and universal probability estimation algorithm named context tree weighting (CTW, [44]), and in [43], where the authors analyzed the performance limits of the probability maximum-likelihood estimator. Importantly, in all the above cases, the estimation procedure becomes computationally feasible when the sequences X^T and Y^T are assumed to be generated according to jointly stationary and ergodic Markov processes [45].

3.3. Model-free generalizations: transfer entropy

Consistent with the GC concept, another information-theoretic functional was independently proposed under the name of transfer entropy (TE), aimed to measure causal dependencies between random processes in dynamical systems [46]. Unlike the DI, the TE only applies to pairs of stationary processes, X_t and Y_t , that jointly satisfy the Markov property, i.e.,

$$P_{Y_{t+1}|X^t, Y^t} = P_{Y_{t+1}|X_{t-K+1}^t, Y_{t-J+1}^t} \quad (5)$$

for any $t \geq \max(J, K)$, where J and K are the order of each process, respectively. Given (5), the TE between processes X_t and Y_t is defined as

$$TE(X \rightarrow Y) = I(Y_{t+1}; X_{t-K+1}^t | Y_{t-J+1}^t), \quad (6)$$

for any $t \geq \max(J, K)$ [47]. Under the usual assumptions of DI rate estimators (stationarity, ergodicity and the Markov property), it can be easily checked that the DI rate converges to the TE in the limit of the sequence length T [48]. Furthermore, when these conditions hold in Gaussian models, it can be shown that both DI and TE coincide with the MVAR version of GC [49]. Similarly to MVAR models, the DI, the TE and other GC-derived measures can also be extended via conditioning to measure conditional causal dependencies in multivariate models. Examples of such multivariate extensions have been theoretically proposed for the DI [50,51], for the TE [52,53] as well as for other GC-derived measures [54].

3.4. Estimation remarks

When estimating model-free GC-derived measures, both the outer expectations and the inner (conditional) probability distributions appearing in (1) are approximated by leveraging a sufficiently large number of temporal samples from the observed time series. Therefore, in this type of estimations, there is a trade-off between the assumptions of stationarity and ergodicity usually holding at short segment lengths and the estimation power requiring lengthy time series. These constraints do not apply in other reviewed methods such as cross-correlation in which samples are obtained from the number of trials over which a certain quantity is averaged. Critically, in neuroscience studies, one may argue that the use of temporal samples (instead of trials) may compromise the inference of the exact times at which spike-train interactions occur. However, over the last years, a few works have shown that interaction times can also be revealed in this framework via delayed versions of the original measures [55] and ad hoc statistical tests (e.g., see Supplementary information in [56]). Finally, the statistical power of model-free GC-derived measures can be assessed by performing nonparametric significance testing of the estimated quantities using methods such as permutation tests [57].

3.5. Application to neuroscience studies

Since the early 2000s, a number of data-driven methods derived within the GC framework have been applied to pairs of simultaneously recorded neurons in order to investigate how information flows between brain areas are associated to cognitive functions.

Because GC was originally aimed to analyze continuous-value time series, the classic MVAR formulation of GC [58] is not a priori suitable to deal with binary spike trains. However, some works circumvented this issue by developing variants of the original method. For instance, in experimental studies about visual information processing [59,60], a non-parametric version of the original GC estimation in the frequency domain [61] was applied to spike trains, thus bypassing the point-process modelling [21]. This

approach was specifically tested with recordings from visual areas, while monkeys were exposed to visual stimuli [59,60]. In this application, Hirabayashi et al. highlighted the temporal recurrence of feedforward and feedback interactions in the same pair of neurons during stimulus presentation [59]. An alternative approach was due to Kim et al. who kept the point process modelling of spike trains [21] and proposed a Poisson-log likelihood version of the original GC measure [58].

The application of the DI to simultaneous single-neuron datasets became specially popular after its adequacy to handle spiking data was demonstrated in [29]. Specifically, in [29], the proposed DI rate estimator was applied to recordings from the primary motor cortex (M1) of a monkey while it performed arm movement tasks according to visual targets. The outcomes of the analysis supported the existence of electrical propagation waves above 10 Hz, which are known to encode information about visual targets in reaching tasks [62]. In addition, a variant of the DI rate estimator introduced in [29] was proposed in [63], which showed an accurate estimation of the conduction delays between neurons in different brain areas during motor tasks performed by rodents and nonhuman primates. On the other hand, time-delayed versions of the CTW-based estimator were elaborated in [56,64] to infer task-driven directional interactions between the thalamus and the somatosensory area 1 (S1) in monkeys performing a tactile detection task [56], and across cortical somatosensory, premotor and motor areas in monkeys performing a tactile dissemination task [64]. Finally, an extension of the CTW algorithm for non-necessarily finite-order Markov processes [65] was used to estimate the DI rate between neural spike trains from the buccal ganglion of *Aplysia* [66].

4. GC limitations: estimation and interpretation

Over the last couple of decades, the GC framework has become one of the main statistical method in neuroscience to analyze neural interactions from a variety of recording modalities including spike trains. Despite its growing popularity, its practical application has also raised some concerns [67–69], about the computational reliability of the estimated outcomes and their biological interpretation. In this section, we will review two sources of criticism about GC-derived measures: those concerning their estimation, and those related to the information flow interpretation of their outcomes.

4.1. Estimation challenges

The original formulation of the GC concept relying on linear Gaussian statistics has been refined in the frequency domain to resolve some of its initial technical limitations such as the bias and high variance of the interactions estimates [58,61]. However, some additional challenges prevail such as the validity of the linearity and stationarity assumptions, or the effect of temporal sampling [69,70,71], which may impair its application to spike train data. In fact, the use of model-free generalizations like the DI or TE resolves the linearity assumption, but it is still susceptible to problems such as the estimation bias or the lack of stationarity in data recordings. Nevertheless, recent works have showed promise in dealing with these later issues. For instance, Schamberg et al. showed that the above reviewed DI estimators are biased when the Markov order of the receiving process Y^T is different from the order of the joint processes (X^T, Y^T) [45]. In addition, they outlined sufficient conditions under which the equal-order Markovian assumption is met and provided a bound for the estimation bias in those cases when such conditions may not be satisfied. To address the non-stationarity problem, Sheikhattar et al. developed a window-based adaptive model that makes uses of point-process

modelling and leverages the sparsity of spiking data [72]. They applied this technique to simultaneous recordings from ferrets to describe time-varying top-down and bottom-up interactions between primary auditory area (A1) and prefrontal cortex (PFC) during a tone detection task.

4.2. Interpretation issues

One of the fundamental criticism about the GC statistical framework in general, and about GC-derived measures in particular, is the interpretation of the inference outcomes as characterizing information flows between neurons. Importantly, a review of the recent literature [67,69,73,71] readily reveals that some of the controversy mainly arises due to the different notions of information flow that researchers adopt in their studies. Hence, we might start asking the conceptual question: what do we understand by information flow?

To begin with, if measuring information flow means detecting the exchange of information between neuron A and neuron B through their synaptic connections, then the GC framework (and also the GLM) alone is in general insufficient to address this question. This is because the GC concept and its information-theoretic generalizations are aimed to infer statistical dependencies between observed variables and, therefore, its application to spike train data characterizes single-neuron interactions only at a phenomenological level. As such, GC-derived measures are susceptible to latent confounding effects arising from limited spatial sampling such as the influence of unobserved neurons. Indeed, given the thousands of neurons that may have an effect on a single postsynaptic neuron, the GC estimates are in general not able to discriminate between anatomically direct or indirect connections. Instead, if we wish to make detailed inferences about synaptic connections or other sources of interactions, mechanistic approaches are required. An example of such approaches is dynamic causal modeling (DCM) [74], a widely established framework to analyze coarser neural data modalities like functional magnetic resonance (fMRI) or electroencephalography (EEG) [75]. Specifically, DCM assumes an underlying causal model with biophysically plausible properties and estimates its parameters via Bayesian inference [74].

Alternatively, we can assume, in a weaker sense, that information flow across or within brain regions is mapped into certain *meaningful* causal dependencies between neuron's spike trains. By meaningful, we may understand that these dependencies map either anatomically direct or indirect neural interactions that are consistent with the processing of external stimuli or internally built actions (“information”) along a functional pathway (“flow”). Under this definition, we may include the biological interpretation employed in most of the studies reviewed in Section 3.5. Since GC-derived measures estimate causal dependencies, they can be used in this context but its application needs to be made with caution. Indeed, one of the main highlighted issues [76,77,67] is the fact that GC-derived measures only capture pairwise dependencies and hence, they conflate different sources of dependency when certain information is shared by more than two variables. This can be illustrated by a simple example given in [67]. Consider two sequences X^T and Y^T , where X^T is a sequence of independent and identically distributed Bernoulli variables with parameter $p = 1/2$, and Y^T is defined as follows:

$$Y_1 \sim \text{Bern}(1/2) \quad (7)$$

$$Y_t = X_{t-1} \oplus Y_{t-1}, \quad 2 \leq t \leq T, \quad (8)$$

where \oplus stands for the XOR operator between binary values⁵. If we compute the DI between X^T and Y^T using the binary logarithm, we

⁵ The XOR operator satisfies $0 \oplus 0 = 0, 0 \oplus 1 = 1, 1 \oplus 0 = 1, 1 \oplus 1 = 0$.

find that each term $I(Y_t; X^t | Y^{t-1}) = 1$ bit, for $1 \leq t \leq T$, and thus, the DI rate (4) equals 1bit, and as a non-zero quantity, it measures that Y^T causally depends on X^T . However, a closer look at the model shows that the second argument of the conditional mutual information in (3), i.e., the truncated sequence X^t , cannot predict alone the variable Y_t for any $t \geq 2$. Hence, the estimated causal dependence is not uncovering a genuine information flow from X^T to Y^T because, in this example, it is the combination of the past of Y^T and X^T which contributes to the present of Y^T .

At the core of the above example, it lies the following theoretical fact: a straightforward application of the conditional mutual information fails in general to describe dependencies between random variables beyond pairwise interactions [77,67] (e.g., in the above example there is a third-order dependence between Y_t, X^t and Y^{t-1}). This is indeed a critical problem in the field since a certain type of higher-than-two order interactions called synergistic have been found in several neuroscience studies [78–80]. To integrate these additional sources of interaction in the analysis, one can resort to the partial information decomposition (PID) framework proposed in [76]. Briefly, the PID decomposes the mutual information that a set of variables A_1, A_2, \dots, A_n has about a variable B , i.e., $I(B; A_1, A_2, \dots, A_n)$, into the information that the variables A_i provide individually (unique information), redundantly (shared information) or only jointly (synergistic information) about B .

More recently, a more conceptual limitation of GC-derived measures is gaining attention in the literature. In our second information flow interpretation, we required that causal dependencies were part of a functional path that processed information content. In practice, if this information content is measurable we can make the requirement more specific and ask the estimated interactions to be statistically associated to an information message (an external stimulus, internal command, etc.), as it is considered in theoretical communication models [81]. In other words, causal dependencies need to be *about a message* [82]. Surprisingly enough, the effect of stimulus on estimated neural interactions has been to date largely neglected or uniquely considered as a source of covariation [21]. However, there is a growing consensus that the relationship between the stimulus (or any internal variable) and the estimated interactions is a necessary condition to support the information flow interpretation [64,56,83,84,82]. For instance, the use of GC-derived measures to analyze single-trial and time-varying neural interactions in monkeys performing perceptual decision-making tasks have showed the modulatory effect of stimulus information and internal percepts into inter-area interactions [64,56].

5. New directions

In the following, we will overview two trends that have made recent progress in tackling some of the GC framework challenges discussed in Section 4. The first is motivated by current technology developments: it assesses whether we can benefit from large scale recordings and dimensionality reduction techniques to estimate functionally relevant neural interactions that are obscure to pairwise statistics. The second concerns the development of theoretical models and measures aimed to integrate the message statistics in order to improve the information flow interpretation.

5.1. Inferring multivariate interactions via dimensionality reduction

Over the last years, neural recording advances have made possible to record up to thousands of neurons simultaneously [85] keeping a pace that is growing at an exponential rate [5]. As a consequence, researchers have started to regard data analysis as a

multi-dimensional problem with one of the key dimensions being the number of simultaneously recorded neurons. In this context, the classical notion of single-neuron activity has been replaced by that of population activity, which has been correlated with sensory stimuli, behavioral variables and between ensembles of simultaneous spike trains from different brain areas [86].

A key aspect of this approach is the use of dimensionality reduction techniques to extract robust and interpretable information from multivariate recording sets [87,88]. Examples of applied techniques are principal component analysis (PCA), factor analysis (FA) or tensor decomposition analysis [89–91], among others. Rather than spike trains, these techniques are typically applied over sequences of firing rates, obtained as the normalized number of spikes in a certain time window, which allows for multivariate Gaussian modelling. Using this framework, most studies have analyzed how distinct information features about stimuli [92,93] or motor actions [94,95] were encoded into lower-dimensional population activity subspaces, i.e. firing-rate subspaces that were of lower rank than the number of recorded neurons. In contrast, less work has been devoted to reformulate the study of spike-train dependencies at a population level and complement the above-reviewed approaches (GLM, GC). Yet, there are some interesting directions pointed in the recent literature [94,96]. For instance, in the context of a motor task performed by macaque monkeys, Kaufman et al. investigated the communication mechanisms under which some information (muscle-related) lying in motor cortical areas flowed to the spinal cord and muscles, while some other (preparatory-related) largely stayed in the cortex [94]. Their analysis showed that the same population of neurons could project different sources of information (muscle or preparatory) into distinct activity subspaces, and these subspaces allowed to selectively route the appropriate information source (in this case, muscle-related) towards target regions such as the spinal cord and the muscles. Using similar methods, Semedo et al. more recently studied the structure of population interactions between the primary visual (V1) and the secondary visual (V2) brain areas in anesthetized macaque monkeys [96]. Their results concluded that V1 makes use of different population subspaces for intra-area and inter-area interactions, respectively. In particular, they showed that the V1-V2 interaction subspace (named *communication subspace*) lying in V1 is of lower dimension and disjoint with respect to the V1 subspace capturing intra-area interactions (see Fig. 2). As a consequence, V2 population activity is related to a small subset of V1 population activity patterns, which differ from the most prevalent patterns shared by V1 neurons. These findings support the hypothesis introduced in [94] that neural population subspaces constitute a mechanism to route information across brain areas. Even though dimensionality reduction is a powerful ensemble of tools to deal with high-dimensional datasets, its current application to neuroscience has some limitations when it comes to inferring results in terms of information flow. To name a few, the lack of directionality or explicit stimulus variables in the models, and the special focus on PCA-related methods relying on Gaussian assumptions and variance maximization. In particular, to overcome the later issue, nonparametric generalizations of PCA such as projection pursuit [97] could be applied to ensembles of non-Gaussian firing rates for which functions other than the variance (e.g., skewness [98], entropy [99]) are optimized in order to unravel interesting lower-dimensional projections. Finally, dimensionality reduction techniques could be applied in non-linear models via the use of embedding methods [100].

5.2. Introducing the message variable in information flow models

The message, that is, the source random variable that needs to be transmitted over a network from an origin to a destination, is

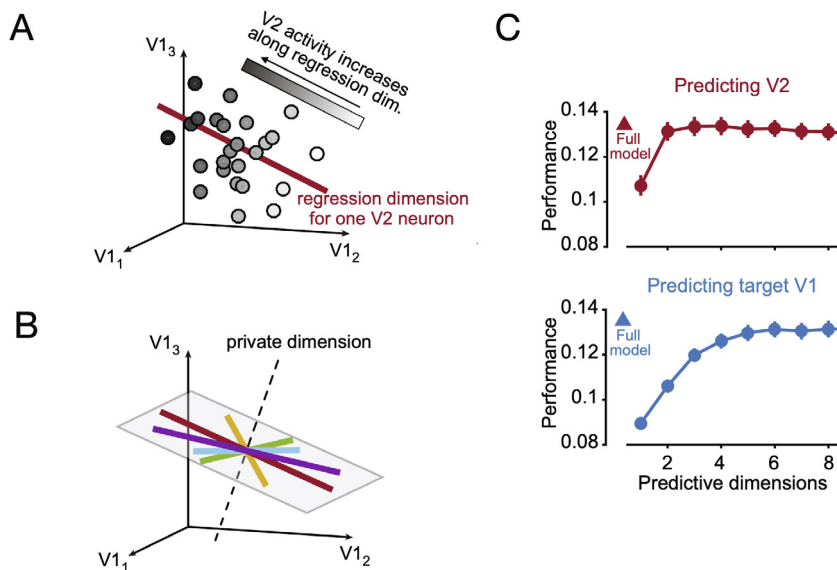


Fig. 2. New directions: Inferring population interactions from multivariate datasets. Extracted from [96] (Fig.3 and Fig. 4) with permission. (A) Graphical depiction of linear regression between a population of V1 (primary visual brain area) neurons and one V2 (secondary visual brain area) neuron. Each circle represents the activity recorded simultaneously in V1 (three neurons) and V2 (one neuron) during an observation sample. The position of the circle represents the V1 population activity and its shading represents the activity of the V2 neuron. The activity of the V2 neuron increases along the regression dimension (red line). (B) Low-dimensional population interaction. The regression dimensions (shown as multiple-color straight lines) for different V2 neurons span a 2-dimensional subspace (the gray plane) of the V1 population space. Thus, two predictive dimensions are sufficient to capture the inter-area interactions between V1 and V2. All dimensions that are not predictive of V2, and therefore lie outside of this subspace, are called private dimensions. (C) Top: The number of predictive dimensions of V2 (red circles) in V1 needed to achieve full predictive performance (i.e., using all V1 neurons, red triangle) is two dimensions. Bottom: In contrast, the number of predictive dimensions of V1 (blue circles) in V1 needed to achieve full predictive performance (i.e., using all V1 neurons, blue triangle) is six dimensions. Error bars shows the standard error of the mean across different datasets. Adapted by permission from Elsevier Ltd. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

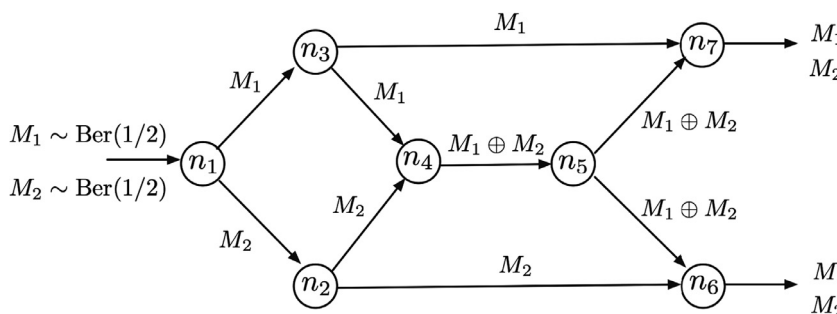


Fig. 3. New directions: Information flow models with message variables and the need to address higher-than-two order dependencies. A depiction of the butterfly network introduced in [102] and used as an example of application in [82], in which higher-than-two dependencies arise among the observed variables. In this case, two binary messages M_1 and M_2 modeled as independent Bernoulli variables with parameter $p = 1/2$ are transmitted from a source neuron (n_1) to two destination neurons (n_6 and n_7) travelling through intermediate nodes (illustrated with the corresponding message displayed on each travelled edge). Along the transmission, all neurons relay their incoming information except n_4 , which performs the XOR operation of their incoming messages. For instance, in this network a third-order dependence may arise between the output activity of n_2 (neuron uniquely conveying M_2 information), n_3 (neuron uniquely conveying M_1) and n_4 (neuron uniquely conveying $M_1 \oplus M_2$). In contrast, it can be checked that all pairs among these 3 variables are marginally (pairwise) independent.

a key component in all theorized communication models [81,101,102], Hence, when aiming to interpret spike trains dependencies as information flow, we may ask: What is the information source that these dependencies convey? To address this question, recent works [82,84] have attempted to develop novel models and measures that infer the existence of information flow (or information transfer) by analyzing the interplay between recorded neural activity and the message variable (e.g., sensory stimulus) that is expected to flow.

From a theoretical perspective and largely inspired by information theory models [81,102], Venkatesh et al. have proposed a formal definition of information flow that explicitly includes the message as a model variable [82]. This definition is formulated in the framework of computational systems, which are defined as time-indexed directed graphs where node “transmissions” are

modeled as random variables associated to their outgoing edges, where “computations” over each transmission are performed at each arriving node, and where there exists a subset of nodes (“input nodes”) whose transmissions depend on the message variable at time $t = 0$. Then, information about the message flows on an edge as long as the mutual information between the corresponding edge random variable and the (discrete-valued) message conditioned on a set of additional edge variables has a positive value. The author’s definition is time-dependent since it assumes varying statistics over different observation time points and can be naturally extended to characterize information paths between pairs of nodes. Importantly, this approach specifically deals with the existence of the high-order dependencies reviewed in Section 4.2, which might arise between the observed edge variables and the message (see Fig. 3 for an exemplary network where this type of

Table 1

Published applications to real spike-train data with open-access online data or software. V1: Primary visual area; V2: Secondary visual area. MT: Middle temporal area; LIP: Lateral intraparietal area; FEF: Frontal eye fields; A1: Primary auditory area; PFC: Prefrontal cortex area; VPL: Ventral posterolateral nucleus of the thalamus; S1: Primary somatosensory area; PMd: Premotor cortex; M1: Primary motor area.

Method	Simultaneous dataset	Online data/software	Year
Cross-correlation [13]	V1 and V2 area neurons from anesthetized macaque monkeys during visual stimulation	doi.org/10.6080/KOB27SHN	2015
GLM [23]	In-vitro ganglion cells from macaque monkeys during visual stimulation	github.com/pillowlab/neuroGLM	2008
GLM [24]	MT and LIP area neurons from macaque monkeys performing a visual task	github.com/jcbyts/mtlipglm	2017
GLM [25]	Rat hippocampus during exploration of an open square field	github.com/NII-Kobayashi/GLMCC	2019
GLM [26]	LIP and FEF area neurons from macaque monkeys performing a visual task	doi.org/10.5061/dryad.gb5mklwk7	2020
GC [72]	A1 and PFC area neurons from ferrets performing an auditory task	github.com/Arsha89/AGC_Analysis	2018
GC-DI [56]	VPL and S1 area neurons from macaque monkeys performing a somatosensory task	github.com/AdTau/DI-Inference	2019
Dim. reduction [94]	PMd and M1 area neurons from macaque monkeys performing a motor task	github.com/ripple-neuro	2014
Dim. reduction [96]	V1 and V2 area neurons from anesthetized macaque monkeys during visual stimulation	https://github.com/joao-semedo/communication-subspace	2019

dependencies are present). Consistent with their definition, the authors provide an information flow inference method consisting of a set of conditional mutual information tests between the stimulus message and the recorded neural activity variables. Although the method might be in practice computationally costly and susceptible to common problems such as the effect of hidden variables, the overall proposal constitutes a valuable effort with theoretical and practical implications (see [82, Section VII]).

At a more practical level, Bim et al. tackled a similar problem and proposed a directed pairwise correlation measure that determines whether a causal dependence between two spike trains is about a certain stimulus feature [84]. In particular, their measure applies the notions of redundancy and uniqueness from PID theory [76,77] as follows: it quantifies the information about the stimulus in the target spike train that is redundant with the information at the driver spike train and unique with respect to the information already available in the past activity of the target spike train. Consequently, this measure simultaneously addresses the presence of some high-order dependencies in the observed data and the required existence of information content during information transfer. However, because it strictly applies to pairs of spike trains, it cannot be a priori generalized to detect the variety of information flow mechanisms that might be present at a network level [102] (see also Fig. 3).

6. Summary and outlook

We have discussed the problem of modeling and inferring single-neuron and population interactions to detect neural information flows from the pioneering use of cross-correlations [8] to the most recent methods [72,82,96]. In particular, we have seen the evolution of model-based and model-free approaches to face technical estimation problems and allow meaningful biological interpretations. A special attention has been paid to the specificities and challenges of a widely established framework such as Granger causality. Finally, we have outlined new research lines that attempt to address some of the reviewed challenges.

As seen, this field has always been constrained by the technical difficulty of isolating the activity of multiple neurons from different brain areas simultaneously [103]. Critically, we are living an epoch of rapid technological advances in neural recordings [104] and the amount of available data requires improving the performance capacity and computing resources of current methods. In this paper, we have mainly referred to electrophysiological recordings (see Table 1 for a summary of applied studies with open-access data or software). However, it is due mentioning that a new generation of imaging recording methods relying on fluorescence molecule indicators [105] have been able to record the activ-

ity of more than 10,000 neurons simultaneously [106]. Consequently, these methods hugely increase the single-neuron recordings' spatial resolution at the expense of reducing the temporal resolution [107] to detect spike trains⁶. In conclusion, similar techniques like the ones reviewed here can be employed to analyze single-neuron interactions from imaging data as there are already examples in the literature [109,110].

Regardless of how neural data is recorded (e.g., electrophysiology or imaging techniques), there are different challenges that need to be tackled in the upcoming years. Below we outlook some of those from the conceptual and estimation angles, respectively. Conceptually, prior to following a model-based or model-free approach, it is critical to understand the limitations of the dataset at hand and appropriately define the notion of information flow that will be investigated in the study. Then, according to the defined notion, it is desirable to choose a proper method (e.g., GLM, DI, TE, PCA-based) and validate to a reasonable extent its assumptions on the data (e.g., trial independence, time series stationarity) to be able to make statistical inferences and interpretations [111].

There are still several statistical estimation issues that require further development such as the problem of non-stationarity data, the curse of dimensionality when aggregating multiple neurons, the observation noise, among others. However, recent developments such as combining data observation with model prior information (e.g., network sparsity, lower dimensionality activity) [72,96], or simultaneously recording single neurons with surrounding aggregated neural activity [3,112] have brought light to the above problems. An important aspect characterizing some of the methods reviewed in the paper is whether they use single or multiple trials to infer interactions associated to information flow. For instance, multiple trials are needed to evaluate dependencies between the information message and neural spike trains [82] because the former is usually variable only across trials. On the other hand, spike-train interactions should be validated at a single-trial level due to its possible variable statistics during repeated trials [56].

Due to the above mentioned limitations, spike-train inference methods are still far from providing a complete description of the spatial and temporal mechanisms by which multiple neurons communicate information between each other. Over the last two decades, we have experienced the rise and consolidation of GLM and GC approaches and we believe that we are about to witness a fruitful evolution of the topic in the next years thanks to novel theoretical [82] and practical insights [72,84,96]. This will eventually deepen our understanding on the inference of neural informa-

⁶ Yet, there are recent improvements along this line [108].

tion flows, widen its application scope and provide a more unified approach to address biological questions by leveraging its connection to interactions estimated at larger recording scales [113–115], to computational models [86,116], or to results obtained from other related paradigms such as neural population coding [117–120] or network science [121].

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Bial Foundation grant 106/18 and by the project "Clúster Emergent del Cervell Humà" (CECH) ref. 001-P-001682. This later project is co-financed by the European Regional Development Fund Operational Program of Catalonia 2014–2020. I am grateful to J. Barbosa and M. Vila-Vidal for the critical reading of manuscript, and to A. Hyafil, Il Memming Parl and Jacob Yates for fruitful discussions about the generalized linear model.

References

- [1] Koch C. *Biophysics of computation: information processing in single neurons*. Oxford University Press; 2004.
- [2] Bialek W, Rieke F, De Ruyter Van R, Warland Steveninck D. Reading a neural code. *Science* 1991;252(5014):1854–7.
- [3] Buzsáki G, Anastassiou C, Koch C. The origin of extracellular fields and currents: eeg, ecog, lfp and spikes. *Nat Rev Neurosci* 2012;13(6):407.
- [4] Rossant C, Kadir SN, Goodman DF, Schulman J, Hunter ML, Saleem AB, Grosmark A, Belluscio M, Denfield GH, Ecker AS, et al. Spike sorting for large, dense electrode arrays. *Nat Neurosci* 2016;19(4):634–41.
- [5] Stevenson IH, Kording KP. How advances in neural recording affect data analysis. *Nat Neurosci* 2011;14(2):139.
- [6] Gerstein G, Clark W. Simultaneous studies of firing patterns in several neurons. *Science* 1964;143(3612):1325–7.
- [7] Perkel DH, Gerstein GL, Moore GP. Neuronal spike trains and stochastic point processes: II. Simultaneous spike trains. *Biophys J* 1967;7(4):419–40.
- [8] Gerstein GL, Perkel DH. Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science* 1969;164(3881):828–30.
- [9] Moore GP, Segundo JP, Perkel DH, Levitan H. Statistical signs of synaptic interaction in neurons. *Biophys J* 1970;10(9):876.
- [10] Aertsen A, Gerstein G, Habib M, Palm G. Dynamics of neuronal firing correlation: modulation of effective connectivity. *J Neurophysiol* 1989;61(5):900–17.
- [11] Brody CD. Correlations without synchrony. *Neural Comput* 1999;11(7):1537–51.
- [12] Brody CD. Disambiguating different covariance types. *Neural Comput* 1999;11(7):1527–35.
- [13] Zandvakili A, Kohn A. Coordinated neuronal activity enhances corticocortical communication. *Neuron* 2015;87(4):827–39.
- [14] Stevenson IH, Rebesco JM, Miller LE, Kording KP. Inferring functional connections between neurons. *Curr Opin Neurobiol* 2008;18(6):582–8.
- [15] Saito Y, Harashima H. Tracking of information within multichannel eeg record causal analysis in eeg. In: Yamaguchi N, Fujisawa K, editors. *Recent advances in (EEG) and (EMG) data processing*; 1981..
- [16] Kamiński MJ, Blinowska KJ. A new method of the description of the information flow in the brain structures. *Biol Cybern* 1991;65(3):203–10.
- [17] Kamiński M, Blinowska K, Szelenberger W. Topographic analysis of coherence and propagation of eeg activity during sleep and wakefulness. *Electroencephalogr Clin Neurophysiol* 1997;102(3):216–27.
- [18] Rosenberg J, Halliday D, Breeze P, Conway B. Identification of patterns of neuronal connectivity—partial spectra, partial coherence, and neuronal interactions. *J Neurosci Methods* 1998;83(1):57–72.
- [19] Sameshima K, Baccalá LA. Using partial directed coherence to describe neuronal ensemble interactions. *J Neurosci Methods* 1999;94(1):93–103.
- [20] Baccalá LA, Sameshima K. Partial directed coherence: a new concept in neural structure determination. *Biol Cybern* 2001;84(6):463–74.
- [21] Truccolo W, Eden UT, Fellous MR, Donoghue JP, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol* 2005;93(2):1074–89.
- [22] Brown EN, Barbieri R, Eden UT, Frank LM. Likelihood methods for neural spike train data analysis. *Comput Neurosci Comprehens Approach*; 2003: 253–286..
- [23] Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky E, Simoncelli EP. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 2008;454(7207):995–9.
- [24] Yates JL, Park IM, Katz LN, Pillow JW, Huk AC. Functional dissection of signal and noise in mt and lip during decision-making. *Nat Neurosci* 2017;20(9):1285.
- [25] Kobayashi R, Kurita S, Kurth A, Kitano K, Mizuseki K, Diesmann M, Richmond BJ, Shinomoto S. Reconstructing neuronal circuitry from parallel spike trains. *Nat Commun* 2019;10(1):1–13.
- [26] Hart E, Huk AC. Recurrent circuit dynamics underlie persistent activity in the macaque frontoparietal network. *eLife* 2020;9:e52460.
- [27] Stevenson IH, Rebesco JM, Hatsopoulos NG, Haga Z, Miller LE, Kording KP. Bayesian inference of functional connectivity and network structure from spikes. *IEEE Trans Neural Syst Rehab Eng* 2008;17(3):203–13.
- [28] Eldawlatly S, Zhou Y, Jin R, Oweiss KG. On the use of dynamic bayesian networks in reconstructing functional neuronal networks from spike train ensembles. *Neural Comput* 2010;22(1):158–89.
- [29] Quinn CJ, Coleman TP, Kiyavash N, Hatsopoulos NG. Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. *J Comput Neurosci* 2011;30(1):17–44.
- [30] Wiener N. *The theory of prediction*. Mod Math Eng..
- [31] Granger CW. Investigating causal relations by econometric models and cross-spectral methods. *Econometr J Econ Soc* 1969:424–38.
- [32] Pearl J. *Causality*. Cambridge University Press; 2009.
- [33] Geweke J. Measurement of linear dependence and feedback between multiple time series. *J Am Stat Assoc* 1982;77(378):304–13.
- [34] Geweke JF. Measures of conditional linear dependence and feedback between time series. *J Am Stat Assoc* 1984;79(388):907–15.
- [35] Marko H. The bidirectional communication theory—a generalization of information theory. *IEEE Trans Commun* 1973;21(12):1345–51.
- [36] Rissanen J, Wax M. Measures of mutual and causal dependence between two time series (corresp.). *IEEE Trans Inf Theory* 1987;33(4):598–601.
- [37] Massey J. *Causality, feedback and directed information*. In: *Proc Int Symp Inf Theory Applic (ISITA-90)*, Citeseer; 1990. p. 303–305..
- [38] Cover T, Thomas J. *Elements of information theory*. NJ: Wiley-Interscience; 2006.
- [39] Permuter HH, Kim Y-H, Weissman T. Interpretations of directed information in portfolio theory, data compression, and hypothesis testing. *IEEE Trans Inf Theory* 2011;57(6):3248–59.
- [40] Tatikonda S, Mitter S. The capacity of channels with feedback. *IEEE Trans Inf Theory* 2008;55(1):323–49.
- [41] Kim Y-H. A coding theorem for a class of stationary channels with feedback. *IEEE Trans Inf Theory* 2008;54(4):1488–99.
- [42] Jiao J, Permuter HH, Zhao L, Kim Y-H, Weissman T. Universal estimation of directed information. *IEEE Trans Inf Theory* 2013;59(10):6220–42.
- [43] Kontogiannis I, Skoulariidou M. Estimating the directed information and testing for causality. *IEEE Trans Inf Theory* 2016;62(11):6053–67.
- [44] Willems FM, Shtarkov YM, Tjalkens TJ. The context-tree weighting method: basic properties. *IEEE Trans Inf Theory* 1995;41(3):653–64.
- [45] Schamberg G, Coleman TP. On the bias of directed information estimators. In: *2019 IEEE international symposium on information theory (ISIT)*. IEEE; 2019. p. 186–190..
- [46] Schreiber T. Measuring information transfer. *Phys Rev Lett* 2000;85(2):461.
- [47] Vicente R, Wibral M, Lindner M, Pipa G. Transfer entropy: a model-free measure of effective connectivity for the neurosciences. *J Comput Neurosci* 2011;30(1):45–67.
- [48] Amblard P-O, Michel OJ. The relation between granger causality and directed information theory: a review. *Entropy* 2013;15(1):113–43.
- [49] Barnett L, Barrett AB, Seth AK. Granger causality and transfer entropy are equivalent for gaussian variables. *Phys Rev Lett* 2009;103(23):238701.
- [50] Amblard P-O, Michel OJ. On directed information theory and granger causality graphs. *J Comput Neurosci* 2011;30(1):7–16.
- [51] Quinn C, Kiyavash N, Coleman T. Directed information graphs. *IEEE Trans Inf Theory* 2015;61(12):6887–909.
- [52] Runge J, Heitzig J, Petoukhov V, Kurths J. Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys Rev Lett* 2012;108(25):258701.
- [53] Lizier JT, Rubinov M. Inferring effective computational connectivity using incrementally conditioned multivariate transfer entropy. *BMC Neurosci* 2013;14(S1):P337.
- [54] Eichler M. On the evaluation of information flow in multivariate systems by the directed transfer function. *Biol Cybern* 2006;94(6):469–82.
- [55] Wibral M, Pampu N, Priesemann V, Siebenhühner F, Seiwert H, Lindner M, Lizier JT, Vicente R. Measuring information-transfer delays. *PLoS One* 8(2):.
- [56] Tauste Campo A, Vázquez Y, Álvarez M, Zainos A, Rossi-Pool R, Deco G, Romo R. Feed-forward information and zero-lag synchronization in the sensory thalamocortical circuit are modulated during stimulus perception. *Proc Natl Acad Sci* 2019;116(15):7513–22.
- [57] Cohen MX. *Analyzing neural time series data: theory and practice*. MIT press; 2014.
- [58] Seth AK, Barrett AB, Barnett L. Granger causality analysis in neuroscience and neuroimaging. *J Neurosci* 2015;35(8):3293–7.
- [59] Hirabayashi T, Takeuchi D, Tamura K, Miyashita Y. Triphasic dynamics of stimulus-dependent information flow between single neurons in macaque inferior temporal cortex. *J Neurosci* 2010;30(31):10407–21.

- [60] Liang H, Gong X, Chen M, Yan Y, Li W, Gilbert CD. Interactions between feedback and lateral connections in the primary visual cortex. *Proc Natl Acad Sci* 2017;114(32):8637–42.
- [61] Dhamala M, Rangarajan G, Ding M. Analyzing information flow in brain networks with nonparametric granger causality. *Neuroimage* 2008;41(2):354–62.
- [62] Rubino D, Robbins KA, Hatsopoulos NG. Propagating waves mediate information transfer in the motor cortex. *Nat Neurosci* 2006;9(12):1549–57.
- [63] So K, Koralek AC, Ganguly K, Gastpar MC, Carmena JM. Assessing functional connectivity of neural ensembles using directed information. *J Neural Eng* 2012;9(2):026004.
- [64] Tauste Campo A, Martínez-García M, Nacher V, Luna R, Romo R, Deco G. Task-driven intra- and interarea communications in primate cerebral cortex. *Proc Natl Acad Sci USA* 2015;4761–6.
- [65] Csizsár I, Talata Z. Context tree estimation for not necessarily finite memory processes, via bic and mdl. *IEEE Trans Inf Theory* 2006;52(3):1007–16.
- [66] Cai Z, Neveu CL, Baxter DA, Byrne JH, Aazhang B. Inferring neuronal network functional connectivity with directed information. *J Neurophysiol* 2017;118(2):1055–69.
- [67] James R, Barnett N, Crutchfield J. Information flows? A critique of transfer entropies. *Phys Rev Lett* 2016;116(23):238701.
- [68] Stokes PA, Purdon PL. A study of problems encountered in granger causality analysis from a neuroscience perspective. *Proc Natl Acad Sci* 2017;114(34):E7063–72.
- [69] Faes L, Stramaglia S, Marinazzo D. On the interpretability and computational reliability of frequency-domain granger causality. *F1000Research* 6.
- [70] Florin E, Gross J, Pfeifer J, Fink GR, Timmermann L. Reliability of multivariate causality measures for neural data. *J Neurosci Methods* 2011;198(2):344–58.
- [71] Barnett L, Barrett AB, Seth AK. Misunderstandings regarding the application of granger causality in neuroscience. *Proc Natl Acad Sci* 2018;201714497..
- [72] Sheikhattar A, Miran S, Liu J, Fritz JB, Shamma SA, Kanold PO, Babadi B. Extracting neuronal functional network dynamics via adaptive granger causality analysis. *Proc Natl Acad Sci* 2018;115(17):E3869–78.
- [73] Stokes PA, Purdon PL. Reply to barnett et al.: regarding interpretation of granger causality analyses. *Proc Natl Acad Sci* 115(29):2018:E6678–E6679..
- [74] Friston KJ, Harrison L, Penny W. Dynamic causal modelling. *Neuroimage* 2003;19(4):1273–302.
- [75] Friston K, Moran R, Seth AK. Analysing connectivity with granger causality and dynamic causal modelling. *Curr Opin Neurobiol* 2013;23(2):172–8.
- [76] Williams PL, Beer RD. Nonnegative decomposition of multivariate information, arXiv preprint arXiv:1004.2515..
- [77] Williams PL, Beer RD. Generalized measures of information transfer, arXiv preprint arXiv:1102.1507..
- [78] Schneidman E, Bialek W, Berry MJ. Synergy, redundancy, and independence in population codes. *J Neurosci* 2003;23(37):11539–53.
- [79] Nigam S, Poojoga S, Dragoi V. Synergistic coding of visual information in columnar networks. *Neuron* 2019;104(2):402–11.
- [80] DJ, Denman, RC, Reid, Synergistic population encoding and precise coordinated variability across interlaminar ensembles in the early visual system, *bioRxiv* (2019) 812859..
- [81] Shannon C. A mathematical theory of communication, *Bell Syst Tech J* 27;1948:379–423 and 623–656..
- [82] Venkatesh P, Dutta S, Grover P. Information flow in computational systems. *IEEE Trans Inf Theory* 2020;66(9):5456–91.
- [83] Pica G, Soltanipour M, Panzeri S. Using intersection information to map stimulus information transfer within neural networks. *BioSystems* 2019;185:104028.
- [84] Bím J, De Feo V, Chicharro D, Bieler M, Hanganu-Opatz IL, Brovelli A, Panzeri S. A non-negative measure of feature-specific information transfer between neural signals. *bioRxiv*; 2020: 758128..
- [85] Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364(6437)..
- [86] Williamson RC, Doiron B, Smith MA, Byron MY. Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction. *Curr Opin Neurobiol* 2019;55:40–7.
- [87] Cunningham JP, Byron MY. Dimensionality reduction for large-scale neural recordings. *Nat Neurosci* 2014;17(11):1500–9.
- [88] Williamson RC, Cowley BR, Litwin-Kumar A, Doiron B, Kohn A, Smith MA, Yu BM. Scaling properties of dimensionality reduction for neural populations and network models. *PLoS Comput Biol* 2016;12(12):e1005141.
- [89] Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi X-L, Romo R, Uchida N, Machens CK. Demixed principal component analysis of neural population data. *Elife* 2016;5:e10989.
- [90] Pandarinath C, O’Shea DJ, Collins J, Jozefowicz R, Stavisky SD, Kao JC, Trautmann EM, Kaufman MT, Ryu SI, Hochberg LR, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat Methods* 2018;15(10):805–15.
- [91] Williams AH, Kim TH, Wang F, Vyas S, Ryu SI, Shenoy KV, Schnitzer M, Kolda TG, Ganguli S. Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron* 2018;98(6):1099–115.
- [92] Mante V, Sussillo D, Shenoy KV, Newsome WT. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 2013;503(7474):78–84.
- [93] Murray JD, Bernacchia A, Roy NA, Constantinidis C, Romo R, Wang X-J. Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc Natl Acad Sci* 2017;114(2):394–9.
- [94] Kaufman MT, Churchland MM, Ryu SI, Shenoy KV. Cortical activity in the null space: permitting preparation without movement. *Nat Neurosci* 2014;17(3):440–8.
- [95] Li N, Daie K, Svoboda K, Druckmann S. Robust neuronal dynamics in premotor cortex during motor planning. *Nature* 2016;532(7600):459–64.
- [96] Semedo JD, Zandvakili A, Machens CK, Byron MY, Kohn A. Cortical areas interact through a communication subspace. *Neuron* 2019;102(1):249–59.
- [97] Friedman JH, Tukey JW. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans Comput* 1974;100(9):881–90.
- [98] Loperfido N. Skewness-based projection pursuit: a computational approach. *Comput Stat Data Anal* 2018;120:42–57.
- [99] Han S, Chu J-U, Park JW, Youn I. Linear feature projection-based real-time decoding of limb state from dorsal root ganglion recordings. *J Comput Neurosci* 2019;46(1):77–90.
- [100] Kantz H, Schreiber T. *Nonlinear time series analysis*, vol. 7. Cambridge University Press; 2004.
- [101] El Gamal A, Cover T. Multiple user information theory. *Proc IEEE* 1980;68(12):1466–83.
- [102] Ahlswede R, Cai N, Li S-Y, Yeung RW. Network information flow. *IEEE Trans Inf Theory* 2000;46(4):1204–16.
- [103] Brown EN, Kass RE, Mitra PP. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat Neurosci* 2004;7(5):456–61.
- [104] Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, Lee AK, Anastassiou CA, Andrei A, Aydın Ç, et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* 2017;551(7679):232–6.
- [105] Stosiek C, Garaschuk O, Holthoff K, Konnerth A. In vivo two-photon calcium imaging of neuronal networks. *Proc Natl Acad Sci* 2003;100(12):7319–24.
- [106] Stringer C, Pachitariu M, Steinmetz N, Carandini M, Harris KD. High-dimensional geometry of population responses in visual cortex. *Nature* 2019;571(7765):361–5.
- [107] Yaksi E, Friedrich RW. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved ca 2+ imaging. *Nat Methods* 2006;3(5):377–83.
- [108] Pachitariu M, Stringer C, Harris KD. Robustness of spike deconvolution for neuronal calcium imaging. *J Neurosci* 2018;38(37):7976–85.
- [109] Stetter O, Battaglia D, Soriano J, Geisel T. Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. *PLoS Comput Biol* 2012;8(8):e1002653.
- [110] Orlandi JG, Stetter O, Soriano J, Geisel T, Battaglia D. Transfer entropy reconstruction and labeling of neuronal connections from simulated calcium imaging. *PLoS One* 2014;9(6):e98842.
- [111] Runge J. Causal network reconstruction from time series: from theoretical assumptions to practical estimation. *Chaos* 2018;28(7):075310.
- [112] Nir Y, Fisch L, Mukamel R, Gelbard-Sagiv H, Arieli A, Fried I, Malach R. Coupling between neuronal firing rate, gamma lfp, and bold fMRI is related to interneuronal correlations. *Curr Biol* 2007;17(15):1275–85.
- [113] Fries P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 2005;9(10):474–80.
- [114] Lisman J, Jensen O. The theta-gamma neural code. *Neuron* 2013;77(6):1002–16.
- [115] Fries P. Rhythms for cognition: communication through coherence. *Neuron* 2015;88(1):220–35.
- [116] Palmigiano A, Geisel T, Wolf F, Battaglia D. Flexible information routing by transient synchrony. *Nat Neurosci* 2017;20(7):1014.
- [117] Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. *Nat Neurosci* 2011;14(7):811.
- [118] Stanley G. Reading and writing the neural code. *Nat Neurosci* 2013;16(3):259.
- [119] Panzeri S, Harvey CD, Piasini E, Latham PE, Fellin T. Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. *Neuron* 2017;93(3):491–507.
- [120] Shamir M. Emerging principles of population coding: in search for the neural code. *Curr Opin Neurobiol* 2014;25:140–8.
- [121] Avena-Koenigsberger A, Misic B, Sporns O. Communication dynamics in complex brain networks. *Nat Rev Neurosci* 2018;19(1):17.
- [122] Geiger D, Verma T, Pearl J. Identifying independence in Bayesian networks. *Networks* 1990;20(5):507–34.