Statistical analysis tools for assessing the functional relevance of higher-order correlations in massively parallel spike trains

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Mohammad Hasan Rostami,
M.Sc. in Physics,
aus Teheran, Iran

Berichter: Universitätsprofessorin Dr. rer. nat. Sonja Grün
Universitätsprofessor Dr. rer. nat. Bjorn Kampa

Tag der mündliche Prüfung: March 08, 2017

To my parents

Abstract

Understanding how the brain processes information is a major goal in neuroscience. A crucial step towards achieving this goal is to elucidate how information is processed by networks of neurons and how this is expressed in their firing activities. It has been hypothesized that neurons communicate through temporally precisely coordinated spiking, a long-standing hypothesis investigated by several studies. This view has been supported by the insight that a neuron would fire most efficiently if it receives synchronous (in a range of a few milliseconds) spike input. Experimental studies have endorsed this hypothesis and shed light on the functional relevance of synchrony in cortical information processing. However, much of this work has been carried out on handfuls of neurons, thus neglecting features of coordinated activity which emerge only at the level of statistically representative neuron populations while remaining invisible in data sampled from small populations.

Over the last few decades electrophysiological techniques have improved significantly, enabling nowadays to record from hundreds of neurons simultaneously. Such advances allow us to study the mechanism behind cortical information processing on a large scale by incorporating the interaction of the constituents of neuronal ensembles. This, however, calls for advanced analysis techniques which can handle increasingly larger numbers of recorded neurons. It is therefore very important to check whether the effectiveness of small-scale data analysis tools also extends to large-scale data.

In this thesis we begin with investigating two widely used statistical analysis tools in statistical neuroscience, namely the Unitary Event (UE) analysis and the Maximum Entropy (ME) model. These analysis tools have been designed to assess the presence, magnitude and neuronal composition of correlations in parallel spike train data, and the relevance of these correlations to cortical information processing. We address the emergent features and shortcomings of these analysis tools when applied to massively (i.e. up to hundreds of) parallel spike trains. Then, in line with the statistical frameworks employed by each of these analysis tools, we present two novel statistical analysis tools, i.e. Inhibited Maximum Entropy (IME) and Population Unitary Events (PUE), that can handle the high complexity of large scale data and are able to provide insights about the correlations within an ensemble of neurons. In particular, they aim at segregating the correlations which emerge as a direct consequence of network connectivity, e.g. shared inputs to pairs of neurons, and those with a functional role, e.g. task-dependent modulation of correlations in relation to behavior. In the application of our newly proposed methods to massively parallel spike train recordings from motor cortex of macaque monkey, we verify the insights that they can bring about the importance of correlations (pairwise and higher order) in cortical information processing.

To summarize, this work proposes two novel statistical tools to study the presence and functional relevance of correlations in massively parallel spike trains. In studying

the mechanism behind cortical information processing, these tools shed new light on the distinction between correlations that emerge as a consequence of network connectivity, and those that have a functional role.

Zusammenfassung

Eines der wichtigsten Ziele der Neurowissenschaften ist es, zu verstehen wie das Gehirn Informationen verarbeitet. Ein entscheidener Schritt hierfür ist herauszufinden, wie Informationen in neuronalen Netzwerken verarbeitet werden und inwiefern sich dieser Prozess in deren Aktivität widerspiegelt. Die Kommunikation von Neuronen beruht hauptsächlich auf dem Austausch von Aktionspotentialen. Eine wichtige und viel diskutierte Hypothese ist die Kommunikation von Neuronen durch zeitlich präzise aufeinander abgestimmte Aktionspotentiale. Diese Annahme wird von der Einsicht gestützt, dass ein Neuron insbesondere dann feuert (d. h. Aktionspotentiale produziert), wenn es viele synchrone Eingangssignale empfängt. Die Synchronizität bezieht sich hier auf ein zeitlich eng begrenztes Intervall weniger Millisekunden. Experimentelle Studien stützen diese Hypothese und geben Aufschluss über die funktionelle Relevanz von Synchronizität in kortikaler Informationsverarbeitung. Die meisten dieser Studien beruhen jedoch auf der Aufzeichung der Aktivität *weniger* Neurone und können somit Eigenschaften, die sich erst aus dem Zusammenspiel *vieler* Neurone ergeben, nicht erfassen. Koordinierte Aktivität wird deshalb erst unter Betrachtung einer statistisch signifikanten Anzahl von Neuronen sichtbar und bleibt in einer zu kleinen Stichprobengröße unerkannt.

Durch die technischen Fortschritte der letzen Jahrzehnte in elektrophysiologischer Messtechnik ist es heutzutage möglich, die Ativität von hunderten Neuronen simultan zu messen. Hierdurch können die Mechanismen kortikaler Informationsverarbeitung, unter Berücksichtigung der Interaktionen zwischen neuronalen Gruppen („assemblies"), in großem Maßstab untersucht werden. Dazu werden neue Analysemethoden benötigt, die mit der zunehmenden Anzahl von aufgenommenen Neuronen umgehen können. Ein wichtiger Schritt ist daher zu untersuchen, ob sich bisherige Methoden, die für eine kleinere Anzahl von Neuronen entwickelt wurden, auf größere Anzahlen von Neuronen anwenden lassen.

Diese Doktorarbeit beginnt deshalb mit der Analyse von zwei bekannten und etablierten statistischen Methoden, der „Unitary Events (UE) Analyse" und der „Maximum Entropy (ME) Analyse". Beide Methoden untersuchen die Existenz, Stärke und neuronale Zusammensetzung von Korrelationen in Datensätzen, die simultan aufgenommene Ketten von Aktionspotentialen (im folgenden „spike trains" genannt) vieler Neurone enthalten. Sie geben Aufschluss über die Relevanz dieser neuronalen Korrelationen für die kortikale Informationsverarbeitung. Wir untersuchen die Eigenschaften und Probleme dieser Methoden, welche sich ergeben, wenn sie auf bis zu hunderten simultan aufgenommenen spike trains angewendet werden.

Des Weiteren stellen wir zwei neuartige statistische Methoden vor, die speziell für die Analyse komplexer und umfangreicher Datensätze entwickelt worden: „Inhibited Maximum Entropy (IME)" und „Population Unitary Events (PUE)". Mit Hilfe dieser Methoden können aus den Korrelationen innerhalb von neuronalen Gruppen geschlossen werden, welche

im kortikalen Netzwerk allgegenwärtig sind. Insbesondere ermöglichen diese Methoden eine Unterscheidung der Korrelationen bezüglich ihrer Ursache. Einerseits können Korrelationen als direkte Konsequenz der Netzwerkkonnektivität auftreten, z. B. durch gemeinsam erhaltene Eingangssignale in Neuronpaaren, zum Anderen können sie funktionale Ursachen wie, z. B. eine aufgabenabhängige Modulation der neuronalen Korrelationen als Folge einer Verhaltensänderung haben. Abschließend wenden wir die neu entwickelten Methoden auf Daten aus Multielektrodenmessungen im Motorkortex von Makaken an. Damit verifizieren wir den praktischen Nutzen der Methoden, d. h. den erzielten Erkenntnisgewinn von neuronalen Korrelationen (sowohl paarweise als auch höherer Ordnung) und deren Bedeutung für kortikale Informationsverarbeitung.

Zusammengefasst stellt diese Doktorarbeit zwei neuartige statistische Methoden vor, die der Untersuchung von Korrelationen in massiv parallelen spike trains dienen. Die Methoden ermöglichen die Analyse von Mechanismen, die der kortikalen Informationsverarbeitung zu Grunde liegen. Sie geben Aufschluss über den Unterschied zwischen aus der Netzwerkstruktur resultierenden und funktionell bedingten Korrelationen.

Contents

I

Introduction

CHAPTER 1

Fundamentals and focus of study

This chapter introduces the biological and theoretical background for proposing the main aim and questions of this thesis. It focuses on the neuronal mechanism behind cortical information processing by highlighting the relevant hypotheses and reviewing the important studies which have been carried out during the last century in this respect.

1.1 Neuronal response and synapse

Thanks to Santiago Ramón y Cajal and his "Neuron doctrine", it is known nowadays that the building blocks of the *nervous system* are neurons: remarkable cells, among all types of cells in the body, with regard to their structures and functions. They, in a nutshell, consist of a cell body called *soma* which receives electrochemical signals from other neurons through ramified structures called *dendrites* and sends out a particular type of electric impulses known as *action potential* via *axons* to other neurons. The axon of a sending (pre-synaptic) cell is linked to the dendrite of a receiving (post-synaptic) cell via synapses.

The neuron interior is separated from the outer environment by a cell membrane which is normally impermeable to charged ions such as Cl^-, $K^+$and Na^+. These ions are present in the inside and the outside of the neuron, but usually at different concentrations. The overall higher concentration of positive charges in the exterior compared to the interior endows a neuron with electrical properties: a voltage difference across its membrane known as *membrane potential*. In absence of electrical inputs (*resting state*) to a neuron, this negative potential fluctuates around a baseline level called *resting potential* ($\sim -70\,\mathrm{mV}$, alhough this value may change across species and animals). To maintain this voltage, or the electrical gradient, neurons actively expell Na^+ and pump in K^+ by so called *ionic pump*s. Charged ions can pass through the cell membrane via protein structures called ion channels, which can open in specific conditions and are selective to a specific type of ions that can pass through, e.g. specific to Na^+ but not Cl^- or K^+. Two important classes or channels are *voltage-gated* channels, whose opening probability depends on the membrane potential itself, and *chemically-gated* channels, opened by chemicals called *neurotransmitters*. When a channel opens, ions of the associated type are free to diffuse against their concentration gradient (either outside the cell or inside it), generating an electrical current and eventually changing the membrane potential around the channel location. Channel opening and consequential changes in membrane potentials mediate the communication between neurons. When a pre-synaptic neuron A sends an input to a post-synaptic neuron B through a connecting synapse (Figure 1.1A), neurotransmitters are released at the pre-synaptic site (the terminal part of the axon of neuron A) and binds to chemically-gated channels at the postsynaptic site (dendrite of B). These channels are selective, in the most common scenario, to Na^+, which is then allowed from the outside to the inside of the cell (because of its chemical gradient), thereby causing an increase (*excitatory post-synaptic potential*, ePSP) in the local membrane potential (*depolarization*). PSPs usually have a magnitude of a few mV at most. If the increase in membrane potential reaches a sufficient level caused *firing potential* or *firing threshold*, for instance because neuron B receives sufficiently many inputs in a short amount of time, then voltage-gated channels on B open, allowing for a much more rapid influx of positively charged ions and

thus for a faster and stronger depolarization. This is the first phase of the generation of action potential, or *spike*.

The stereotypical shape of the action potential is shown in Figure 1.1B. The membrane potential has a rapid excursion from the resting potential toward higher values because of the opening of voltage-gated Na^+channels. This causes opening of more channels until the membrane reaches a maximum value and the Na^+ channels get inactivated. At this point the K^+ voltage-gated channels open and the efflux of K^+ restores the membrane potential. A remarkable property of the action potential, which should be noted here, is that its frequency characterizes the intensity and duration of cell stimulation rather than its amplitude. Therefore, it is said to be an "all-or-none" signal.

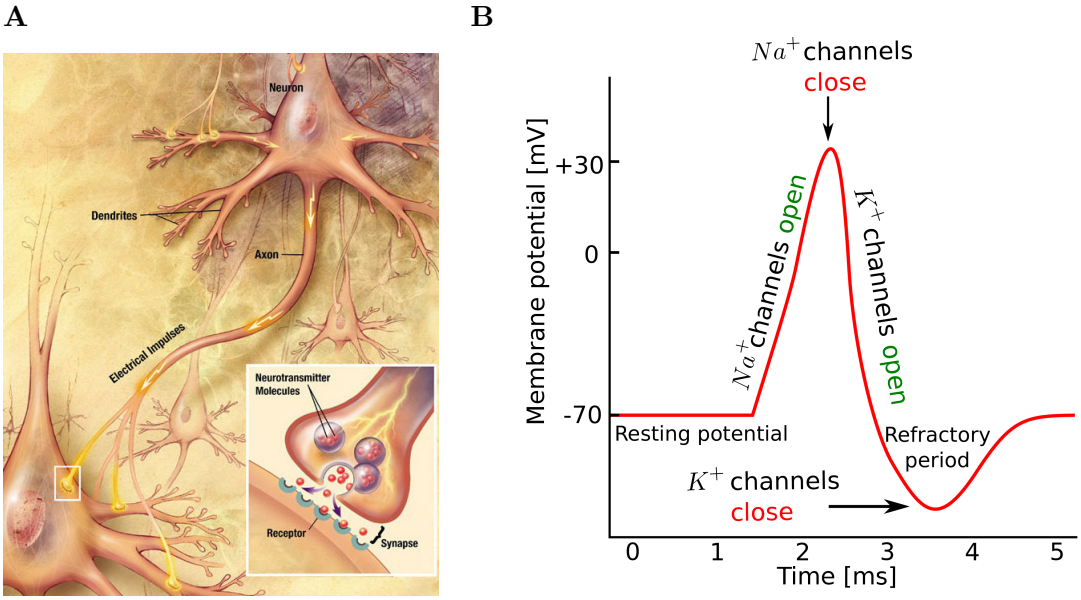

Figure 1.1: **Sketch of a biological neuron and an action potential. A)** Structure of a typical neuron and a chemical synapse (source: US National Institutes of Health, National Institute on Aging `http://www.nia.nih.gov/alzheimers/publication/alzheimers-disease-unraveling-mystery/preface`). **B)** Stereotypical shape of the action potential. Synaptic inputs to the neuron raise the membrane potential from the resting potential. This rise happens because of the opening of Na^+ channels and it continues until the potential reaches a maximum value. Then the Na^+ channels get inactivated. At this point the K^+ channels open and restore the membrane potential.

1.2 Theories of neuronal code

Neurons are commonly believed to be the basic units of information processing in the brain. There is a huge number of neurons in the brain, e.g., of the order of 10^{11} in the human

brain. These neurons communicate with one another to exchange information to cause feeling, acting, learning, remembering, etc. For instance a simple sensory–motor task like reaching and grasping an object needs the cooperation of many brain areas, from sensory neurons which locate the position of the object, till the motor neurons which execute the movement. Our brain manages such kind of complicated tasks in our everyday life within a few tens to hundreds of milliseconds, which brings up the question of how the information is (that efficiently) represented and processed in the brain by the activity of neuronal cells. This question boils down to understanding the relation between the neuronal responses and stimuli, i.e. how stimulus attributes (e.g. light, sound, smell, etc.) or motor actions (e.g. the direction of an arm movement) are represented by the response of neurons, in particular by the action potentials or spikes. Therefore, by characterizing the relation between neuronal response and stimuli we can try to understand how neurons represent information in terms of spikes, which is the central question in "neural coding". This very fundamental question remains unanswered, despite tremendous progress in understanding various cognitive processes.

In the following we will first introduce the notion of *local* and *distributed* neuronal coding scheme and the complex of *binding problem*. Then we will discuss the proposed solutions for that problem and the principal debate on this issue, which is known as *rate coding* versus *temporal coding* paradigm.

1.2.1 Local and distributed coding, binding problem

In this section, we first introduce the concept of *local* and *distributed coding* schemes by answering the following question: How do the responses of the individual or ensemble of neurons are used for representation of the external world in our brain?

Early studies on the stimulus selectivity of neurons (from sensory nerve fibers of frog muscles) by Adrian [6] and on the response properties of neurons by Hubel and Wiesel [168] were the basis of the *local coding* hypothesis, according to which neurons respond to the features (e.g. form, color, motion, depth, and position) of the perceived objects; elementary features are coded by neurons in the early sensory pathway and the more complex ones in the higher sensory areas. In the extreme case, this theory states that there is a single neuron which will be activated at the presence of one's grandmother, thus, often called *grandmother cell hypothesis* [19]. Let us take a concrete example to make this clear. Assume that you see a red Newton apple. You immediately perceive this apple which is "red" and "Newton" type. How do you do that? One possibility is to have one neuron specialized to detect the "red-Newton-apple". Therefore activation of this neuron results in the perception of the red Newton apple. Such a possibility is a simple example of the local coding hypothesis. Now considering different colors, e.g. 5, and different varieties, e.g. 5, of apples one would need 5×5 neurons to only be able to perceive different apples.

Thinking about the enormous number of possible features leads us to doubt this local coding hypothesis, simply because there are not enough neurons in the brain to cover all possible concepts one may come across. Apart from this *trivial* argument, the other criticism to this hypothesis is that coding by a single neuron is risky, i.e. the death of one neuron causes the death of one concept such as the concept of grandmother.

Alternatively, in the *distributed coding* hypothesis, different features are distributed in different streams like those of "what" and "where" streams in the visual cortex. This also supports the idea originally proposed by Donald Hebb according to which different feature dimensions are encoded in the co-activation of different groups of neurons called *cell assemblies* [157]. A cell assembly is an anatomically dispersed set of neurons with excitatory connections. The activity of the cell assembly can therefore be maintained by mutual excitation without continuous sensory stimulation (see [152] for a review). In this paradigm the same neurons may be used in different cell assemblies; hence there can be more groups of neurons than there are single neurons. Coming back to the red Newton's apple example, instead of 5×5 neurons we can code each color or type of apples with only 5 neurons: the co-activation of sub-groups of these neurons represents e.g. the red Newton apple or the green Newton apple, etc. Thus this is much more efficient, and prevents the problem of combinatorial explosion of the grandmother cell theory.

The distributed representation is appealing because it provides a powerful paradigm for explaining the mechanism behind perception. However, the question then arises: how can superposition of activation of assemblies be segregated into individual assemblies? And, at a more global level, how can local features in different cortical areas form a unified whole to represent coherent percepts and actions? The best-known example is shown in Figure 1.2. This picture contains black patches which, by looking at them for the first time, seem completely disordered. But after a while our brain perceives an order in this disorder: the Dalmatian dog in the middle of the picture. The question of how this is done by our brain is known as the *binding problem* or *Gestalt psychology.* Another example of the binding problem can be seen in Figure 1.3. Here the picture has more than one interpretation. It can be interpreted either as a rabbit or a duck, and our interpretation alternates between the rabbit and the duck but we never see both of them at the same time.

1.2.2 Temporal coding and rate coding

In the previous section we introduced the binding problem. Here we discuss two rather opponent solutions which were proposed to this problem: *rate coding* and *temporal coding.*

Milner [246] and von der Malsburg [350–352], independently from each other [353], suggested the idea of *binding-by-synchrony* (also called *temporal correlation hypothesis*) as a possible solution to the binding problem. Binding-by-synchrony is based on the cell assembly hypothesis postulated by Hebb [156] with time as an additional dimension

Figure 1.2: **Illustration of the binding problem or Gestalt psychology**. There is a Dalmatian dog hidden in the black-and-white spots. Adapted from [341].

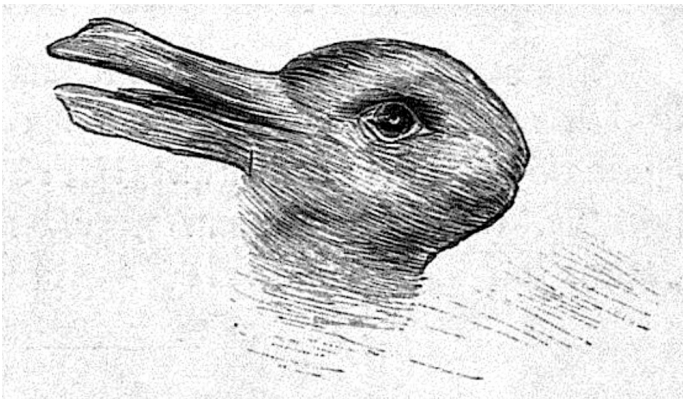

Figure 1.3: **Rabbit and Duck**. Adapted from the 23 October 1892 issue of Fliegende Blätter ("Rabbit and Duck") `http://digi.ub.uni-heidelberg.de/diglit/fb97/0147&ui_lang=eng`.

to represent dynamical relations (sometime called *dynamical cell assembly hypothesis*). The basic idea is that a group of neurons forms a cell assembly by synchronizing their spiking activity within a few millisecond-time scale. The synchronization is dynamically modulated and allows a particular cell to participate in different cell assemblies. The binding-by-synchrony was first supported in the experimental works of Gray and Singer [127, 128, 313].

Considering the biological nature of neurons, the temporal coding hypothesis postulates that neurons operate as coincidence detectors [1, 2]. This implies that the postsynaptic neurons are sensitive to the precise arrival timing of the presynaptic inputs (on a millisecond time scale). This concept in line with the dynamical cell assembly hypothesis brought a vigorous framework to shed light on the mechanism of cortical information processing.

Another somewhat opposite view has been proposed in the rate coding paradigm, where a neuron is seen as an integrator. This paradigm states that most, if not all, incoming presynaptic inputs contribute to the generation of action potentials on the postsyanptic side. In fact the neuron integrates the inputs from the presynaptic side over a temporal window of tens or hundreds of milliseconds and this can lead to the generation of an action potential if there are enough input spikes. Thus, a neuron operates as an integrator and the precise timing of afferent signals is not important. This implies that the firing rate of a neuron conveys information, and the correlations between the activity of neurons have no functional meaning and might only emerge as an epiphenomenon of the network structure [222, 302].

The two rather opposing paradigms of rate coding and temporal coding have been widely studied during the last few decades. Experimental evidence has been provided for both of these hypotheses, and the pros and cons of both have been argued about in theoretical studies [18, 44, 303, 313, 323]. Taking all these studies together, it seems that the brain might use both rate and temporal codes as complementary contributions.

1.3 Correlations & behavior

The question of how information is represented and processed by the brain is still a matter of an ongoing debate (see previous section). Nevertheless, correlated activity between pairs of cells has been observed early on [265] and immediately the question arose whether there is a functional interpretation of this observation [104], or correlated activity is merely an unavoidable epiphenomenon, because neurons are connected and therefore influence one another [303]. In this section we will review briefly the major hypotheses that have been studied during the last two decades trying to shed light on the role of correlation in cortical information processing.

Theoretical considerations hypothesized that synchronous activation of neurons may be used to bind representations of features into more complex percepts [29, 127, 313,

352]. Another hypothesis is that correlations may be used to efficiently gate information [296]. These functional interpretations can also be approached in a bottom up, dynamical view, by asking the question: What makes a neuron fire? Cortical neurons typically show two tendencies [167]: they integrate, but they are also efficient in detecting synchrony; from which the terms "integrator" and "coincidence detector" [2, 202] have been coined (see Section 1.2.2). Moreover, these two dimensions of information processing are not completely orthogonal, because correlation transmission may be co-modulated with changes of firing rate, and in theory [60, 307] in the regime of Gaussian fluctuating membrane potentials. Synchronous synaptic events, however, have been hypothesized to explain highly non-Gaussian features of membrane potential fluctuations in sensory cortex [66]. These can cause a considerable gain of correlation transmission, supporting synchrony as a robust and noise-suppressing channel of signal transfer [299].

In another view, by studying chaos in neuronal networks, it has been suggested that the chaotic activity arising in recurrent neuronal networks [325, 347] may be employed as an argument for correlations being an epiphenomenon: Estimating that a single spike causes a barrage of downstream spikes shows that cortex may operate deeply in the chaotic regime, so that a single spike can hardly be used to process information in a reliable way [222]. In line with this argument is the view of weakly correlated activity of the cortex as a rich set of basis vectors from which arbitrary linear readouts can be constructed [343]. Correlations between spike trains impair this basis set by introducing redundancy. The signal-to-noise ratio of such a signal, in the simplest way of the population-averaged activity, is therefore bounded by correlations [303, 364]. The proposed linear readout, inspired by the accumulation of synaptic input by a single neuron, has, however, to be contrasted with non-linear readout mechanisms, which indeed can assess information represented in correlations [324]. A more differential investigation of these ideas has shown that correlations are not necessarily information-limiting, but that limitations arise only from a specific subset of correlations that relate to the tuning properties of the neurons [248].

However, the above-mentioned hypotheses cannot to date explain a salient experimental finding of the relevance of transient, short-lasting correlations (in a range of few milliseconds), called *synchrony.* Such synchronous events have been observed in motor cortex [196, 278, 340] at time points of expected, task-relevant information, but also in primary visual cortex in relation to saccades (eye movements) [170, 233].

1.3.1 Forms of correlations: signal correlation, noise correlation, synchrony

In the previous sections we have discussed different views on the importance of correlation in cortical information processing. However, the term "correlation" in the neuroscientific literature has been referred to distinct phenomena, causing sometimes confusion about

what it is meant by "correlation". Here we explain three commonly used measures of pairwise correlations with their basic neuroscientific standpoint in the field of neural coding.

The experimental paradigm of repeatedly presenting a stimulus and measuring the average response, and the assumption that information is entirely contained in the average firing rate, suggests the characterization of neurons by their tuning curves [168]. As a consequence, the terms "signal correlation" and "noise correlation" have been established [73]: In a pair of neurons, signal correlation measures the similarity of their tuning curves, while noise correlation (sometimes called spike count correlation) quantifies the co-variability of their responses, given the same stimulus. Both of these measures are typically quantified by the Pearson correlation coefficient and have been based on responses measured over the duration of typically hundreds of milliseconds.

Another form of correlation is "synchrony", which expresses the temporal relationship between neurons on a short time scale of few milliseconds. A very first technique, proposed by [265], to measure the synchrony (also delayed synchrony) between two spike trains is the *cross-correlogram (CCG)*. The CCG measures the tendency of two spike trains to jointly emit spikes with zero lag, used for quantifying synchrony, or by a certain time delay, used for quantifying delayed synchrony (see [141] Chapter 17 for an extensive explanation).

1.4 Principled statistical models: towards higher order correlations (HOC)

The correlation measures mentioned above, i.e. the Pearson correlation coefficient for measuring noise or signal correlation and the CCG, are simple, descriptive statistics for studying the correlations between a pair of neurons. Their simplicity has made them a convenient choice to investigate the role of correlations in cortical information processing. However, such pairwise correlations provide a limited view of the collective dynamics of neuronal populations [33, 141, 206, 236]. To clarify what we mean by limited view, let us give a concrete example, adapted from [141]. Consider the two scenarios sketched in Figure 1.4A and Figure 1.4B. In each scenario, the spiking activity of 3 neurons, $s_1(t)$, $s_2(t)$ and $s_3(t)$, are shown together with the input to each neuron. The firing rate of each neuron and the pairwise correlations for all combinations of pairs are similar in both scenarios. However, in Figure 1.4A, all the neurons receive one shared common input (y_{123}) whereas in Figure 1.4B for each pair of neurons there is a different shared input (y_{12}, y_{13}, y_{23}). Assuming strong synapses in both scenarios, this clear difference appears only by measuring the triplet synchrony or third order correlation between spike trains. Now, considering a large network of neurons with a structured (not random) connectivity, these scenarios might be more complicated, and only statistical measures which go beyond pairwise correlations can provide a complete picture. Thus, to avoid the limited view sketched above in studying

the functional role of pairwise correlations in cortical information processing, one needs to carefully investigate higher order correlations between neurons.

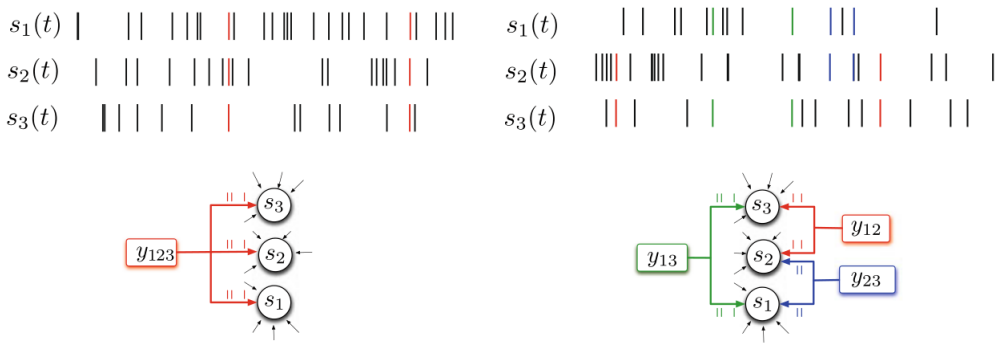

Figure 1.4: Illustrative example of the importance of higher-order correlation (see text for details; adapted from Figure 12.1 of [141]).

1.4.1 Signature of assemblies: time-locked synchrony to behavior

Precise spike coordination within the spiking activity of multiple neurons, e.g. synchrony, is considered to be a signature of cell assemblies [1, 103, 152, 314]. This argument rests on the premise that synchronized presynaptic spikes enhance reliably the generation of output spikes [2, 70, 201, 299]. Therefore, the detection of cell assemblies and their functional relevance can be focused on the detection of time-locked synchrony in relation to behavioral context [94, 278, 295]. Statistically, this needs principled statistical models to study synchrony of different orders among spike trains. These statistical models should be able to quantify the synchrony among spike trains and relate it to a behavioral task (or stimulus). This, however, is harder than it sounds, because to demonstrate that information coding is solely carried out by synchronous spikes — or spike patterns – one needs to subtract the effect of firing rates. This is a crucial step to ensure that the increase in the rate of synchronous spikes is not caused by the increase in the firing rates and, therefore, to justify the conclusion that the brain has to pay attention to correlations to extract extra information that is carried out by synchrony.

A prominent method which was specifically developed to quantify the synchrony between spike trains is the Unitary Events method [135, 136, 138]. This method uses a statistical approach, by testing the null hypothesis that neurons are mutually independent (see Chapter 2). A notable feature of this method is its capability of a time resolved analysis, thereby enabling to relate the temporal modulation of synchrony to sensory stimuli and behavioral events.

1.4.2 Synchrony and information-theoretic perspective

Another approach to study the role of correlations in cortical information processing comes from an information theoretic perspective. In this context, the synchronous spikes act as an extra channel of information beside the information carried by the firing rates. An explicit assumption here is that if there is information in spike trains, e.g. originated by synchrony, the brain uses it. Therefore, if we measure the amount of information by considering different aspects of spike trains, e.g. marginal statistics, pairwise or higher order correlations, we will be able to relate this to the stimulus or behavior. In the other words, if the brain uses correlations to accomplish a task, we have to be able to understand the underlying coding scheme by characterizing the relation between information in synchrony among neuronal responses and stimuli (or behavioral context). In this section, we first elaborate briefly on a big picture of this approach which is based on a probabilistic view of the relation between stimuli and neuronal responses using Bayesian theory [182]. Then we will explain a common quantitative method used in this framework for studying the relevance of correlations.

Understanding the relation between neuronal response on one side, and what we could call the "state" of the brain – the complex combination of behavior, stimuli, memory, and thought, which still partly escapes definition and measurement [12] – on the other side, is a major goal in the study of cortical information processing [192, 216]. We would like to achieve this understanding at a probabilistic level at least. In very hand-waving terms, we could say it amounts to assigning probabilities of the form

$$P(\text{brain activity}|\,\text{state}). \tag{1.1}$$

Given the practically uncountable patterns of activity in the brain or even in just a small region of it, and the continuous spectrum and even vagueness of "states", assigning such probabilities is practically impossible and will likely stay that way for the next few decades. In Bayesian theory we deal with a vague or too large set of probabilities by introducing one or more statistical models, which simplify the problem and give it well-defined contours. For example, we can introduce a set of models M , each of which includes some multi-dimensional parameter $\boldsymbol{\alpha}$, in such a way that they informationally *screen off* every "brain activity" from every "state", making them conditionally independent [57, 262, 263, 326]:

$$P(\text{activity}|\,\boldsymbol{\alpha}, M, \text{state}) = P(\text{activity}|\,\boldsymbol{\alpha}, M) \qquad \text{for all activity and states.} \tag{1.2}$$

Then the inverse also holds: $P(\text{state}|\,\boldsymbol{\alpha}, M, \text{activity}) = P(\text{state}|\,\boldsymbol{\alpha}, M)$. By the rules of marginalization and total probability we can then rewrite the probability $P(\text{activity}|\,\text{state})$

as

$$P(\text{activity}|\,\text{state}) = \sum_M \int P(\text{activity}|\,\boldsymbol{\alpha}, M)\, P(\boldsymbol{\alpha}, M|\,\text{state})\,\mathrm{d}\boldsymbol{\alpha}. \tag{1.3}$$

The advantage of this approach appears in the mutual information conditional on the model M and $\boldsymbol{\alpha}$:

$$I(\text{activity}, \text{stimuli}|\,\boldsymbol{\alpha}, M) = 0,$$

or, paraphrasing Caves [47]: "the mutual information between state and brain activity flows through the model M and parameters $\boldsymbol{\alpha}$". In this *divide et impera* approach we deal more easily with $P(\text{activity}|\,\boldsymbol{\alpha}, M)$ and $P(\boldsymbol{\alpha}, M|\,\text{stimuli})$ separately than with the full probability (Eq. 1.1), provided the parameter $\boldsymbol{\alpha}$ has much fewer dimensions than the "activity" and "state" spaces. This parameter then constitutes a coarser but sufficient description of the activity, or of the state (stimuli, behaviour, memory, thought processes), or of both. An example of the first case could be the mean activities and pairwise correlations of a neuronal population; an example of the second could be the orientation of a light-intensity gradient on the retina, or ambient temperature. In the first case, if the model M can be interpreted and motivated neurobiologically, then it is a "neural code" [34, 72, 86, 259, 260, 275].

The abstract viewpoint just outlined [36, 58, 59, 92, 217, 262, 263] can be made concrete by the maximum-entropy method: the model M is chosen to be a maximum-entropy model [51–53, 101, 142, 162, 165, 174–176, 180–182, 186, 243, 304, 305, 315, 317, 342, and references therein], and the parameter $\boldsymbol{\alpha}$ to be the empirical means and the pairwise empirical correlations of the activity of a neuronal population. Does such a choice of model and parameters give reasonable predictions $P(\text{activity}|\,\boldsymbol{\alpha}, M)$? This question has been asked repeatedly in the neuroscientific literature of the past few years. Some studies [33, 49, 126, 297, 338] have tested the suitability of the maximum-entropy distribution – (Eq. 1.2) from our perspective – for various experimental and simulated "activities" and "states". Some studies [21, 96, 230, 235, 310, 312] have tested the suitability of pairwise correlations or of higher-order correlations – our parameter $\boldsymbol{\alpha}$. Some studies have done both at the same time [105, 227, 228, 293]. This will be the topic of Chapter 3 of this thesis where we explain the maximum entropy model in detail and examine its aptness for studying the relevance of correlations in massively parallel spike trains.

1.5 Advances in experimental recordings: needs for advanced statistical tools

Over the last five decades, the number of simultaneously recorded neurons has grown exponentially. To a good approximation, this number has doubled every 7 years [329] since the development of the first implantable multi-electrode array in the 1950s. Nowadays, microelectrode arrays, e.g. the Utah array (Blackrock Microsystems, Salt Lake City, UT,

USA), enable to chronically record extracellular activity from hundreds of individual neurons in both anesthetized and behaving animals [25, 42, 254, 279]. Consequently, these advances in the neuronal recording allow, for the first time, to test the relevant hypothesis behind neuronal coding in a large population of neurons and, therefore, to gain a rigorous understanding of information processing in the brain. However, this requires data analysis techniques which can handle large numbers of simultaneously recorded neurons. These analysis techniques must be developed taking explicitly into account the relevant questions that need to be studied. For example, in studying the cell assembly hypothesis (discussed in the previous sections), which is the most experimentally and theoretically studied hypothesis in cortical information processing and is the focus of this thesis, we would require analysis tools which enable to consider emergent features when studying large populations of neurons. One of these features that has been extensively studied in theoretical works is the correlation between ensembles of neurons that arises from direct or recurrent synaptic connections as well as shared inputs, even in the simplest network models without any temporally modulated input [158, 159, 204, 266, 276, 303, 333, 344]. This feature becomes accessible when analyzing large population of cortical neurons and, consequently, violates the assumption of independence between neurons which is commonly presumed in classical analysis tools.

To illustrate the last point, consider the real case scenario of recording from cortical neurons. Using, for example, the Utah array we are able to record from hundreds of neurons. The recorded spikes could solely be the result of the *ongoing* activity [12], e.g. when there are no stimuli or behavioral tasks; or they could reflect the composition of ongoing and task-related activity (see Figure 1.5). From theoretical studies [159, 204, 303, 332, 333] we can hypothesize that during ongoing activity the recorded spike trains show correlations merely because of the connectivity between them, which we call *baseline correlation*. Whereas when these ensemble of neurons are engaged in a specific task there will be extra correlations, such as those hypothesized by Hebb, which are synchronous action potentials appearing at behaviorally relevant time points, which we call *functional correlation*. Taken all together, designing statistical tools with a null hypothesis which includes the baseline correlation that is expected in the absence of any task-related stimulus or behavior is essential.

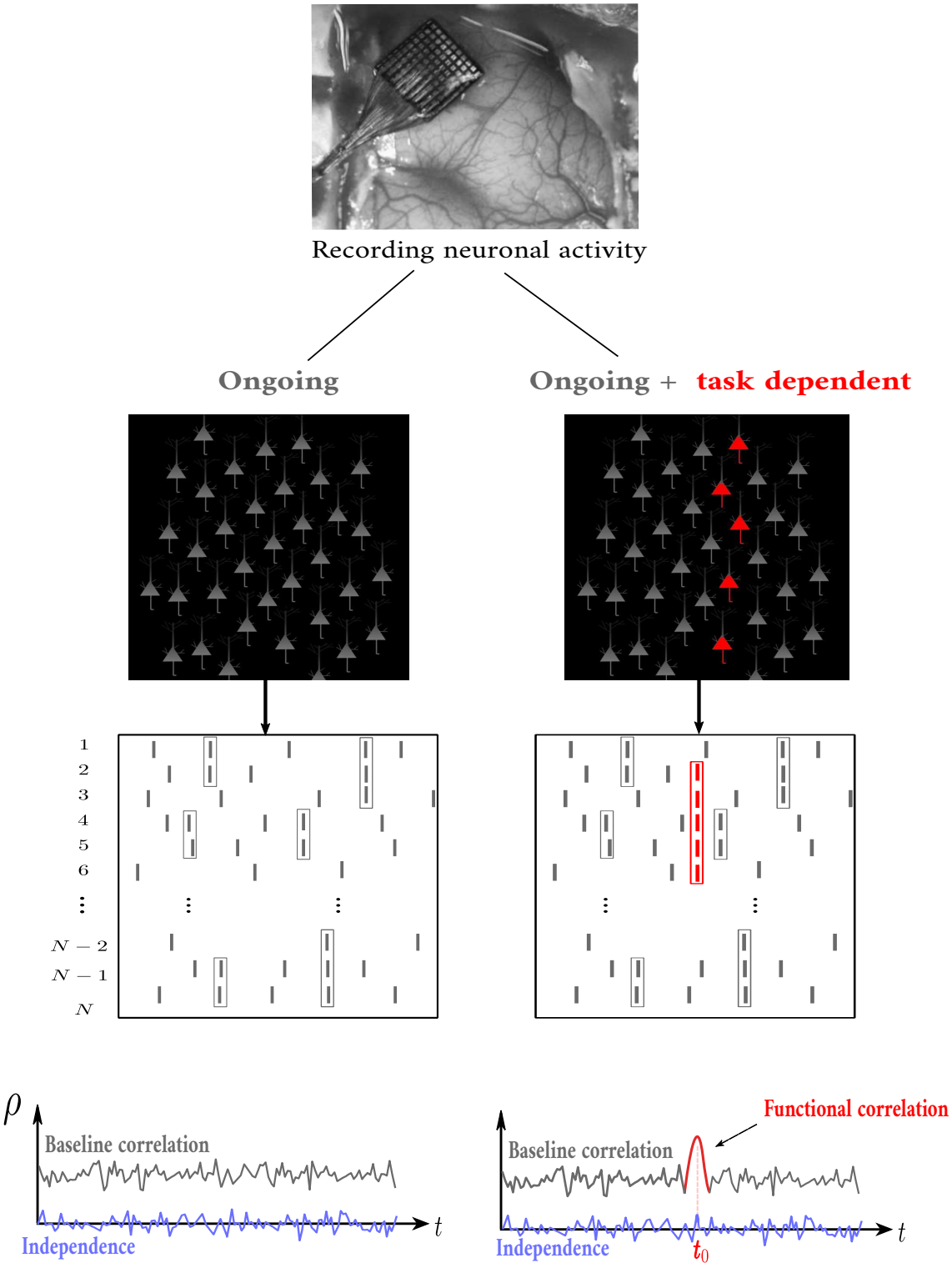

Figure 1.5: *(Caption on the next page)*

Figure 1.5 *(previous page)*: **Baseline and functional correlations in activity of an ensemble of neurons.** Using multi-electrode array, e.g., Utah array (Blackrock Microsystems, Salt Lake City, UT, USA), the activity of hundreds of neurons can be recorded simultaneously. Top panel shows a Utah array implanted chronically in the motor cortex of monkey (adapted from [279]). Following the cell assembly picture, the recorded activities of neurons are either the results of so called "ongoing" activity [12] (gray color; left branch), i.e. in the absence of stimulus or behavioral task, or the results of the ongoing activity plus the task dependent activity where the activity of a portion of the neurons (assembly) is relevant for the task (red color; right branch). The ongoing activity of neurons exhibits correlation due to the connectivity (marked spikes in gray squares). This correlation is beyond chance correlation (coincidence) that we might observe given the firing rates of neurons. A time-resolved statistical measure ρ, which takes into account the chance correlation by assuming independence among neurons, fluctuates around zero when measured on the activity of (unconnected) independent neurons (bottom panel; blue color), but measuring ρ on an ensemble of connected neurons exhibits a positive offset, what we name baseline correlation, which is not accounted for in the statistical measure ρ (bottom panel; gray color). Now, considering the same recordings while the animal is engaged in a behavioral task or represented by a stimulus (right branch), extra correlations will be appeared because of the activation of the cell assemblies related to the stimulus or task, what we name functional correlation. The functional correlation (appeared in the sketch at t_0) is, therefore, time-locked to the behavioral task and can be detected in our statistical measure as a significance increase on top of the baseline correlation (bottom panel; red color).

Furthermore, these data analysis techniques will be along with computational challenges. These challenges range from handling the big amount of data to the combinatorial explosion of the complexity of the data from large numbers of neurons and the corresponding increase in the number of degrees of freedom. Therefore, it is necessary to improve the efficiency of data analysis algorithms and to develop hardware to facilitate them. These techniques should also be made publicly available, to assist the analysis of huge amounts of data which are being produced every day in different laboratories around the world.

1.6 Aim and overview of the thesis

The aim of this thesis is to provide statistical tools for analyzing (pairwise or higher order) synchrony among massively parallel spike trains (MPST), available nowadays by modern electrophysiological techniques. These analyses can contribute to our understanding of how cortical information processing is carried out in the brain. In particular they can test the cell assembly hypothesis in large population of neurons by considering the complexity of high-dimensional data. To this aim, we first investigate the aptness of the existing

commonly used methods in spike train analysis, namely the Unitary events (UE) analysis [134–136, 139] and the Maximum Entropy (ME) model [21, 96, 230, 235, 310, 312], to be applied to MPST. This investigation will carefully consider the severe computational or statistical problems that might arise in high dimensional data. After identifying the emergent features as well as the limitations and shortcomings of these analysis tools, we will propose two novel statistical analysis tools which can substitute the UE and ME methods in the search for the functional relevance of correlations in MPST.

The rest of this thesis is structured as follows: In Part II the results of our studies are presented. In Chapter 2, the UE method is inspected by replicating the results of [278] to ensure that the original research is reproducible, and to provide an open source implementation of this method in Python. Then, using this python implementation, the UE method will be applied to a recording from a monkey performing a center-out-reaching task to check the systematic occurrences of UEs across ensemble of neuron pairs. These analyses demonstrate that the UE method applied to a large population of cortical neuron pairs allows to uncover a baseline correlation which is known through theoretical studies to be a direct consequence of shared input in network of neurons. Considering the ubiquity of baseline correlations in the brain, we will discuss the importance of providing analysis tools which enable to separate the baseline correlation from the correlations that have a functional role.

In Chapter 3, the ME method will be inspected by investigating its aptness for MPST. By employing large populations of neurons (159 neurons) recorded from motor cortex of a macaque monkey during "ongoing activity", i.e. with the monkey sitting on a chair without performing any task, we carefully investigate the results of the pairwise ME method and their interpretation. We identify a failure of this method in applications to MPST which has not been noted in previous literature. To correct this failure and its ensuing problems, we, then, introduce a new method, named Inhibited Maximum Entropy (IME). IME is a parametric-model based method similar to ME but based on a biologically realistic model that accounts for features of the experimental data which have been neglected in the pairwise ME method.

In Chapter 4, we present a new method, the Population Unitary Event (PUE) analysis, which is specifically designed to assess higher-order (i.e., beyond pairwise) synchrony in MPST by accounting for the baseline correlation. This method is calibrated on test data which imitate realistic scenarios of experimental data, and we demonstrate that the method copes with the variability of neuronal responses over time and across trials and neurons, variability which are usually observed in physiological data [140]. We also show that the method has a high degree of specificity and sensitivity, as well as a high tolerance against relatively small sample sizes, which allows its application in a time resolved manner to detect time dependent changes of the synchrony, and infer their order. Furthermore, we

demonstrate the application of this method to MPST recorded by a 10×10 electrode Utah array from motor cortex of awake monkey performing a delayed reach-to-grasp task.

Finally we discuss the findings of this thesis in Part III by summarizing our results and outlining possible future work based on our studies.

II

Results

CHAPTER 2

Baseline and functional correlations

The first part of the following chapter has been published as the following manuscript: **Rostami V**, Ito J, Denker M, and Grün S (2016) [Re] Spike Synchronization and Rate Modulation Differentially Involved in Motor Cortical Function, ReScience, volume 3, issue 1, article number 3 2017, DOI 10.5281/zenodo.583814. The second part of this chapter is an unpublished work in collaborations with Junji Ito, Joachim Confais, Alexa Riehle, Moritz Helias and Sonja Grün.

Author contributions: Under supervision of Junji Ito and Sonja Grün, the author performed the method implementation, the replication of the results of [278], all the analysis, simulations and interpretation of the results.

2.1 Introduction

To test whether temporal coordination of spiking activity is indeed relevant for neuronal information processing, advanced data analysis methods are required that perform correlation analysis between simultaneously recorded single unit spike trains. The Unitary Events (UE) analysis [134–137] is one of those methods which is able to extract significant spike synchrony between neuronal activities that is beyond what is expected by chance (given the firing rates of neurons) and enables us to follow the dynamics of this coordination. The tricky issue with such analyses is that one has to take into account that experimental spike data are typically non-stationary in the sense that the firing rates are not stationary in time and not homogeneous across trials, and that the statistics of the individual spike trains deviate from those of a Poisson process [140, 141]. If such features are ignored there is a considerable danger of occurrence of false positives and thus wrong interpretation of the data. Changes in the firing rates are the most prominent generators of false positives if ignored. The original UE method (used in [278]) considers this aspect by performing the analysis in a sliding window fashion. In later versions of the analysis method also other features were corrected for by considering the respective features in the null-hypothesis of the tests [139, 223, 233, 269], either by using extended analytical descriptions of the null-hypotheses or by use of surrogate methods [140, 141].

The UE method had been applied to experimental parallel spike data from, e.g., the motor cortex of awake behaving non-human primates [64, 65, 122, 139, 196, 278] and to data from visual cortices [170, 233]. Generally it was found that UEs, i.e., synchronous spike events across neurons that are in excess of the expectation, occur in relation to behavior, e.g., when the animal expects a signal to occur but did not [278]. This finding, in particular, demonstrates the core feature of the method: by performing a time resolved analysis, the method accounts for changes in the firing rates and captures modulations of significant spike synchrony in time. It was also shown that the time of occurrence of UEs may change to a new requested timing in the behavior during a learning process [196].

In this chapter we first implement the UE method in Python and replicate the results of [278]. A publicly available and open source implementation of the UE method had not been available. Only very recently, we reimplemented the UE analysis as part of the Electrophysiology Analysis Toolbox[1] (Elephant), a Python library that provides implementations for the analysis of electrophysiological data. A reimplementation of the method, rather than releasing the existing code, was justified by two factors. First, the custom data object model used to represent the primary data and metadata in the original analysis code was not documented. Therefore, any data represented in a specific file format had to be converted by implementing a custom data loading routine for this data model. Our new implementation of the UE method is part of the Elephant and

[1] `http://neuralensemble.org/elephant/`

Neo[2] libraries. These provide generic functions and classes for standardized neuronal data representation and basic analyses. In particular, the new UE reimplementation is based on the internal data object model provided by the Neo library [97], a package for representing electrophysiology data in Python. Neo provides support for reading a wide range of neurophysiology proprietary file formats, and supports writing to a subset of these formats, including non-proprietary formats such as HDF5. Second, the original implementation has experienced several updates throughout the development process and after its publication in order to include improvements and extensions of the method to accommodate more features of experimental data. Since no version control was employed in this development process, the original code used in Riehle et al. (1997) [278] was lost at some point. Considering that no systematic testing was performed each time the code base of the Unitary Event analysis was updated after the original publication to check whether the code gives the same results as before, it was not clear whether the latest version could exactly reproduce the results of [278].

Then, using our Python reimplementation, we apply the UE method on a larger population of neuron pairs than the number of neurons used in [278]. For doing that we employ a meta statistic across neuron pairs which was suggested in [196] that allows to study the systematic occurrences of UEs across large neuronal sampling. As a result we demonstrate that this method reveals correlations between pairs of neurons which are known, in theoretical studies [158, 159, 204, 266, 276, 303, 333, 344], to be present between connected cells, because of their connectivity. These correlations are interpreted as baseline correlations since they occur owing to the structure of the network, independent of functions in the network.

2.2 Replication of Riehle et al (1997) by an open source implementation of the Unitary Events method

In this section we illustrate the successful reproduction of the results shown in [278] using our new Python implementation of the UE method. In particular, we reproduce Figure 2 and Figure 4A of the original paper, which represent the central results of the original study. The remainder of the original paper consisted of more example analyses of individual data sets and a meta statistics across many data sets, all of which were based on the same analysis method and thus do not provide additional insight when reproduced. In the original publication the authors used two different UE implementations: one implemented by Sonja Grün (SG) in IDL [3], and the other in Matlab (Mathworks, Nattick, MA) implemented later by Markus Diesmann (MD) and SG. The IDL implementation is not available anymore. At a later point we were provided with a Matlab implementation of the extended UE

[2] `http://neuralensemble.org/neo/`

[3] `http://www.harrisgeospatial.com/ProductsandSolutions/GeospatialProducts/IDL.aspx`

method, however it does not preserve the original implementation used in [278] and was not considered in this study.

For the reproduction of the original results we contacted and communicated with Alexa Riehle (AR), CNRS-AMU, Marseille, and MD, Research Centre Jülich. AR is the first author of the original publication and performed the experiments and data analysis. MD is the third author of the original publication and contributed with the Matlab implementation and quality checks of the software implementations. The reproduction of the results of [278] would not have been possible without contacting these authors, since the information in the original publication is not sufficient for reproducing the results. AR provided us with the original data and information, including an old written report by MD, which was crucial to reproduce Figure 4A.

2.2.1 UE analysis and Python implementation

For reproducing the results of [278] we use our reimplementation of the Unitary Events analysis method in Python which is made available in the *unitary_event_analysis* module of the Elephant library (accepted pull-request: [287]). The method is accompanied by unit tests for individual functions (test coverage: 88.54%) and documented as part of the library documentation. The structure and algorithm of our UE implementation is explained in the pseudo-code shown below.

In the following, we will explain the algorithm in detail. The primary data entering the UE method is a set of parallel, i.e., simultaneously recorded, spike trains, recorded in one or multiple trials. In Elephant, an individual spike train of a particular neuron in a particular trial data is represented as a `Spiketrain` object in the data object model provided by the Neo library, which stores the time points of spike occurrences along with additional information describing the spike train, such as the start and end time. The UE method consumes a nested list of `Spiketrain` objects relating to the spike trains of individual neurons in the individual trials. In order to perform the necessary calculations, the spike data must be converted to an alternative time-binned representation (line 1 in the pseudo code) where the parallel spike trains are stored as binary sequences of ones (marking time bins containing at least 1 spike) and zeros (bins with no spike). The bin size is a parameter provided as input (parameter *bin_size* in the pseudo code) to the analysis, and defines the temporal precision in detecting spike synchrony. A spike pattern is defined as a specific vector of zeros and ones in one time bin of a given trial across all neurons. Thus, the number of possible spike patterns is given by $2^N - N - 1$, where N is the number of neurons. In order to limit the analysis to a set of patterns of interest, the input parameter *pattern_hash_values* specifies the patterns to consider in the analysis in the form of a hash value which uniquely represents each spike pattern (line 2 in the pseudo code). The hash value is obtained by interpreting the binary spike pattern as a binary number, where the n-th neuron is represented as bit $n - 1$.

Algorithm of the Unitary Events method: Black color shows the part of the code used for reproduction of the results of Riehle et al (1997) and the gray color indicates the additional features which are now available in our reimplementation but not used specifically here.

input : `spike_trains`, `bin_size`, `window_size`, `window_step`, `pattern_hash_values`, `method`

output: time series of: suprise `S`, number of empirical coincidences `n_emp`, number of expected coincidences `n_exp`, firing rate of each neuron `rate`

```
Diescretize the input spike train data (spike_trains) and represent
  them as binary sequences (‘BinnedSpikeTrain()‘ from coversion
  module in Elephant)
for each hash value from pattern_hash_values do
    Define the sliding window positions over the whole length of the
      data given window_size and window_step (‘_winpos()‘ and
      ‘_bintime()‘)
    for each window position do
        Calculate n_emp (‘n_emp_mat()‘ and ‘n_emp_mat_sum_trial()‘)
        if method is “analytic_TrialAverage” or
          “analytic_TrialByTrial” then
            if method is “analytic_TrialAverage” then
                calculate n_exp based on “analytic_TrialAverage”
                  (‘_n_exp_mat_analytic()‘ and ‘n_exp_mat_sum_trial()‘)
            end
            else if method is “analytic_TrialByTrial” then
                Calculate n_exp based on “analytic_TrialByTrial”
                  (‘_n_exp_mat_analytic()‘ and ‘n_exp_mat_sum_trial()‘)
            end
            Generate Poisson distribution with the mean n_exp
        end
        else if method is “surrogate_TrialByTrial” then
            Generate n_exp distribution based on surrogates
              (‘_n_exp_mat_surrogate()‘ and ‘n_exp_mat_sum_trial()‘)
        end
        Calculate p-value of n_emp given the distribution of n_exp
    end
    Calculate the surprise as a function of time from the p-values at
      each window position (‘jointJ()‘)
end
```

The UE analysis is performed in a sliding window fashion, i.e. the data in each window are analyzed separately. This approach is chosen to account for potential changes of the firing rates in time and to follow the dynamics of the correlation. Here, a sliding time window is defined based on trial time meaning that a certain window position includes the activity of all neurons in all trials in a certain time interval of the trial. In the algorithm, at each position of the window, the data contained in the window are extracted in order to compute the significance of the specified patterns (line 3 in the pseudo code). For doing that we require the empirical number of occurrences of each pattern as well as its expected number given the firing rates. While the empirical number is directly extracted from the data (line 5 in the pseudo code), the method to compute the expected number can be chosen using the input parameter `method` depending on the assumptions regarding the data (line 6 and 15 in the pseudo code). For cross-trial homogeneous data, selecting "*analytic_TrialAverage*" as `method` (line 7 in the pseudo code) computes the expected number by estimating the rate of each neuron from the average spike count across trials [135, 136]. For data with cross-trial variability, the "*analytic_TrialByTrial*" method (line 10 in the pseudo code) should be used, which accounts for cross-trial changes in the firing rates by computing the expected number of occurrences of the spike pattern based on the product of single-trial estimates of firing rates [139]. Both methods perform a parametric test, where the number of occurrences of the spike pattern is assumed to be a Poisson-distributed. As an alternative non-parametric option, selecting "*surrogate_TrialByTrial*" as `method` (line 15 in the pseudo code) will numerically compute the distribution of the expected number by implementing the null-hypothesis based on surrogate spike trains [140]. In the following, we will explain these methods in greater detail.

In case of selecting *"analytic_TrialAverage"* - as used in the remainder of this study for the reproduction of [278] - the number of spikes per neuron within the sliding window is summed across trials and divided by the number of trials and bins contained, thus yielding the average probability p_i to have a spike of neuron i in a bin of the time window. The probability to find a particular pattern by chance in a bin is then computed by multiplication of the relevant probabilities, i.e. for a pattern [1,0,1] the probability p of occurrence is given by $p_{[1,0,1]} = p_1 * (1 - p_2) * p_3$. In this expression, $(1 - p_2)$ is the probability for contributing no spike to the pattern. The expected number of pattern occurrences, computed as the product of the occurrence probability p of the pattern, e.g. $p_{[1,0,1]}$, multiplied by the number of bins (across all trials) covered by the sliding window. The distribution of coincidences is given by a Poisson distribution with the mean equal to the expected number.

Alternatively, for the *"analytic_TrialByTrial"* method, the firing probability of each neuron is calculated in a trial-by-trial manner based on the spike counts per trial. The probability p of finding the pattern by chance is calculated by summing the products of the firing probabilities obtained individually from each trial. As for the trial-averaging method,

the expected number of pattern occurrences is given by multiplication with the number of bins of the trial in the sliding window, used as the mean of a Poisson distribution to obtain the distribution of expected coincidences.

As a third alternative, a surrogate method for estimating the expected number can be selected using "*surrogate_TrialByTrial*" for the parameter `method`. In this Monte-Carlo approach, a surrogate version of the spike trains is generated repeatedly, and from each surrogate the number of occurrences of the pattern of interest is counted. The method by which surrogates are generated from the input spike trains is spike time randomization of the spikes per trial and per neuron within the sliding window. The pattern counts obtained from this procedure form the distribution of the expected number of occurrences of the pattern, thus implementing the null-hypothesis under the constraints implied by the surrogate method.

The distribution obtained by either of the three methods above is then used for the significance test of the pattern on the basis of the empirical occurrence count. The p-value resulting from the test is then transformed by a logarithmic transformation to the surprise value (line 20 in the pseudo code), which indicates by positive or negative values more or less occurrences of the pattern than expected by chance, respectively. If the p-value is below a fixed prescribed level (e.g. below 5%, which corresponds to a surprise value exceeding 1.27), the occurrences of the spike pattern under investigation in the sliding window are marked as UEs for that pattern. This procedure is performed for each pattern of interest, and in each sliding window.

In the present study, in order to reproduce the original results we used the "*analytic_TrialAverage*" method, which reflects the analysis performed in the original publication [278]. The *"analytic_TrialByTrial"* and *"surrogate_TrialByTrial"* methods are extensions of original UE method, which were developed after the original publication and introduced in subsequent works [139, 140].

The reimplementation of the UE method is based on the data object model provided by the Neo library, upon which the Elephant library is based. The Neo library provides loading routines for a variety of data formats, including proprietary and generic data formats. The data sets available for reproducing Figures 2 and 4A of [278] were tab-separated ASCII text files containing two columns of integers (informally often referred to as 'GDF-format'): the first column provides event codes (behavioral events or neuron IDs), and the second column contains the time of the occurrence of these events (time stamps). The units of the time stamps are not contained in the data file. We partly extracted metadata information, in particular the time units and the meaning of the event codes, from a Matlab routine (provided by AR) operating on the GDF data file. However, only after further communication with AR we were able to identify the exact meaning of the content of the data files. Using this information, we wrote a new loading routine that loads the GDF data as Neo data objects.

Our reimplementation uses the `conversion` module of Elephant for converting the spike data (represented as a series of time stamps) into the binary sequence to guarantee a unique, global binning mechanism for all analysis methods provided in Elephant. The bin size to be set for the analysis was extracted from the original publication. However, defining the time point to start the binning of each single trial data required to know the alignment event in each trial and how much time before this event (pre-time) is considered. Since this information was not documented in the original paper, we tried several possibilities until we got an agreement with the original figures as will be shown in the Results.

To check if our Python implementation produces the same results as the implementation(s) used in the original publication, we compare each of our figures visually in detail with the original figures. This is also the test if we used the correct data files since the names of the data sets are not mentioned in the original publication. A direct numerical comparison to the original results plotted in these figures was not possible as this data is not available.

2.2.2 Reporducing the results of Riehle et al (1997)

For the reproduction of the original results in [278], we focus on reproducing Figure 2A-F and Figure 4A. Figure 2 is chosen because it represents the main result of the study and includes the UE analysis that underlies all subsequent analyses. Figure 4A is chosen because this is an example of the application of the UE method to data with more than 2 neurons. In terms of complexity of the code the implementation of the UE analysis for three or more neurons is considerably more demanding than for only two neurons. With this example we show that our implementation is capable of performing the UE analysis for the generic case of arbitrary numbers of neurons.

We apply our reimplementation of the UE method to preprocessed versions of the spike train data available to us after communication with AR, which in part were identical to those used in the original analysis. Also, we learned from AR that Figure 2 was generated by the Matlab implementation of the UE method while all remaining figures of the original publication, including Figure 4A, were generated by the older implementation in IDL.

Let us start with the reproduction of Figure 2 of the original publication. We first give a brief description of the experiment (see the original publication for details). After the monkey was presented with the preparatory signal (PS) he had to sit still and wait for a response signal (RS) to start his arm movement (i.e. equivalent to a GO signal). The duration of the waiting period was randomly selected on a trial-by-trial basis to be either 600, 900, 1200 or 1500 ms. In Figure 2 of [278] only trials of the longest waiting period (1500 ms) were used for the analysis. In these trials, times marked as expected signals ES1, ES2, and ES3 corresponded to the ends of the three shorter waiting periods, at which the monkey could have gotten the RS signal but did not. As the monkey was trained to recognize and distinguish the four waiting periods, but was not informed of the randomly

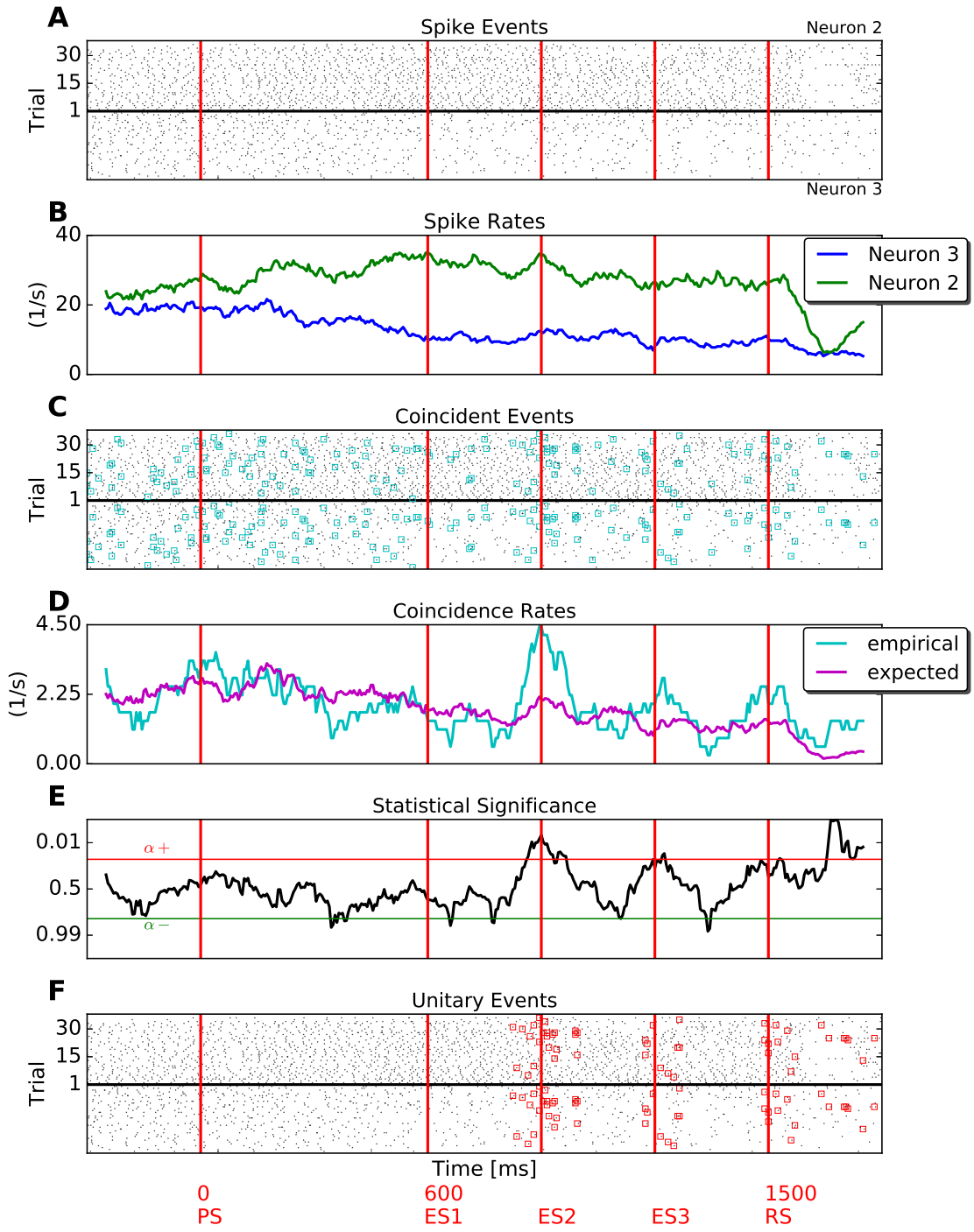

Figure 2.1: *(Caption on the next page)*

Figure 2.1 *(previous page)*: Initial attempt to reproduce Figure 2 of the original publication with trial alignment to PS. (**A**) Raster plot of two neurons (neuron 2: top of panel; neuron 3: bottom of panel) in 32 trials (sorted identically for both neurons). (**B**) Average firing rate of each neuron calculated across trials in a sliding window of length 100 ms in steps of 5 ms. (**C**) Same raster plot as in panel A with spike coincidences (i.e., pattern $[1, 1]$) between the two neurons marked by cyan squares. (**D**) Empirical (cyan) and expected (magenta) number of coincidences calculated in a time-resolved manner (parameters of sliding window identical to panel B). (**E**) Time course of the surprise measure, calculated in same sliding windows as in panel B. Surprise values that correspond to positive and negative significance levels $\alpha+ = 0.05$ and $\alpha- = 0.95$ are shown with by horizontal red and green lines, respectively. (**F**) Same raster plot as in panel A with significant coincidences, i.e. UEs, marked by red squares.

selected period for a given trial, ES1-ES3 were time points at which the monkey expected that a signal could occur.

Since the data file for Figure 2 of [278] contained the data as a continuous recording of one recording session (*winny131.gdf*; 2 neurons, and behavioral events), we extract the trials by cutting the data in a time window around specific trigger events that belong to trials of the longest waiting period, such that the complete trial is contained in the cut-out. In a subsequent step, the spike times in the individual trials are aligned to the trigger event, such that spike times in each trial are given as relative to the trigger.

The original publication does not provide information which trigger event was used for alignment. In this experiment, 2 triggers that occur in every trial could serve as trigger events, the preparatory signal PS (event code 114) and the response signal RS (event code 124). We noticed that the time interval between PS and RS for the longest trials was not identical across the respective trials and varied by $\pm$ 1 ms. Given the UE method is applied on a time scale of 5 ms, the analysis results therefore are expected to depend on whether trials are aligned to PS or RS. Thus, we decide to generate the results for both alignments .

Figure 2.1 and Figure 2.2 show the results of performing the UE analysis for both PS- and RS-aligned data, respectively. Here, the analysis parameters are set to the identical values as reported in the original publication (bin size: 5 ms, analysis time window size: 100 ms, time step of the sliding window: 5 ms, significance level $\alpha = 0.05$). The comparison of the two figures to the original figure shows agreement in the raster displays (panel A) and the time-resolved, trial-averaged firing rate estimates (B). However, although the graphs of the number of coincidences per sliding window (panel D) and the surprise measure (panel E) are similar in their overall general behavior, they differ in the details. Thus, indeed the choice of the alignment influences the analysis result. In order to test if one of the two alignments is in agreement with the original publication, we perform a detailed visual

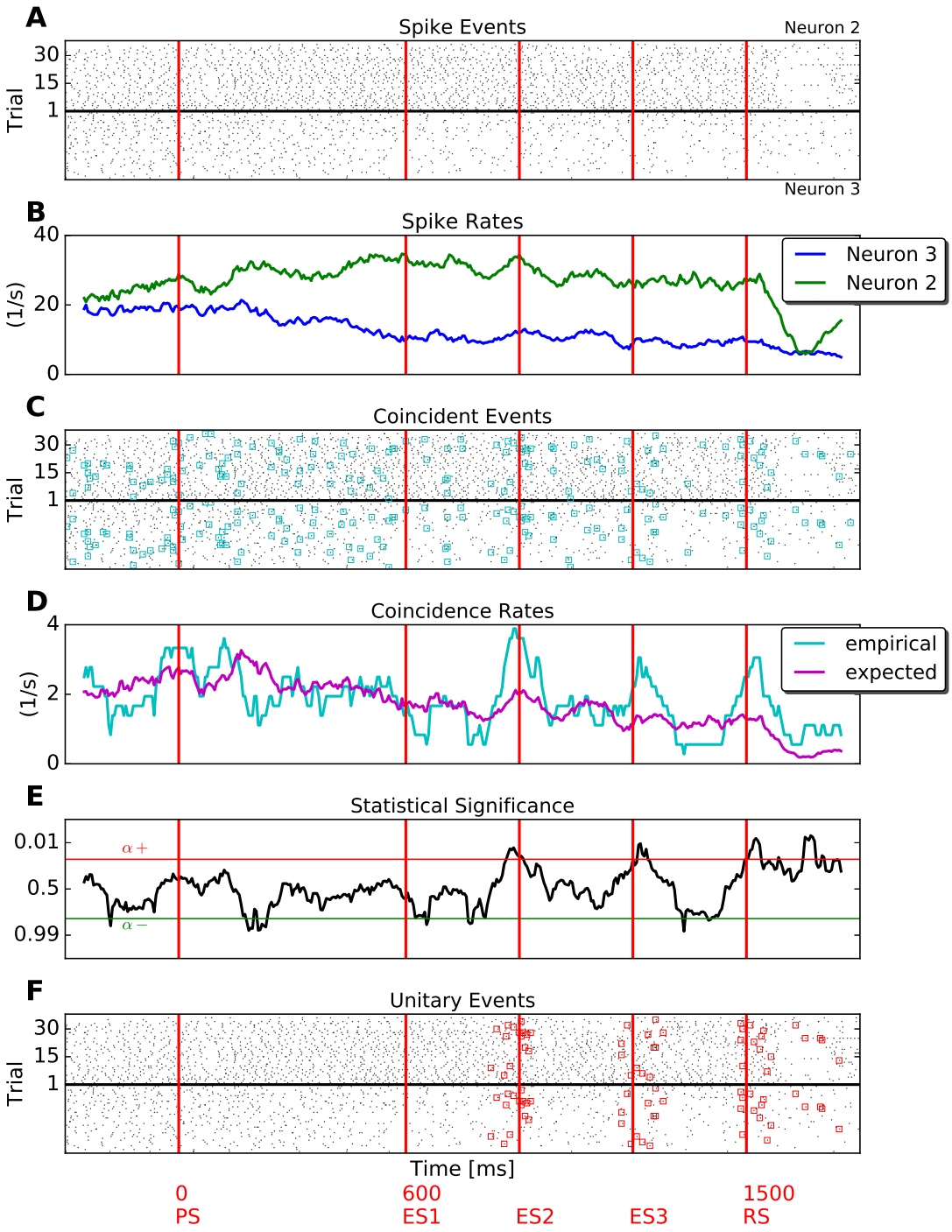

Figure 2.2: Reproduction of Figure 2 of the original publication with trial alignment to RS. The same conventions as in Figure 2.1 apply to the respective panels.

comparison of our two figures and the original one on the basis of the spikes marked as coincident (panels C) and as part of a UE (panel D). We notice that when aligning to PS (Figure 2.1), the marked spikes do not agree in all details with the original figure. However, Figure 2.2 with alignment of the trials to RS agrees completely with the original figure. Although we cannot check the identity to a further extent because the exact numerical values of the original results are not available anymore in electronic form, we believe that this visual inspection provides enough support for a reproduction of the original result by the reimplementation of the UE method.

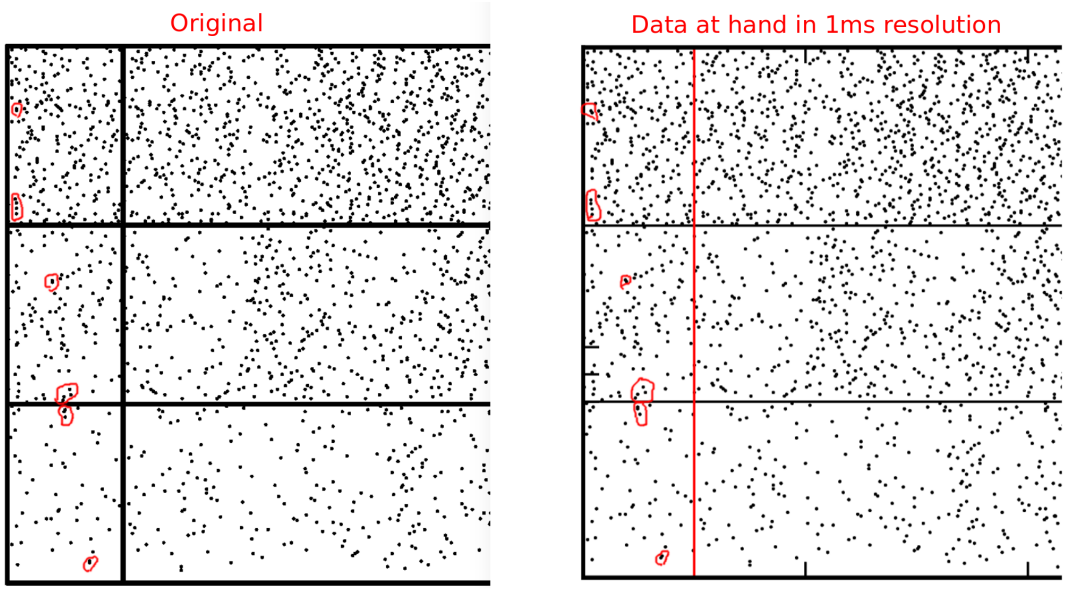

Figure 2.3: Close-ups of the original raster display in Figure 4A of the original publication (left) and the first data file available at hand for reproduction (right) reveal slight differences in the positions of some spikes. Data on the right are aligned (similar to the data on the left) to ES1 (event code 15 in the GDF data file). The time before the alignment event is chosen as 700 ms, and a bin size of 5 ms is used. The red marks indicate spike clusters with identified differences between the left and the right panel, where at least one spike is shifted in the right panel compared to the left panel.

As a next step we aim at reproducing Figure 4A of [278]. This figure contains the result of the analysis of three neurons recorded simultaneously, in contrast to Figure 2 where only two neurons are considered. We analyze the original data provided by AR with the parameter values given in [278] and compare our result to the original figure. Figure 4A of the original publication contains the raster displays of the data in the top panel, the raster displays with the marked coincident spikes (blue marks) in the middle panel, and the raster displays with the marked spikes that are part of a UE (red marks) in the bottom panel. We find that the UE result is different, as the UEs occur at different times and between different neurons compared to the original publication. Thus we check

whether the spike times of the individual spikes are identical between the original and our results. Figure 2.3 shows a segment of the raster plot of the original figure and the corresponding segment of our reproduced figure. We compare the positions of the single spikes and find that there are small discrepancies between the two raster plots in some of the spike times. Figure 2.3 demonstrates examples of clusters of spikes marked in red that should be identical in both raster plots but contain a few individual spikes that are slightly shifted in comparison to the original figure by a very small amount.

This leads us to the suspicion that the data are binned in a fashion that is not consistent with the data shown in the original publication. Personal communication with AR revealed that while Figure 2 had been generated by the Matlab implementation of the UE analysis, Figure 4 had been generated by the IDL implementation (see Introduction). A report by MD written before the time of the original publication summarized a comparison of the IDL and the Matlab implementations, and concluded that both were correct implementations of the method, but differed in their results due to a slightly different implementation of the down-sampling and binning of the raw data (recorded at 10 kHz). In the workflow for the IDL implementation, as illustrated in Figure 2.4 (the leftmost branch of the diagram), the raw data were first down-sampled to a temporal resolution of 0.5 ms (by a program `2gdf`) and then further rounded to 1 ms resolution integer values inside the IDL implementation. The data available to us had a resolution of 1 ms, which must have been a result of another down-sampling procedure than the one for the IDL implementation. This explains the difference in the raster displays, and this difference is likely also the cause that we were initially not able to reproduce the original UE result.

In our reproduction of Figure 2 of the original paper we use preprocessed data available in 1 ms resolution, that likely experienced the `alexa2gdf` program for conversion as shown in Figure 2.4 (the rightmost branch), before data are loaded into our reimplementation of the UE analysis. However, according to the aforementioned report by MD we only have a chance to reproduce Figure 4A of [278] if we have the original data or a version of them with a time resolution lower than 1 ms available (see the relevant ASCII drawing from the report in Figure 2.4). The original raw data with 0.1 ms resolution are presumably only available on a storage medium and format that at present we are not able to read and interpret. However, after we contacted AR she found the data ("jenny201_345_preprocessed.gdf") of Figure 4A of [278] with a time resolution of 0.5 ms (left box in the second row of the sketch shown in Figure 2.4).

We loaded this data at 0.5 ms resolution into Python and converted the data from the 0.5 ms to the required 1 ms resolution by the mathematical operation $\left\lfloor x + \frac{1}{2} \right\rfloor$, called "rounding half up". In numerical software packages, including Matlab, this operation is typically implemented by a function named round(). However, the round() implementation of NumPy (version 1.11.0) performs an even rounding, i.e., values exactly halfway between two integers are rounded to the nearest even integer. Indeed, the latter implementation

Figure 2.4: Illustration of the data preprocessing workflow, taken from the 1997 report of MD (in German) on the comparison of the first UE implementation in IDL (left branch) and the second implementation in Matlab [69] (right branch). Data entering both analysis branches have a resolution of 0.1 ms (top box). In the IDL branch spike times t in the original data are first transformed to a resolution of $h = 0.5$ ms (by the `2gdf` program, middle left) using the method of binning $\lfloor t/h \rfloor$. Then the data are read into the "IDL UE Software" (lower left box) and therein converted to $h = 1$ ms resolution by the method of "round half up" $\lfloor t/h + 1./2 \rfloor$ prior to analysis. Alternatively, one can load the 0.5 ms resolution data into the "MATLAB UE Software" (lower right box). Here the bin width is a parameter of the analysis and thus data can be converted to a 1 ms resolution but results are different from the IDL branch. Results are only identical if a prior transformation "T" (diagonal in center) performs a round half up and no further binning is done in the MATLAB program. At a later point in time the `alexa2gdf` converter function (written in MATLAB) became available such that data in the original 0.1 ms resolution could directly be converted to the 1 ms resolution by binning.

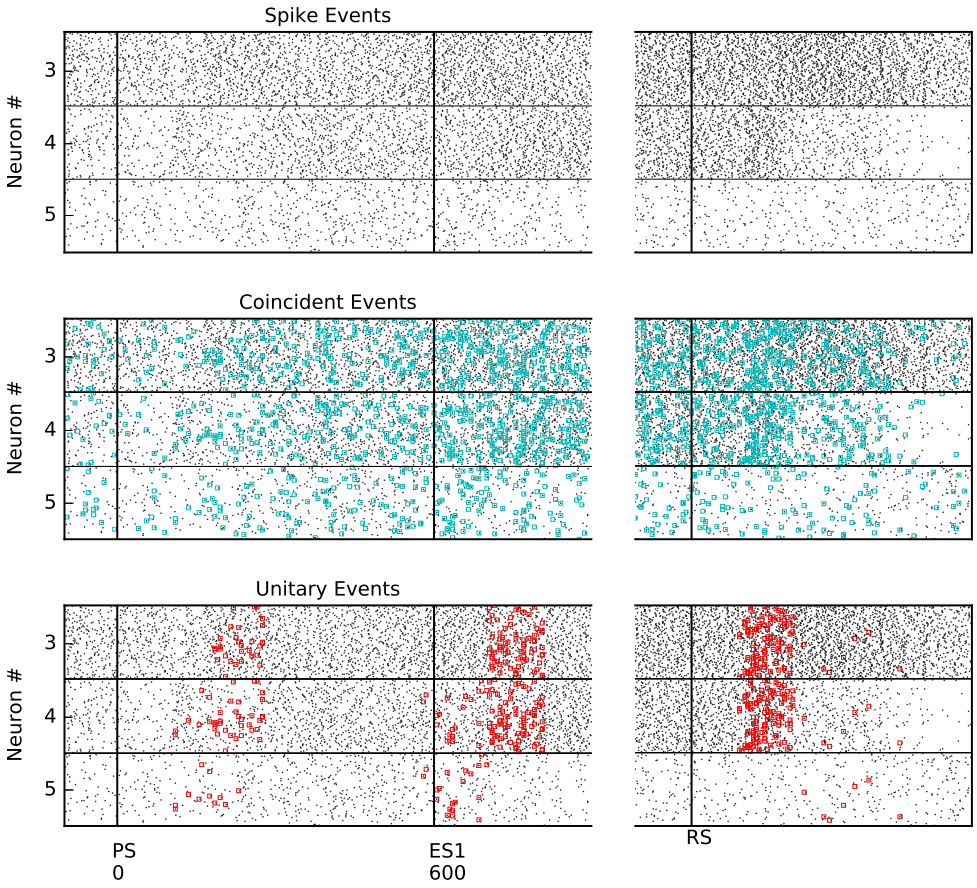

Figure 2.5: Reproduction of Figure 4A of the original publication. The left part of the figure shows the UE analysis result for the data aligned to ES1 (event code 15 in the GDF data file) with a time before the event (pre-time) set to 699 ms. The right part of the figure shows the analysis of the same data aligned to RS and with a pre-time of 99 ms. The left and right parts of the figure show 96 and 128 trials, respectively.

of rounding did not reproduce the result of the original publication. Thus, we used the expression `floor(x+0.5)` to perform rounding as it is implemented by Matlab. The procedure completely reproduces panel A of Figure 4 in the original publication (see Figure 2.5).

2.3 Population statistics uncovers baseline correlation in the cerebral cortex

In the previous section we have implemented the UE method as open source software and reproduced the result of [278] where they found behavior related, dynamic occurrence of excess synchrony. Now we are interested in searching for UEs in a large population of neurons. However, the UE analysis does not scale well due to the combinatorial explosion of the number of spike patterns, and due to the massive multiple testing problem involved.

In Kilavik et al. 2009 [196] the systematic occurrences of UEs in relation to behavior were studied by evaluating the UEs of many, sequentially recorded pairs of neurons. For this purpose a variant of the UE analysis was developed for the population statistics across many pairs of neurons by extending the expectancy of the occurrences of coincidences under the independence assumption of neurons as follows:

$$n_{exp}^{pairs} = \sum_{k=1}^{K} \sum_{m=1}^{M} p_i^{m,k} \cdot p_j^{m,k} \cdot \frac{T_w}{h} \tag{2.1}$$

with K: number of pairs, M : number of trials, $p_i^{m,k}$: firing probability of neuron i in k-th pair and m-th trial, T_w : duration of window, h : bin width.

To evaluate the significance of the synchrony across all pairs, first, the expected number of coincidences, n_{exp}^{pairs}, is calculated for a small portion of data (within a window size of about hundred of milliseconds). Then, the empirical coincidence count, n_{emp}^{pairs}, within each window is calculated by counting the actual number of coincidences for each pair and summing the number across pairs. The significance of the empirical coincidences is derived by calculating the p-value ψ of the deviation of n_{emp}^{pairs} from n_{exp}^{pairs}, i.e. the probability of getting n_{emp}^{pairs} or larger values assuming a Poisson distribution with the mean being n_{exp}^{pairs}. For a better visualization the surprise measure S, which is the logarithmic transformation of the p-value: $log_{10}[(1-\psi)/\psi]$, is used. Finally, applying a sliding window analysis enables to capture the time modulation of the surprise which can be related to the behavior.

We applied the above-mentioned extended UE analysis to a dataset recorded from monkey motor cortex while performing a center-out reaching task (see [50] for details of the experiment). The details of the analysis are as follows: only neurons with a firing rate of at least 5 Hz throughout the whole epoch of the analysis were chosen. Pairs of neurons were analyzed with neurons recorded on different electrodes only, leading to the total of

81 pairs. To assess the effect of the number of pairs on the surprise, we used a bootstrap procedure within each analysis window as follows:

1. We chose a number x of pairs to be analyzed: either $x = 20, 40, 60$ and 81.
2. We sampled x pairs randomly from the total number of available pairs (81) allowing duplication.
3. We calculated the values of n_{emp}^{pairs} and n_{exp}^{pairs} across the x pairs.
4. For a fixed value of x we repeated steps 2 and 3 for 1000 times.
5. We averaged the values of n_{emp}^{pairs} and n_{exp}^{pairs} , respectively, across the 1000 repetitions and calculated the surprise value from the averaged values $\bar{n}_{emp}^{pairs}$ and $\bar{n}_{exp}^{pairs}$.

With this procedure, we studied the surprise of populations of $x = 20$, 40, 60 or 81 pairs of neurons. Figure 2.6A shows the time modulation of the obtained surprises for different numbers of neuron pairs included in the analysis. The surprise is not modulated around zero but has a positive offset which depends on the number of pairs in the analysis. Including more pairs in the analysis increases the offset as it is shown in Figure 2.6A right panel.

In a next step we evaluated the time modulation of synchrony by introducing a new measure which does not depend on the number of pairs. This measure, called R, is the ratio between summed empirical and summed expected coincidences over pairs calculated in each analysis window:

$$R = \frac{n_{emp}^{pairs}}{n_{exp}^{pairs}}. \tag{2.2}$$

The time modulation of R, computed on the same data and with the same analysis parameters as in Figure 2.6A, is shown in Figure 2.6B. It demonstrates that, similar to what we observed in the time modulation of the surprise, there is a positive offset in R as well (mean value of R is above 1, where $R = 1$ is expected for independent data).

2.4 Flaw of UE or feature of the data?

The offset that we observed in the temporal modulation of the surprise S and the ratio R brought up the question whether this offset is the result of a not yet known flaw of the UE analysis, or if it reflects a feature of the data captured by the UE method. To answer this question we test the method with various artificial and surrogate data.

2.4.1 Test on independent and correlated Poisson data

First, we ask the question: How do the surprise S and the ratio R change as a function of the number of neuron pairs when they are derived from independent and correlated

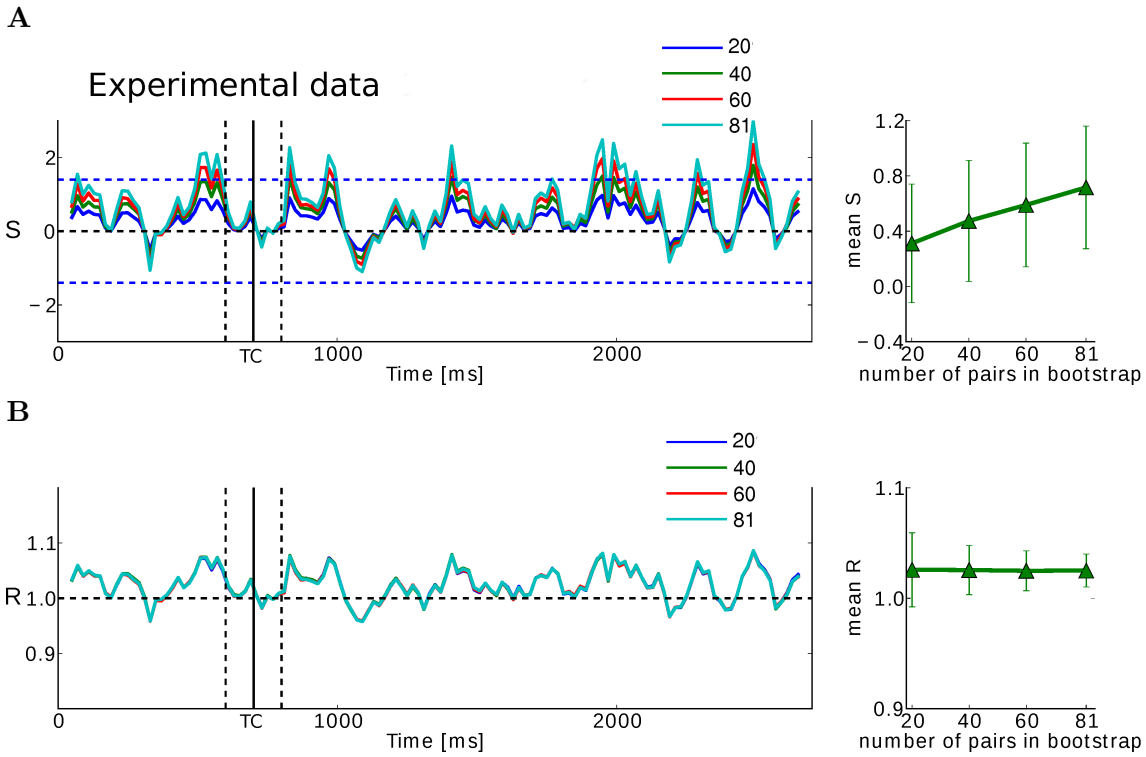

Figure 2.6: **Modulation of surprise and R resulted from experimental data.** Applying the extended UE analysis on the data recorded from the motor cortex of monkey in the delay trials before performing an arm movement task in which there is a temporal cue (TC) at 700 ms (see [50] for details of the experiment). (**A**) Left: Temporal modulation of surprise, summed across trials and different number of pairs. The window of analysis is 100 ms with a shift of 20 ms. The blue dashed line indicates the 5% significance level. 1000 bootstrapping was done to take into account the bias due to the sampling from different pairs. Each color corresponds to a different number of pairs included in the analysis. Right: Surprise as a function of included pairs in the analysis. The error bars are bootstrap errors on the estimation of mean value of surprise in time. (**B**) The same as in A for R.

Poisson spike data, respectively? To answer that, we apply the extended UE analysis to independent and correlated artificial spike data and compare the behavior of the surprise S and the ratio R as a function of the number of neuron pairs in each. Both data sets, the independent and the correlated data, are generated using a compound Poisson process (CPP) model [141, 327, 328]. The CPP model (as explained in Appendix 5.1.1) is a stochastic model to generate synthetic spike data with known correlation structure. The correlation structure is defined by a set of parameters called amplitude distribution $f_A(\omega)$, which represents the occurrence probabilities of spike synchrony events of different orders. The parameter ω, called amplitude, takes an integer value between 1 and N, and f_A fulfills $\sum_{\omega=1}^{N} f_A(\omega) = 1$ as it represents a probability distribution. The amplitude distribution is used to generate a marked Poisson process [75, 322] $s_{cp}(t) = \sum_j \delta(t - t_j) \cdot a_j$, called carrier process, where $\{t_j\}$ is a Poisson process of rate λ_{cp} and a_j is a random number between 1 and N drawn from f_A. From this carrier process a population of N correlated point processes x_i $(i = 1, 2, \cdots, N)$ is constructed by copying the event of the carrier process at t_j into each of a_j processes chosen from $\{x_i\}$. The specific process IDs which receive a spike copy are drawn randomly from $\{1, 2, \cdots, N\}$. Thus synchronous events of order ω are induced whenever events of amplitude ω occur in the carrier process. The mean firing rate of the individual processes is $\lambda = \int_0^T s_{cp}(t)dt/(NT)$, which depends on the rate of the carrier process λ_{cp} and the amplitude distribution f_A.

Using the CPP model we generate independent and correlated datasets with duration of $T = 1\,\mathrm{s}$ and $N = 100$ number of neurons. The firing rate of each neuron, in both datasets, is $\lambda = 20\,\mathrm{Hz}$. The amplitude distribution for the independent data has one entry only at $\omega = 1$ and for the correlated data it has entries at $\omega = 1$ and 20 with $f_A(20) = 5 \times 10^{-4}$.

Figure 2.7 shows the behavior of the expected and empirical coincidences as well as the ratio R and the surprise S as a function of the number of pairs in the analysis. The blue and red colors indicate the results for the independent and the correlated data sets, respectively.

The results of the independent data set (blue) show that increasing the number of pairs increases the empirical coincidences as well as the expected number of coincidences, but their ratio, i.e. R value, remains around 1. Similar to the R value, the surprise also does not increase with increasing the number of pairs and the mean value of the surprise remains around 0. For the correlated data (red), the empirical coincidences take higher values than the expected. The R value converges to a fixed value above 1 after around 500 pairs. The surprise increases more and more as we include more pairs in the analysis. This test shows that the Unitary Event analysis is exactly doing what it is supposed to do: Adding more pairs causes an accumululation of the evidence that the data are correlated, which increases the surprise.

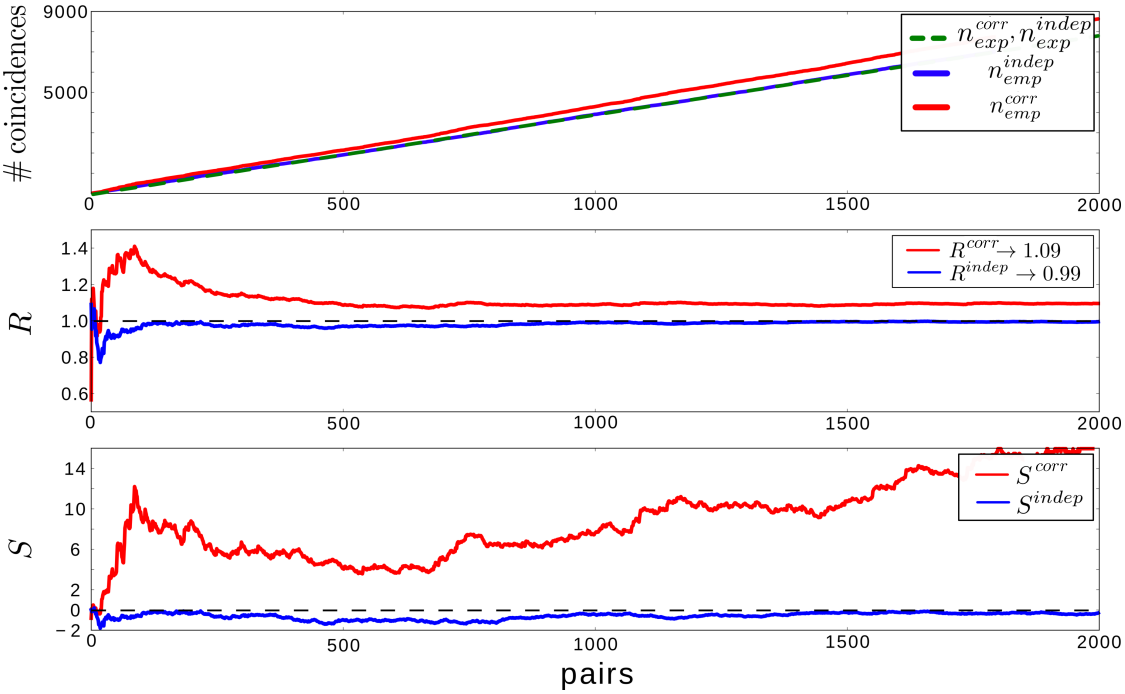

Figure 2.7: Testing the UE analysis with the independent and the correlated Poisson data.

2.4.2 Analysis of experimental data based on surrogate data

As a second test, to confirm that the positive offset in the surprise S and the ratio R observed in experimental data (Figure 2.6) reflects the correlation in these data, we do the following:

First, we generate surrogate data in which we aim to destroy the correlation in the experimental data set without changing the the firing rate of the neurons in each analysis window. For this purpose we randomize the spike times of each neuron in each analysis window (see [140] for an extensive discussion on surrogate methods). Then, we apply the extended UE method on these surrogate data, calculate R and S, and check whether the offset is still present. Figure 2.8 shows the results after the randomization. To account for the bias by of sampling from different number of pairs, we compute the bootstrap statistics of S and R (as explained in Section 2.3). The results show that the positive offset in S and R disappears by destroying the correlation between pairs of neurons. This is another confirmation that the UE analysis performs as expected: when there is no correlation, the accumulation of pairs does not increase the observed level of synchrony.

We conclude that the UE analysis is exactly doing what it is supposed to do. Thus the offset in the surprise and the R values, observed in our dataset is a real phenomenon. This was not observed before because such a huge number of neuron pairs had not been analyzed.

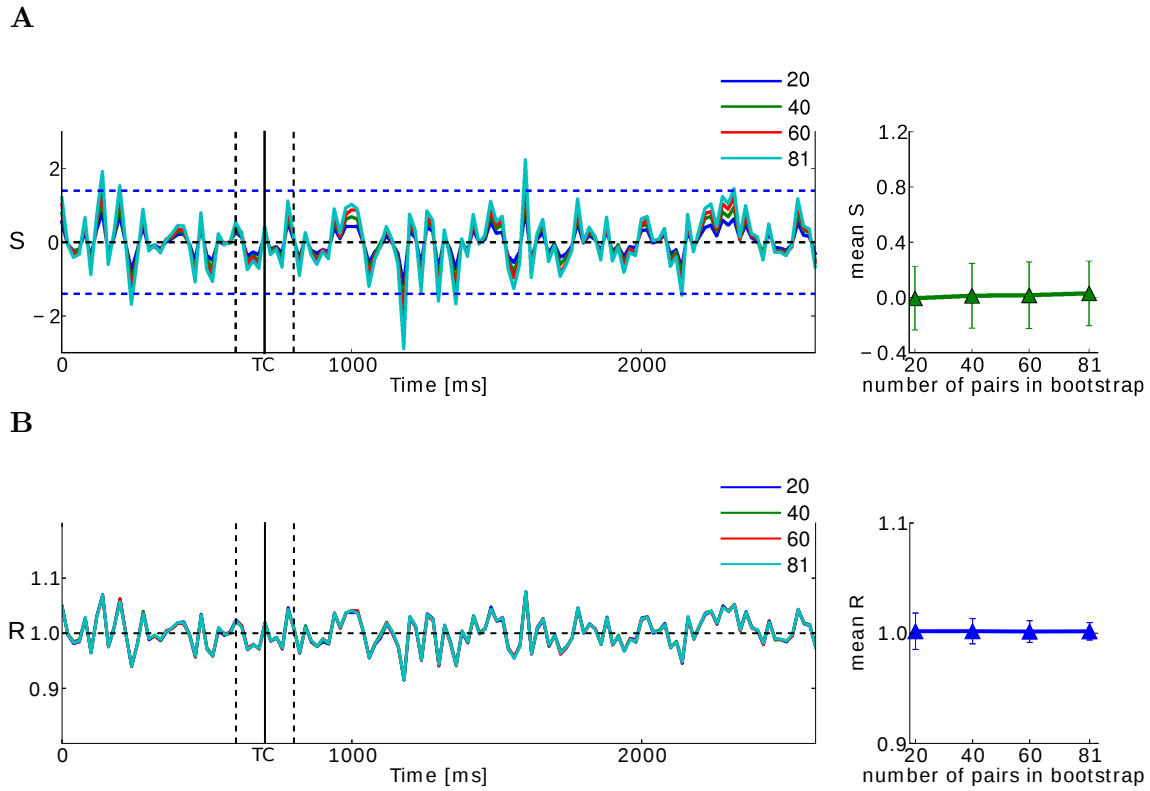

Figure 2.8: **Modulation of surprise and R resulted from surrogate data.** Testing the UE analysis with surrogate generated from experimental data by randomization of spike times. The figure is plotted similar to Figure 2.6.

2.5 Prediction from random balanced network

Anatomical data of the cortex show that there is a large connectivity convergence on individual neurons, and also a large divergence from individual to other neurons. A random balanced network model [38] considers this connectivity, and includes no other specific connectivity such as, e.g., the one proposed for the synfire chain [1]. We next test if such a network model without any specific connectivity structure, but merely random connectivity with known parameters, also exhibits the offset when using the extended UE analysis.

We simulate a randomly connected sparse network of inhibitory and excitatory neurons, so called random balanced network (RBN) [38], with $N = 10000$ neurons (80% excitatory and 20% inhibitory). For the simulation we used NEST [106], which is a simulator for spiking neural network models. The parameters of the simulation were set to values that result in an asynchronous irregular (AI) network dynamic. The AI dynamics often observed in the spontaneous activity of cortical neurons [250, 348], in which individual neurons have irregular spikes with stationary population activity (see [38] for explanation of different dynamics that appear in RBN). We are interested in such a particular dynamics of RBN because it holds the assumption of stationary firing rate and irregularity of spike times,

which are explicitly presumed in the null-hypothesis of the UE method. A raster plot of arbitrarily chosen 2000 neurons from our simulation is shown in Figure 2.9 top panel. By choosing a small bin of $1ms$, we calculate the population activity of neurons, which is the total number of active neurons within each bin (Figure 2.9 bottom panel).

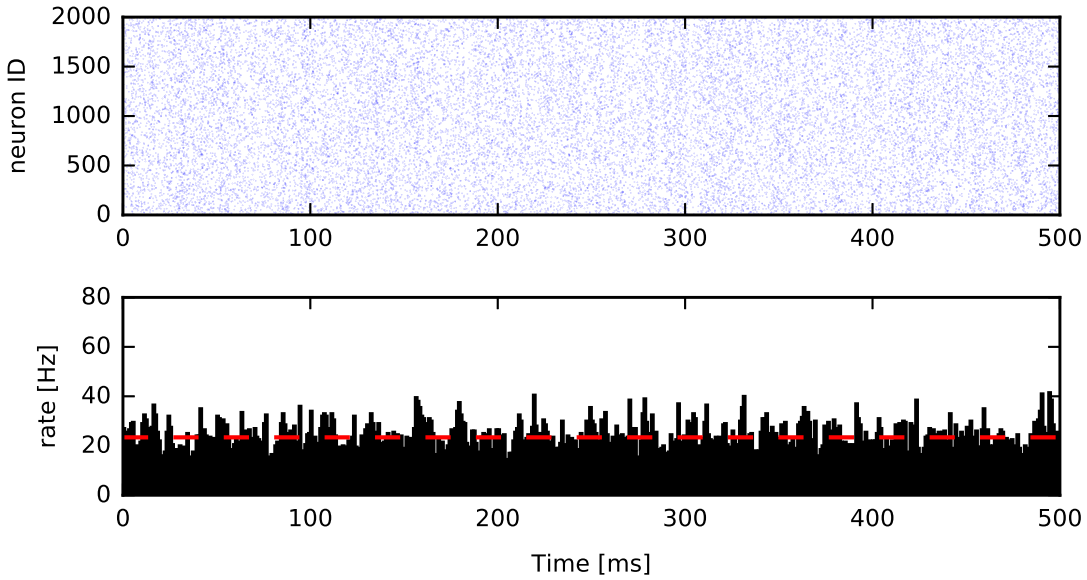

Figure 2.9: **Spike data simulated by RBN.** Snippet of raster plot (top) and population activity (bottom) of spike data generated from RBN network of 10000 neurons with 80% excitatory and 20% inhibitory (500 ms activity of 2000 randomly sampled neurons are shown here). The red dashed line in the bottom panel indicates the mean firing rate of neurons. The parameters for the network simulation were set as follow: The relative strength of inhibitory synapses $g = 6$ and the transmission delay $d = 0.5$ ms to gain asynchronous irregular (AI) dynamics.

Figure 2.10A shows the temporal modulation of the surprise obtained from the RBN data for different number of pairs included in the analysis. It shows a positive offset which increases by increasing the number of neuron pairs included in the analysis. Figure 2.10B demonstrates the temporal modulation of R with an offset larger than 1 irrespective of the number pairs included in the analysis. The magnitude of the offset in R is similar to the value observed in the experimental data (Figure 2.6).

This observation provides strong evidence that the extended UE analysis detects correlations which may be a consequence of direct or recurrent synaptic connections as well as shared inputs [158, 159, 204, 266, 276, 303, 333, 344], and refer to the correlation within a RBN as baseline correlation.

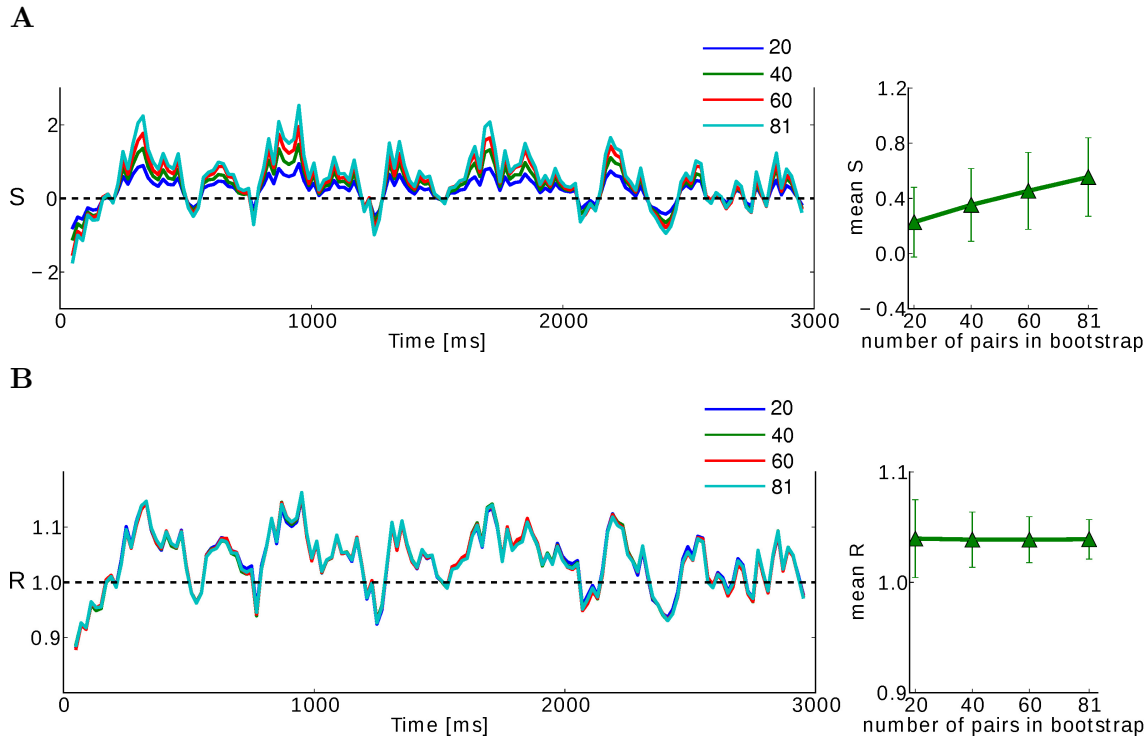

Figure 2.10: **Modulation of surprise and R resulted from RBN data.** Testing the UE analysis on the RBN data generated by the network simulation explained in Figure 2.9. The analysis parameters are similar to Figure 2.6.

2.6 Summary

We were able to reproduce the original results of [278] by applying a completely new reimplementation of the Unitary Events analysis method in Python to the original data. The method involves a number of numerical computations and is very sensitive. Therefore this reproduction of the result is a strong indication that the analysis in the original study was performing correctly and that the analysis program used there faithfully implemented the UE method. The UE software which we developed in Python and used here contains also extra features later developed for improving the statistics, i.e. accommodating cross-trial inhomogeneity, by calculating the expectation on the *TrialByTrial* basis for the rate estimation [139] or the possibility to use different surrogate methods for implementing the null-hypothesis [140, 141].

Furthermore, by applying the extended UE method on the data recorded from motor cortex of the monkey performing a center-out-reaching task, we captured the baseline correlation on the activity of real cortical neurons owing to the connectivity and shared input to a pair of neurons. We tested the method with synthetic data and surrogate data (generated from the experimental data) and showed that the Unitary Event analysis correctly captures the existence of baseline correlation. Moreover, by studying

the correlation structure in the random balanced network in an asynchronous irregular regime, we observed that the surprise S shows an offset and is not modulated around 0. R also shows a higher value than 1 and is close to the value of the offset of R observed in the experimental data.

CHAPTER 3

Inhibited maximum entropy (IME) model

The following chapter and the supplementary materials related to this chapter (Appendix 5.2) have been published as two preprints: i) **Rostami V**, Porta Mana PGL and Helias M, Pairwise maximum-entropy models and their Glauber dynamics: bimodality, bistability, non-ergodicity problems, and their elimination via inhibition. (2016) arXiv:1605.04740 [289], ii) Porta Mana PGL and Torre E and **Rostami V**, Inferences from a network to a subnetwork and vice versa under an assumption of symmetry. (2015) BioRxiv [271]. These works have been submitted to peer-reviewed journals and are under review.

Author contributions: Under supervision of PGL Porta Mana and Moritz Helias, the author perfomed the literature research, conceptual and mathematical derivation, simulation and analyses of the simulated as well as experimental data, model implementations and defining the new model. Sonja Grün provided useful feedback on the obtained results and on defining the biological questions that needed to be answered during the development of this project.

3.1 Introduction

Pairwise maximum-entropy models have been used in recent neuroscientific literature to construct a probabilistic description of the neuronal activity [21, 96, 230, 235, 310, 312]. The obtained probability distribution, by construction including the single-unit and pairwise statistics of the observation, could indeed help us to solve the segregation of baseline and functional correlation described in Chapter 2. In assigning to every observed activity pattern a probability, we obtain a measure of "surprise" for each such pattern; this measure is related to the logarithm of the probability and thus to Shannon's entropy [22, 112, 116, 117, 342]. Periods of low probability correspond to large surprise: these patterns cannot be explained by the statistical properties that entered the construction of the probability distribution. In this way, we are able to effectively differentiate expected, less surprising events from those that are unexpected, surprising, and potentially functionally meaningful.

Computing the maximum-entropy distribution from moment constraints – usually called the *inverse problem* – is very simple in principle: it amounts to finding the maximum of the entropy, which is a convex function and hence optimization is simple [82, 242, 282]: The maximum can be searched for with a variety of methods (downhill simplex, direction set, conjugate gradient, etc. [273, ch. 10]). The convex function, however, involves a sum over exp(number of neurons) terms. For 60 neurons, that is roughly twice the universe's age in seconds, but modern technologies enable us to record *hundreds* of neurons simultaneously [24, 42, 254, 279]. Owing to the exponential explosion for such large numbers of neurons, the convex function cannot be calculated (also not numerically) anymore, but therefore must be "sampled" usually via Markov chain Monte Carlo techniques [11, 31, 171, 203, 212, 225, 273]. In neuroscience the Glauber dynamics (also known as Gibbs sampling) [111, 225, chap. 29] is usually chosen as the Markov chain whose stationary probability distribution is the maximum-entropy one.

Boltzmann learning [4, 37, 161] is such an iterative combination of sampling and search for the maximum, and is still considered the most precise method of computing a maximum-entropy distribution.

A different approach is to approximate the convex function with an analytic expression, and to find the maximum directly via the study of the derivatives of this approximation. The mean-field [153, 252, 257], Thouless-Anderson-Palmer [257, 334], and Sessak-Monasson [300, 301] approximations are examples of this approach. These approximations are valid only in limited regions of the domain of the original convex function, and their goodness is usually checked against a Boltzmann-learning calculation (as e.g. in [293]).

Moment-constrained maximum-entropy models have also been used [140, 226] as generators of surrogate data, again via a Glauber dynamics. Such surrogates are used

to implement a null hypothesis to estimate the significance of correlations between spike trains [95, 102, 139, 140, 223, 267, 268].

The pairwise maximum-entropy model has been used for *experimentally* recorded activities of populations of a couple hundreds neurons at most, so far; but its success (or lack thereof) cannot be automatically extrapolated to larger population sizes. Roudi et al. [292] gave evidence that the maximized Shannon entropy and other comparative entropies of such a model may present qualitatively different features above a particular population size.

In this chapter we discuss a feature of the pairwise maximum-entropy model that may be problematic or undesirable: the marginal distribution for the population-averaged activity becomes *bimodal*, and one of the modes may peak at high activities. In other words, maximum-entropy predicts that the population fluctuates between a regime with a small fraction of simultaneously active neurons, and another regime with a higher fraction of simultaneously active neurons; the fraction of the second regime can be as high as 90%. This feature of the maximum-entropy model seems to have been observed before [33, 229, 230], but never remarked upon.

We also provide evidence that this bimodality is not just a mathematical quirk: it is bound to appear in applications to populations of more than a hundred neurons.It makes the pairwise maximum-entropy model problematic, for several reasons. First, in neurobiological data the coexistence of two regimes appears unrealistic – especially if the second regime corresponds to 90% of all units being active.

Second, two complementary problems appear with the Glauber dynamics and the Boltzmann-learning used to find the model's parameters. If the minimum between the two probability maxima is shallow, the activity alternately hovers about either regime for *sustained* periods, which is again unrealistic, and hence rules out this method to generate meaningful surrogate data. If the minimum between the two maxima is deep, the Glauber dynamics becomes practically *non-ergodic,* and the pairwise model *cannot be calculated at all* via Boltzmann learning or via the approximations previously mentioned [212, § 2.1.3][225, chap. 29]. This case is particularly subtle because it can go undetected: the non-ergodic Boltzmann learning still yields a reasonable-looking distribution, and this distribution gives back the moments used as constraints when re-checked with Monte Carlo sampling. However, *this distribution is not the sought pairwise* maximum-entropy *distribution*: the two differ quantitatively also for low activities.

The plan of the rest of the chapter is the following: after some mathematical and methodological preliminaries we show the appearance of the bimodality problem of the maximum-entropy model in the context of an experimental dataset of the activity of 159 neurons recorded from macaque motor cortex. Then we use an analytically tractable homogeneous pairwise maximum-entropy model (called "reduced" model for reasons explained later) to give evidence that the bimodality affects larger and larger ranges of datasets as

the population size increases. We show that experimental datasets of neural-activity are likely to fall within the bimodality ranges.

After analyzing the appearance of bimodality and the conditions for it, we also propose a way to eliminate it: using a slightly modified pairwise maximum-entropy distribution, which does not suffer from the bimodality problem. This modified distribution can be interpreted as arising from the principle of *minimum relative entropy* (also called minimum discrimination information) [5, 14, 43, 48, 51, 143–145, 162, 165, 176, 207–211, 277] with respect to a neurobiologically motivated reference prior, or as a maximum-entropy distribution with an additional constraint. The most important property of this modified distribution is its stationarity under a modified Glauber dynamics that includes a minimal *asymmetric inhibition*. This gives a neurobiological justification for using the modified distribution and its Glauber dynamics. We also show that the modified maximum-entropy distribution is the actual one obtained via Boltzmann learning or other approximations in the non-ergodic case – and thus could be the distribution actually computed in papers that used such techniques.

3.2 Methods

Our study uses three main mathematical objects: the pairwise maximum-entropy distribution, a "reduced" pairwise maximum-entropy distribution, and the Glauber dynamics associated with them. We review them here and give some additional remarks and references from the probability calculus viewpoint (see Appendix 5.2.5 for a comparison of the pairwise maximum-entropy in the probability calculus and the statistical mechanic).

3.2.1 Pairwise maximum-entropy model

First let us make mathematically clear what we mean by "activity": a set of sequences of spikes of N neurons during a finite time interval $[0, T]$. These spike sequences are discretized: we divide the time interval into n bins of identical length Δ equal to T/n, indexed by t in $\{1, \dots, n\}$. For each neuron i, the existence of one or more spikes in bin t is represented by $s_i(t) = 1$, and lack of spikes by $s_i(t) = 0$. With this binary representation, the activity of our population at time bin t is described by a vector: $\boldsymbol{s}(t) \coloneqq \big(s_i(t)\big)$. We will switch freely between vector and component notation for this and other quantities.

Time averages are denoted by a circumflex: $\widehat{\cdot}$, and *population* averages by an overbar: $\overline{\cdot}$. The activity summed over the population at time t, or time-resolved population-summed

activity, is denoted by $S(t)$, and the *population-averaged activity* by $\bar{s}(t)$:

$$\begin{aligned} S(t) &\coloneqq \sum_{i=1}^{N} s_i(t) \in \{0, 1, \dots, N\}, \\ \bar{s}(t) &\coloneqq \frac{1}{N} \sum_{i=1}^{N} s_i(t) \equiv \frac{1}{N} S(t) \in \{0, 1/N, \dots, 1\}. \end{aligned}$$

The time-averaged activity of neuron i is denoted by m_i:

$$m_i \coloneqq \widehat{s_i(t)} \coloneqq \frac{1}{T} \sum_{t=1}^{n} s_i(t), \tag{3.1}$$

and the time average of the product of the activities of the neuron pair ij, called raw covariance or *coupled activity*, is denoted by g_{ij}:

$$g_{ij} \coloneqq \widehat{s_i(t)\, s_j(t)} \coloneqq \frac{1}{T} \sum_{t=1}^{n} s_i(t)\, s_j(t). \tag{3.2}$$

These time averages will be used as constraints for the maximum-entropy model..

The *pairwise* maximum-entropy statistical model [33, 235, 297, 312] assigns a time-independent probability distribution for the population activity $\boldsymbol{s}(t)$ of the form (time is therefore omitted in the notation):

$$\begin{aligned} P_{\mathrm{p}}(\boldsymbol{s} \mid \boldsymbol{\mu}, \boldsymbol{\Lambda}) &= \frac{1}{Z_{\mathrm{p}}(\boldsymbol{\mu}, \boldsymbol{\Lambda})} \exp\Bigl(\sum_i \mu_i s_i + \sum_{i<j} \Lambda_{ij} s_i s_j\Bigr), \\ Z_{\mathrm{p}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) &\coloneqq \sum_{\boldsymbol{s}} \exp\Bigl(\sum_i \mu_i s_i + \sum_{i<j} \Lambda_{ij} s_i s_j\Bigr); \end{aligned} \tag{3.3}$$

the Lagrange multipliers $\boldsymbol{\mu}(\boldsymbol{m}, \boldsymbol{g})$ and $\boldsymbol{\Lambda}(\boldsymbol{m}, \boldsymbol{g})$ are determined by enforcing the equality of the time averages (Eq. 3.1) and (Eq. 3.2) with the single- and coupled-activity expectations, with their definitions

$$\mathrm{E}_{\mathrm{p}}(s_i) \coloneqq \sum_{\boldsymbol{s}} s_i\, P_{\mathrm{p}}(\boldsymbol{s}), \qquad \mathrm{E}_{\mathrm{p}}(s_i s_j) \coloneqq \sum_{\boldsymbol{s}} s_i s_j\, P_{\mathrm{p}}(\boldsymbol{s}) \tag{3.4}$$

(or in matrix form $\mathrm{E}_{\mathrm{p}}(\boldsymbol{s}) \coloneqq \sum_{\boldsymbol{s}} \boldsymbol{s}\, P_{\mathrm{p}}(\boldsymbol{s})$ and $\mathrm{E}_{\mathrm{p}}(\boldsymbol{s}\boldsymbol{s}^{\intercal}) \coloneqq \sum_{\boldsymbol{s}} \boldsymbol{s}\boldsymbol{s}^{\intercal}\, P_{\mathrm{p}}(\boldsymbol{s})$):

$$\mathrm{E}_{\mathrm{p}}(s_i) = m_i \quad \text{and} \quad \mathrm{E}_{\mathrm{p}}(s_i s_j) = g_{ij}. \tag{3.5}$$

Noting that $\mathrm{E}_{\mathrm{p}}(s_i) = P_{\mathrm{p}}(s_i = 1)$, and $\mathrm{E}_{\mathrm{p}}(s_i s_j) = P_{\mathrm{p}}(s_i = 1, s_j = 1)$, we see that the constraints above are equivalent to fully fixing the single-neuron probabilities of P_{p} and fixing one of its two-neurons probabilities.

By introducing the covariances $\boldsymbol{c}$ and Pearson correlation coefficients $\boldsymbol{\rho}$,

$$\begin{aligned} c_{ij} &\coloneqq \mathrm{E}(s_i s_j) - \mathrm{E}(s_i)\mathrm{E}(s_j), \\ \rho_{ij} &\coloneqq \frac{c_{ij}}{\sqrt{\left[\mathrm{E}_{\mathrm{p}}(s_i^2) - \mathrm{E}_{\mathrm{p}}(s_i)^2\right]\left[\mathrm{E}_{\mathrm{p}}(s_j^2) - \mathrm{E}_{\mathrm{p}}(s_j)^2\right]}}, \end{aligned} \tag{3.6}$$

the constraints above are jointly equivalent to

$$\mathrm{E}_{\mathrm{p}}(s_i) = m_i \quad \text{and} \quad c_{ij} = g_{ij} - m_i m_j \tag{3.7}$$

or

$$\mathrm{E}_{\mathrm{p}}(s_i) = m_i \quad \text{and} \quad \rho_{ij} = \frac{g_{ij} - m_i m_j}{\sqrt{(m_i - m_i^2)\,(m_j - m_j^2)}} \tag{3.8}$$

Note that the covariance constraints $c_{ij} = g_{ij} - m_i m_j$ by themselves are not convex, i.e., they do not define a convex subset in the probability simplex on which the entropy is maximized. The Lagrange-multiplier method does not guarantee the uniqueness of the solution if *only* the covariances are constrained. Uniqueness has to be checked separately [52, 53, 82, 99, 243]. On the other hand, the constraints $\mathrm{E}_{\mathrm{p}}(s_i) = m_i$ and $\mathrm{E}_{\mathrm{p}}(s_i s_j) = g_{ij}$ are separately convex, thus their conjunction $\mathrm{E}_{\mathrm{p}}(s_i) = m_i \bigwedge \mathrm{E}_{\mathrm{p}}(s_i s_j) = g_{ij}$ is convex too, and the bijective correspondence of the latter with $\mathrm{E}_{\mathrm{p}}(s_i) = m_i \bigwedge c_{ij} = g_{ij} - m_i m_j$ guarantees that the latter set of constraints is convex as well. What we have said about the covariances $\boldsymbol{c}$ also holds for the correlations $\boldsymbol{\rho}$.

It is important to remember that $(\boldsymbol{m}, \boldsymbol{g})$ are physically measurable quantities, independent of the observer, whereas $\big(\mathrm{E}_{\mathrm{p}}(s_i), \mathrm{E}_{\mathrm{p}}(s_i s_j)\big)$ depend on the observer's uncertainty, quantified by her probability assignment, and are not physically measurable. Therefore the constraints (3.5) are not trivial definitions ("$\coloneqq$") or equivalences ("$\equiv$"). In fact, in particular situations it does not make sense to enforce some of the constraints [270]; this also depends on *what is our uncertainty about*. Let us explain this point. The maximum-entropy distribution represents our uncertainty about the population activity for each of the *given* time bins, $t \in \{1, \dots, n\}$; this can be shown by symmetry and combinatorial arguments within the probability calculus [51–53, 93, 142, 179–181, 186, 270, 317]. Sometimes the maximum-entropy distribution is also used to represent someone's uncertainty about a *new* observation about a new time bin, e.g. t equal to $n + 1$. But such use implies additional assumptions and a particular prior that are not always justified [270]. Here is an example: suppose the time average of the coupled activity of neurons 1 and 2 vanishes: $g_{1\,2} \coloneqq \frac{1}{T}\sum_{t=1}^{n} s_1(t)\, s_2(t) = 0$ (this happens for a couple of pairs in our data). If we enforce the constraint $\mathrm{E}_{\mathrm{p}}(s_1 s_2) = g_{1\,2} = 0$, then maximum-entropy says that it is *impossible* that neurons 1 and 2 spike together: $P_{\mathrm{p}}(s_1 = 1, s_2 = 1) = 0$ (the corresponding Lagrange multiplier $\Lambda_{1\,2} = -\infty$). This prediction makes sense if we are speaking about any of our n time bins – in fact, $g_{1\,2} = 0$ means that neurons 1 and 2 have never spiked together in

our data, so the prediction is right. But it is an unreasonable prediction about a future or past time bin that is not part of our data: just because neurons 1 and 2 have not spiked simultaneously in our n data bins, we cannot conclude that it is *impossible* for them to spike or have spiked simultaneously in the future ($t > n$) or in the past ($t < 0$). Therefore, when some constraints assume extreme values, as $g_{1\,2} = 0$ in our example, it is not meaningful to use the maximum-entropy model for *new* predictions outside the given dataset. In this case it is more appropriate to use the full (Bayesian) probability calculus [26, 28, 100, 225, 270, 356, 363], possibly with maximum-entropy ideas on a more abstract level (space of prior distributions) [45, 46, 283–285, 316]. Contrary to what is sometimes stated in the literature, it is not true that the maximum-entropy model can only be used if the time sequence of activities is "stationary". This model represents a guess about the activities in the sequence, given *time-average* information. This guess, therefore, has to be time-invariant by symmetry: any time-dependent information has been erased by the time averaging. In other words, it is our guess which is "stationary", not the physical data; but it is still a good guess, given the time-independent information provided. With time-dependent constraints we would obtain a time-dependent maximum-entropy distribution [cf. 160]: this application of the maximum-entropy principle is called "maximum-calibre" [123, 178, 185, 187, 219, 274].

3.2.2 Reduced maximum-entropy model

If the time-averaged activities $\boldsymbol{m}$ are homogeneous, i.e. equal to one another and to their population average $\bar{m}$, and the $N\,(N-1)/2$ time-averaged coupled activities $\boldsymbol{g}$ are also homogeneous with population average $\bar{g}$, $\bar{g} \coloneqq \frac{2}{N\,(N-1)} \sum_{i<j} g_{ij}$, then the pairwise maximum-entropy distribution has homogeneous Lagrange multipliers by symmetry: $\mu_i = \mu_{\mathrm{r}}$ and $\mathit{\Lambda}_{ij} = \mathit{\Lambda}_{\mathrm{r}}$. It reduces to the simpler and analytically tractable form

$$\begin{aligned} P_{\mathrm{r}}(\boldsymbol{s}|\,\mu_{\mathrm{r}}, \mathit{\Lambda}_{\mathrm{r}}) &= \frac{1}{Z_{\mathrm{r}}(\mu_{\mathrm{r}}, \mathit{\Lambda}_{\mathrm{r}})} \exp[\mu_{\mathrm{r}} N\bar{s} + \tfrac{1}{2}\mathit{\Lambda}_{\mathrm{r}}\, N\bar{s}\,(N\bar{s} - 1)], \\ Z_{\mathrm{r}}(\mu_{\mathrm{r}}, \mathit{\Lambda}_{\mathrm{r}}) &\coloneqq \sum_{\boldsymbol{s}} \exp[\mu_{\mathrm{r}} N\bar{s} + \tfrac{1}{2}\mathit{\Lambda}_{\mathrm{r}}\, N\bar{s}\,(N\bar{s} - 1)], \end{aligned} \tag{3.9}$$

which assigns equal probabilities to all those activities $\boldsymbol{s}$ that have the same population average $\bar{s}$. In this homogeneous case, the values of the multipliers are equal to their averages: $\mu_i = \mu_{\mathrm{r}} = \bar{\mu} \coloneqq \frac{1}{N}\sum_i \mu_i$ and $\mathit{\Lambda}_{ij} = \mathit{\Lambda}_{\mathrm{r}} = \bar{\mathit{\Lambda}} \coloneqq \frac{2}{N\,(N-1)} \sum_{i<j} \mathit{\Lambda}_{ij}$.

This simpler distribution could be interpreted as an approximation of the pairwise maximum-entropy one, achieved by disregarding population inhomogeneities of the constraints m_i and g_{ij}. . But it is also an exact maximum-entropy distribution in its own right, obtained by only constraining the expectations for the *population sums* of the single

and coupled activities,

$$\sum_i s_i = S = N\bar{s}, \qquad \sum_{i<j} s_i s_j = S\,(S-1)/2 = N\bar{s}\,(N\bar{s}-1)/2,$$

to be equal to their measured time averages:

$$\mathrm{E_r}(N\bar{s}) = N\bar{m} \quad \text{and} \quad \mathrm{E_r}(N\bar{s}\,(N\bar{s}-1)) = \frac{N\,(N-1)}{2}\bar{g} \coloneqq \sum_{i<j} g_{ij} \tag{3.10}$$

(or equivalently constraining the population averages.)

For this reason we call the model (Eq. 3.9) a *reduced* (pairwise) maximum-entropy model. If the time-averages are homogeneous, then $\mu_\mathrm{r} = \bar{\mu} = \mu_i$, $\Lambda_\mathrm{r} = \bar{\Lambda} = \Lambda_{ij}$ and the reduced and full pairwise model coincide. But in the inhomogeneous case the multipliers of the reduced model are *not* equal to the averages of the pairwise one: $\mu_\mathrm{r} \neq \bar{\mu}$, $\Lambda_\mathrm{r} \neq \bar{\Lambda}$.

It is straightforward to derive the probability distribution for the population average $\bar{s}$ in this model, owing to its symmetry: if the average is $\bar{s}$, then the total number of active neurons must be $N\bar{s}$ in the population, and there are $\binom{N}{N\bar{s}}$ ways in which this is possible, all having equal probability given by (Eq. 3.9). Therefore,

$$\begin{aligned} P_\mathrm{r}(\bar{s}|\,\mu_\mathrm{r}, \Lambda_\mathrm{r}) &= \frac{1}{Z_\mathrm{r}(\mu_\mathrm{r}, \Lambda_\mathrm{r})} \binom{N}{N\bar{s}} \exp[\mu_\mathrm{r} N\bar{s} + \tfrac{1}{2}\Lambda_\mathrm{r}\, N\bar{s}\,(N\bar{s}-1)], \\ Z_\mathrm{r}(\mu_\mathrm{r}, \Lambda_\mathrm{r}) &\coloneqq \sum_{\bar{s}} \binom{N}{N\bar{s}} \exp[\mu_\mathrm{r} N\bar{s} + \tfrac{1}{2}\Lambda_\mathrm{r}\, N\bar{s}\,(N\bar{s}-1)]. \end{aligned} \tag{3.11}$$

This probability distribution $P_\mathrm{r}(\bar{s})$ can, in turn, also be obtained applying a minimum-relative-entropy principle [5, 43, 48, 51, 143–145, 176, 207–211, 277], i.e. minimizing the relative entropy (or discrimination information)

$$H(P, P_0) \coloneqq \sum_{\bar{s}} P(\bar{s}) \ln \frac{P(\bar{s})}{P_0(\bar{s})} \tag{3.12}$$

of $P(\bar{s})$ with respect to the reference distribution $P_0(\bar{s}) = 2^{-N}\binom{N}{N\bar{s}}$ while constraining the first two moments of $P_\mathrm{r}(S)$, or equivalently its first two factorial moments [272], $(\mathrm{E}(S), \mathrm{E}(S\,(S-1)/2))$.

It is easy to see that in this model, by symmetry, we also have

$$\mathrm{E_r}(s_i) = \mathrm{E_r}(\bar{s}), \qquad \mathrm{E_r}(s_i s_j) = \mathrm{E_r}\Big(\frac{N\bar{s}\,(N\bar{s}-1)}{N\,(N-1)}\Big), \tag{3.13}$$

$$c_{ij} = \bar{c} = \mathrm{E_r}\Big(\frac{N\bar{s}\,(N\bar{s}-1)}{N\,(N-1)}\Big) - \mathrm{E_r}(\bar{s})^2, \qquad \rho_{ij} = \bar{\rho} = \frac{\bar{c}}{\mathrm{E_r}(\bar{s}) - \mathrm{E_r}(\bar{s})^2}, \tag{3.14}$$

and $\Big(\mathrm{E_r}(\bar{s}), \mathrm{E_r}\Big(\frac{N\bar{s}\,(N\bar{s}-1)}{N\,(N-1)}\Big)\Big)$, $(\mathrm{E_r}(\bar{s}), \bar{c})$, $(\mathrm{E_r}(\bar{s}), \bar{\rho})$ are equivalent sets of constraints (but $\bar{c}$ and $\bar{\rho}$ by themselves are not convex).

This reduced maximum-entropy model is mathematically very convenient, because the Lagrange multipliers $\mu_\mathrm{r}, \Lambda_\mathrm{r}$ can be easily found numerically (with standard convex-optimization methods like downhill simplex, direction set, conjugate gradient, etc. [273, ch. 10]) with high precision even for large (e.g., thousands) population sizes N.

Summarizing: the reduced maximum-entropy model can be seen as: 1. the form taken by the pairwise maximum-entropy model in the case of homogeneous single and couple activities; 2. an approximation to the pairwise maximum-entropy model in the case of inhomogeneous single and couple activities; 3. a maximum-entropy model in its own right, that uses less information than the full pairwise model.

3.2.3 Glauber dynamics

Next we will use the Glauber dynamics for estimating the parameters of our maximum-entropy model and address specific adaptations that become necessary for our context. The pairwise and reduced maximum-entropy distributions discussed before do not give complete information about the dynamical or kinematical properties of the population activity They are, however, identical in form to the stationary distribution of an asynchronous Glauber dynamics [111] with symmetric "couplings" $\boldsymbol{\Lambda}$, sometimes interpreted as symmetric synaptic couplings, and "biases" $\boldsymbol{\mu}$, or sometimes interpreted as either a threshold or external input controlling the base activity of individual neurons. In the reduced maximum-entropy model these parameters are homogeneous: $\Lambda_{ij} = \Lambda_\mathrm{r}$, $\mu_i = \mu_\mathrm{r}$. The full and reduced maximum-entropy distributions give some information about this particular dynamics, like the appearance of metastable or most probable population-average states.

If we assume that our uncertainty about the *evolution* of the population activity can be modelled by the Glauber dynamics of a binary network, we can choose the $\boldsymbol{\mu}, \boldsymbol{\Lambda}$ parameters determined by the constraints (Eq. 3.5) and thus generate surrogate data that – if the dynamics is ergodic – have infinite-time-average activities as our initial experimentally observed data.

3.3 Results

3.3.1 The problem: bimodality, bistability, non-ergodicity

We first show how the bistability problem subtly appears with a set of experimental data, then explore its significance for larger population sizes.

Example with experimental data

Our data consists of the activity of a population of $N = 159$ neurons recorded from motor cortex of macaque monkey for 15 minutes using a 100-electrode "Utah" array and the experimental setup as described in [279]. , but with a different behavioral design. Here the monkey was awake and alert, but without performing a task during the recording. We choose this behavioral protocol for retrieving 'resting state' (or ongoing) [12] data to characterize the "ground" state (in contrast to a task or functional state).

Figure 3.1A shows a raster plot (2 sec out of 15 min for better visibility) of the activities $\boldsymbol{s}(t)$ of all the simultaneously recorded neurons . The population histogram (or time-resolved population-summed activity) $S(t)$ for this period is shown underneath. The distributions of the time-averaged single and coupled activities m_i, g_{ij}, and the corresponding empirical covariances c_{ij} measured in the full data set of 15 min are shown in panels B, C, D. The population averages of these quantities are

$$\bar{m} \approx 0.0499, \quad \bar{g} \approx 0.00261, \quad \bar{c} \approx 0.000135, \quad \overline{\rho} \approx 0.00319. \tag{3.15}$$

Let us now find, via Boltzmann learning, the Lagrange multipliers of the pairwise maximum-entropy model constrained by the empirical single and coupled activities of the experimental data and shown in Figure 3.1B–C. At each iteration, the sampling phase of the Boltzmann learning has 10^6 timesteps; an example is shown in Figure 3.2A. We decided to use a large number of time steps (n=1000) as compared to Roudi et al. [293] (n=200) or Broderick et al. [37] (n=400) to make sure that the learning in our case really converged. We obtain the Lagrange multipliers (μ_i, Λ_{ij}) whose distributions are shown in Figure 3.2D. The final single and coupled activities are shown in Figure 3.2C, and compar them to the experimental data. We find that the model parameters are highly correlated with the experimental ones and thus describe the data well. Sampling the maximum-entropy distribution with the obtained Lagrange multipliers, we obtain the population-average probability distribution (Figure 3.2B, red). It disagrees in the tail from the distribution based on the experimental data (dashed), which we will address next.

The results of the Boltzmann learning do not show any inconsistency at this point. But now we sample the distribution for a much longer time, i.e. 5×10^7 steps. The result is shown in Figure 3.3A. Here we find that after roughly 2×10^6 steps, the whole population (i.e. all neurons) jumps to a high-activity regime and remains there until the end of the sampling. Thus, we have discovered that the Glauber dynamics has an additional metastable high-activity regime. Then we wondered if there are even more metastable regimes. Therefore we started the dynamics from different activity levels of the population-summed activity, and observed two metastable regimes (see Figure 3.3B) -

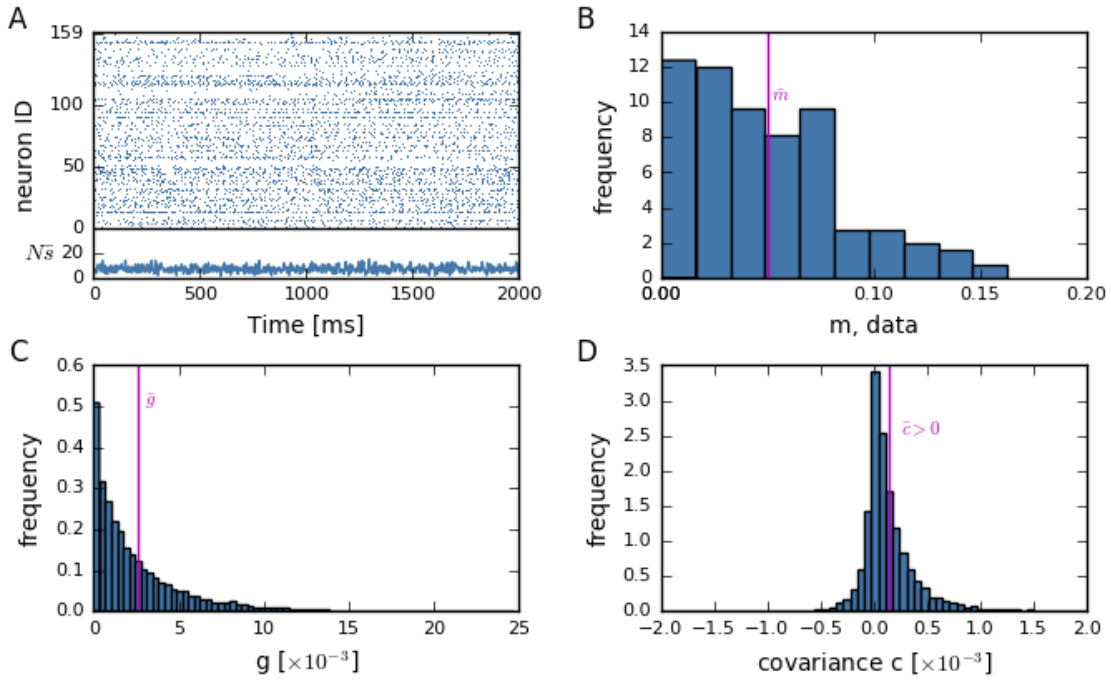

Figure 3.1: **Experimental data and their empirical first- and second-order statistics.** (**A**) Example raster display (snippet of 2 sec of the total data of 15min) of 159 parallel spike recordings of macaque monkey during a state of "ongoing activity". The experimental data are recorded with a 100-electrode "Utah" array (Blackrock Microsystems, Salt Lake City, UT, USA) with 400 μm interelectrode distance, covering an area of 4×4 mm (session: s131214-002). The population-summed activity $S(t)$ shows the number of active neurons within each time bin t of width $\Delta = 3$ ms. (**B**) Population distribution of the time-averaged activities m_i of each of the neurons i, Figure 3.1. The vertical line marks the population average, $\bar{m} := \frac{1}{N}\sum_i m_i$ (**C**) Population distribution of the time-averaged r aw covariances (coupled activities) g_{ij}, Figure 3.2. The vertical line marks the population average, $\bar{g} := \frac{2}{N(N-1)}\sum_{i<j} g_{ij}$ (**D**) Population distribution of the covariances $c_{ij} = g_{ij} - m_i m_j$. The vertical line again marks the population average, $\bar{c} := \frac{2}{N(N-1)}\sum_{i<j} c_{ij}$ with a slightly positive average correlation. (Histograms bins in B, C, D computed with Knuth's rule [198] and calculated over the full recordings (15 minutes)). Data courtesy of A. Riehle and T. Brochier.

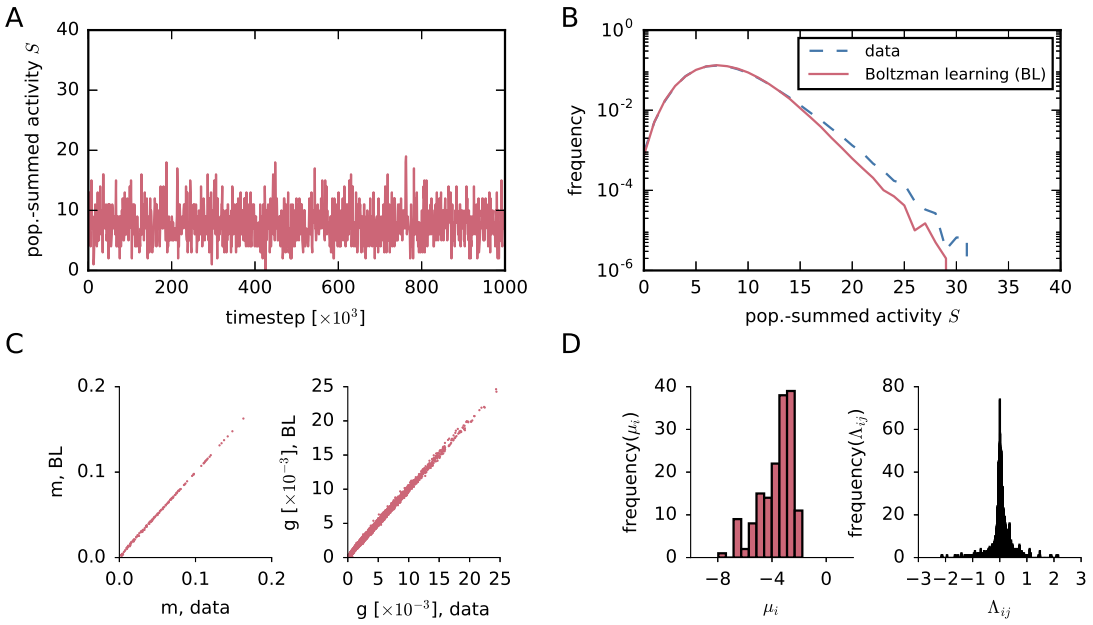

Figure 3.2: **Results of Boltzmann learning**. (**A**) Population-summed activity $S(t)$ of $N = 159$ neurons, obtained via Glauber dynamics in 10^6 timesteps. The couplings Λ_{ij} and biases μ_i of the Glauber dynamics are the Lagrange multipliers, shown in panel D, found by Boltzmann learning (10^6 samples) from the experimental rate time-averages of all neurons m_i and all raw pairwise correlations g_{ij} of Figure 3.1. (**B**) Red, solid: Probability distribution of the population-summed activity, sampled via the Glauber dynamics. Blue, dashed: empirical distribution of the population-summed activity from the dataset shown in A. (**C**) Time averages m_i and g_{ij} obtained from Boltzmann learning (10^6 samples), versus the experimental ones. (**D**) Population distribution of the Lagrange multipliers μ_i and Λ_{ij} obtained via Boltzmann learning (10^6 samples). (Histogram bins in D computed with Knuth's rule [198].)

one at a high activity and one at a low activity. This means that the actual distribution associated with the Lagrange multipliers of Figure 3.2D is *bimodal.*

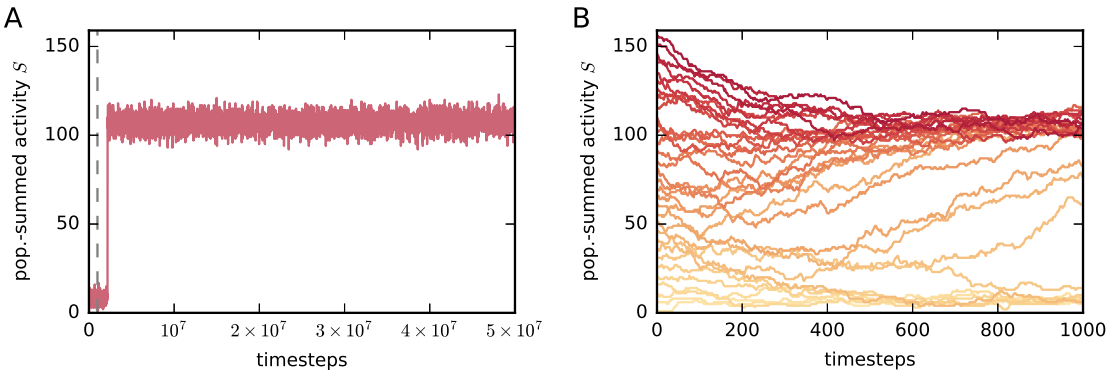

Figure 3.3: **Longer sampling: bistability.** (**A**) Population-summed activity $S(t)$ obtained via the Glauber dynamics, as in Figure 3.2A, but with longer sampling: 5×10^7 timesteps. The dashed grey line marks the end of the previous sampling of Figure 3.2A. (**B**) Population-summed activities $S(t)$ obtained from several instances of the Glauber dynamics. Each instance starts with a different initial population activity $\boldsymbol{s}(0)$, i.e. having different initial population sums $S(0)$, and are represented by different red shaded colors, from $S(0) = 0$ (light red) to $S(0) = N$ (dark red). Note that there are two convergence levels - one at 100 and of approx 15.

The second metastable regime has important implications and causes quite a few issues:

- Our Boltzmann learning had actually not yet converged to the final Lagrange multipliers. This can be seen from the different time-averages of the single and coupled activities obtained if sampled long enough to allow the exploration of both metastable regimes.

- Therefore the Lagrange multipliers shown in Figure 3.2D are not correct, because they were obtained before convergence. Hence *the probability distribution obtained from the initial Boltzmann learning is not the true pairwise maximum-entropy distribution.*

- For practical purposes the Glauber dynamics is *non-ergodic.* In other words to sample the probability distribution around both modes and estimate their relative heights, we would need to observe many jumps between the two metastable regimes. The time required to observe one such jump seems to be larger than 5×10^7 timesteps (we did not wait for longer), which is impractically long: *thus we cannot find the true pairwise distribution* within reasonable times.

- The Sessak-Monasson [300, 301] solution can be shown to be very close to the erroneous Boltzmann-learning.

The reason why the initial result shown in Figure 3.2 seemed self-consistent is that the sampling phase was too brief compared to the time needed to explore the full distribution. For the latter the required time is so long that the dynamics is *non-ergodic* for computational purposes. This non-ergodicity effectively truncates the sampling at states $\boldsymbol{s}$ for which $\bar{s} \lesssim \theta$, where θ is the population-averaged activity at the trough between the two metastable regimes. In other words, the Lagrange multipliers $\boldsymbol{\mu}, \boldsymbol{\Lambda}$ that we found belong to the "truncated" distribution

$$P_{\mathrm{t}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\theta) \propto \begin{cases} \exp(\sum_i \mu_i s_i + \sum_{i<j} \Lambda_{ij} s_i s_j), & \bar{s} \leqslant \theta, \\ 0, & \bar{s} > \theta, \end{cases} \tag{3.16}$$

which reproduces the experimental time averages for the single and coupled activities:$\mathrm{E}_{\mathrm{t}}(s_i) = m_i, \qquad \mathrm{E}_{\mathrm{t}}(s_i s_j) = g_{ij}$. Using these Lagrange multipliers in the true maximum-entropy distribution (Eq. 3.3) would, however, not reproduce these experimental time averages.

Everything is self-consistent as long as we use the truncated distribution P_{t}, but this is not the pairwise distribution P_{p}. This remark will be important later on.

Now the question is whether the correct maximum-entropy distribution is also bimodal, or the Boltzmann learning simply encountered a bimodal distribution during its search of the correct one in the space of probabilities.

We make an educated guess by examining the analytically tractable reduced maximum-entropy model P_{r}, (Eq. 3.9). Using the population-averaged single and coupled activities as constraints, $\mathrm{E}_{\mathrm{r}}(\overline{s_i}) = \bar{m}$ and $\mathrm{E}_{\mathrm{r}}(\overline{s_i s_j}) = \bar{g}$ from (Eq. 3.15), we numerically find the Lagrange multipliers of the reduced model:

$$\mu_{\mathrm{r}} \approx -3.259, \qquad \Lambda_{\mathrm{r}} \approx 0.03859. \tag{3.17}$$

Note that in this case there is no sampling involved – the distribution can be calculated analytically – so the values (Eq. 3.17) are correct within the numerical precision of the maximization procedure (interior-point method [273, chap. 10]). The values of the expected single and couple activities, re-obtained by explicit summation (not sampling) from the corresponding reduced maximum-entropy distribution, agree with the values (Eq. 3.15) to seven significant figures.

The resulting reduced maximum-entropy distribution for the population-summed activity, $P_{\mathrm{r}}(S|\,\mu_{\mathrm{r}}, \Lambda_{\mathrm{r}})$, is shown in Figure 3.4A, together with the experimental time-frequency distribution of our dataIts corresponding Glauber dynamics with two metastable regimes is shown in Figure 3.4B. It shows a second maximum at roughly 90% activity. An exact analysis of small-population cases, and an analysis of large-population cases with a maximum-entropy model constrained by the population variance of the second moments

(not shown here), corresponding to constraining $\mathrm{E}(S)$, $\mathrm{E}(S^2)$, $\mathrm{E}\left(\frac{S\,(S-1)}{N\,(N-1)} - \left(\frac{S\,(S-1)}{N\,(N-1)}\right)^2\right)$, show that if a reduced maximum-entropy model is bimodal, the full inhomogeneous model is also bimodal, with a heightened second mode shifted towards lower activities with respect to the reduced model.

The bimodality encountered in the Boltzman learning, the bimodality of the reduced maximum-entropy model, and the bimodality of the full maximum-entropy model for small populations are evidence that the correct, full pairwise maximum-entropy distribution for our data should be bimodal.

We will now propose a solution to eliminate the bimodality. The basic idea behind can be grasped after having first presented an intuitive explanation of how the bimodality arises.

3.3.2 Intuitive understanding of the bimodality: Glauber dynamics and mean-field picture

From the point of view of a network with couplings $\boldsymbol{\Lambda}$ and biases $\boldsymbol{\mu}$ whose evolution is described by a Glauber dynamics, the bimodality and associated bistability appear because the couplings $\boldsymbol{\Lambda}$ are on average positive, thus making the network dominantly excitatory. This positivity of the couplings appears simply because the average correlation $\bar{c}$ between the neurons is positive (Figure 3.1D).

A possible mechanism to suppress the second peak of high activity is by effective negative feedback: Such a negative input must be activated if the sum of the active units is high, and in turn must inhibit them, much in the same way as inhibition stabilizes the low activity state in neuronal networks [10, 347]. As a consequence, the couplings received by this hypothetical inhibitory input must have a positive sign, while its projections back are negative (illustrated in Figure 3.7A). This asymmetry, however, is in contradiction to the symmetry required for the couplings of a Glauber dynamics. Thus, self-regulary feedback loops, possible in networks with asymmetric couplings, are impossible in this case.

A naive mean-field analysis also confirms this. In such an approximation we imagine that each neuron is coupled to a field that is representing the mean activities of all other neurons [153][252, ch. 4][32, ch. 6] (from the point of view of entropy maximization, we are replacing the maximum-entropy distribution with one representing independent activities, having minimal Kullback-Leibler divergence from the original one [9, 258, 331, chs 2, 16, 17]). Given the couplings $\boldsymbol{\Lambda}$ and biases $\boldsymbol{\mu}$, the mean activities $\boldsymbol{m}$ must satisfy N self-consistency equations

$$\frac{1}{2} + \frac{1}{2}\tanh\left(\frac{1}{2}\sum_{j}^{j\neq i}\Lambda_{ij}m_j + \frac{1}{2}\mu_i\right) = m_i. \tag{3.18}$$

In the homogeneous case they reduce to the equation $\tanh[(N-1)\Lambda_{\mathrm{r}}\bar{m}+\mu_{\mathrm{r}}]=\bar{m}$ and correspond to the intersection of two functions of $\bar{m}$:the line $\bar{m}\mapsto\bar{m}$, and the curve $\bar{m}\mapsto\tanh[(N-1)\Lambda_{\mathrm{r}}\bar{m}+\mu_{\mathrm{r}}]$ that depends parametrically on $(\mu_{\mathrm{r}},\Lambda_{\mathrm{r}})$. See Figure 3.4D: for the Lagrange multipliers of our data, these curves intersect at two different values of $\bar{m}$, meaning that there are two solutions to the self-consistency equation, corresponding to two different mean activities. These approximately correspond to the maxima of the probability distribution for the population average in Figure 3.4A.

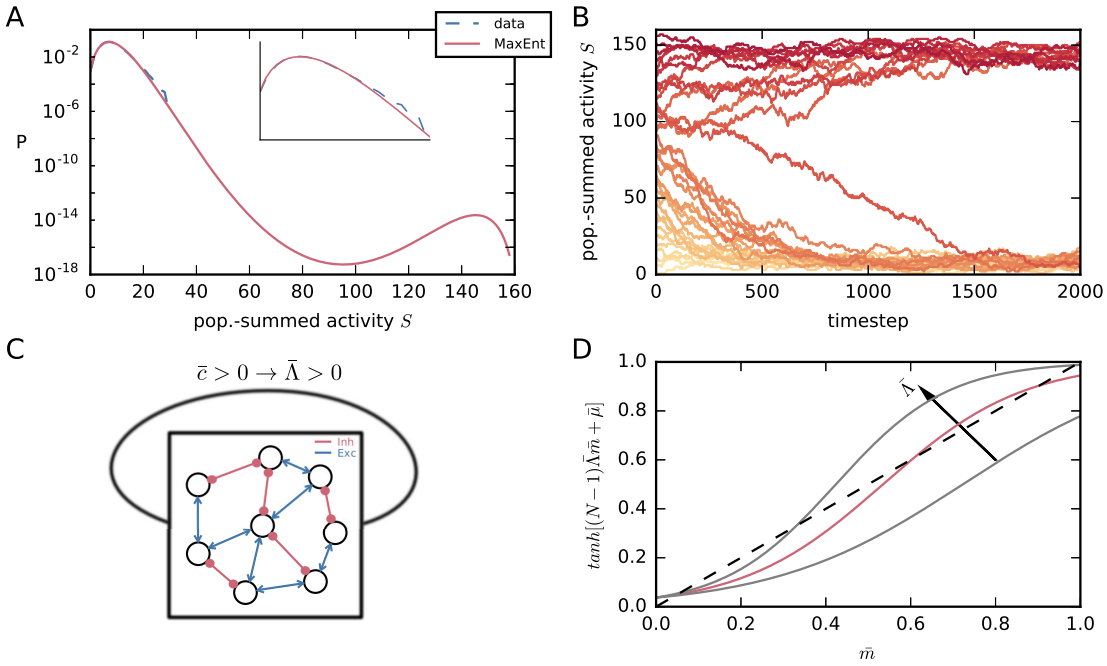

Figure 3.4: **Reduced maximum-entropy model and mean-field picture.** (**A**) Red, solid: Probability distribution for the population-summed activity, $P_{\mathrm{r}}(S)$ given by the reduced model for our dataset (Eq. 3.15); note the two probability maxima. Blue, dashed: empirical distribution of the population-summed activity from our dataset. (**B**) Population-summed activities $S(t)$ obtained from several instances of the Glauber dynamics associated with the reduced model, with homogeneous couplings, $\Lambda_{ij}=\Lambda_{\mathrm{r}}$, and biases, $\mu_i=\mu_{\mathrm{r}}$, of (Eq. 3.17). As in Figure 3.3, each instance starts with a different initial population activity $\boldsymbol{s}(0)$, having different initial population sum $S(0)$, and is represented by a different red shaded curve, from $S(0)=0$ (light red) to $S(0)=N$ (dark red). (**C**) Illustration of a self-coupled symmetric network that is self-excitatory on average. Arrow-headed blue lines ($\to$) represent excitatory couplings; circle-headed red lines ($-\circ$) represent inhibitory couplings. (**D**) Self-consistency solution of the naive mean-field equation, illustrated for different Λ_{r}. Larger Λ_{r} causes two additional intersections, corresponding to one additional unstable and one additional stable solution. The red curve corresponds to the Λ_{r} calculated from our experimental data (Eq. 3.17).

3.3.3 Bistability ranges and population size

Now we aim to understand what are the parameter regimes leading to bimodality, and whether or not the bimodality will disappear for larger neuronal populations or become even more prominent? Further, is the appearance of bimodality peculiar to our experimental dataset, or can it be also expected in other experimental datasets of neuronal activities? We aim to answer these two questions in the following two sections to understand whether or not this is a general problem.

We again make an educated guess using the reduced maximum-entropy model and its distribution for the population average $\bar{s}$ $P_{\mathrm{r}}(\bar{s}|\,\mu_{\mathrm{r}},\varLambda_{\mathrm{r}})$, (Eq. 3.11). An elementary study of the convexity properties (second derivative) of the distribution shows that it can have one minimum in the interior, $0<\bar{s}<1$, or none, depending on the values of the parameters $(\mu_{\mathrm{r}},\varLambda_{\mathrm{r}})$. The distribution has two probability maxima if it has one such minimum for some value $\bar{s}_{\mathrm{m}}$, $0<\bar{s}_{\mathrm{m}}<1$. The conditions for this are

$$\left.\frac{\mathrm{d}P_{\mathrm{r}}(\bar{s}|\,\mu_{\mathrm{r}},\varLambda_{\mathrm{r}})}{\mathrm{d}\bar{s}}\right|_{\bar{s}=\bar{s}_{\mathrm{m}}}=0,\qquad \left.\frac{\mathrm{d}^2P_{\mathrm{r}}(\bar{s}|\,\mu_{\mathrm{r}},\varLambda_{\mathrm{r}})}{\mathrm{d}\bar{s}^2}\right|_{\bar{s}=\bar{s}_{\mathrm{m}}}>0,\quad 0<\bar{s}_{\mathrm{m}}<1, \tag{3.19}$$

These conditions can be solved analytically and give the critical ranges of the multipliers $(\mu_{\mathrm{r}},\varLambda_{\mathrm{r}})$ for which bimodality occurs, parametrically in $(\bar{s}_{\mathrm{m}},\varLambda_{\mathrm{r}})$:

$$\begin{cases} 0<\bar{s}_{\mathrm{m}}<1, \\ \varLambda_{\mathrm{r}}>\Psi'[1+(1-\bar{s}_{\mathrm{m}})N]+\Psi'(1+\bar{s}_{\mathrm{m}}N), \\ \mu_{\mathrm{r}}(\bar{s}_{\mathrm{m}},\varLambda_{\mathrm{r}})=\varLambda_{\mathrm{r}}/2-\bar{s}_{\mathrm{m}}N\varLambda_{\mathrm{r}}-\Psi[1+(1-\bar{s}_{\mathrm{m}})N]+\Psi(1+\bar{s}_{\mathrm{m}}N), \end{cases} \tag{3.20}$$

where $\Psi(x)\coloneqq \mathrm{d}\ln\Gamma(x)/\mathrm{d}x$ and Γ is the Gamma function [3, ch. 6][255, chs 43, 44]. We then express the population-averaged single neuron activities $\mathrm{E}_{\mathrm{r}}(\bar{s})$ and the Pearson correlation $\overline{\rho}$, typically used in the literature, in terms of $(\mu_{\mathrm{r}},\varLambda_{\mathrm{r}})$ using the definitions (Eq. 3.14) and the probability (Eq. 3.11). Finally we obtain the bimodality range for $\big(\mathrm{E}_{\mathrm{r}}(\bar{s}),\overline{\rho}\big)$, parametrically in $(\bar{s}_{\mathrm{m}},\varLambda_{\mathrm{r}})$ within the bounds (Eq. 3.20) numerically.

The results are shown in Figure 3.5A for various values of the number of neurons N. For each N, a curve is associated for the Pearson correlation coefficient as a function of the expected activity. The values of $\big(\mathrm{E}_{\mathrm{r}}(\bar{s}),\overline{\rho}\big)$ above such curves yield a bimodal distribution in the homogeneous case. Most important, Figure 3.5A shows that the maximum-entropy distribution will be bimodal for larger ranges of mean activities and correlations, as the population size N increases. Empirical population-averaged quantities e.g. $\big(\mathrm{E}_{\mathrm{r}}(\bar{s}),\overline{\rho}\big)$, on the other hand, should not change with population size if they are sampled from a biologically homogeneous neural population (see Appendix 5.2.4 for an extensive explanation and the mathematical proof; see also Figure 5.3). This means that even if maximum-entropy does not predict a bimodal distribution for the measured activities and correlations of

a particular small sample, it will predict a bimodal distribution for a larger sample in a similar experimental setup.

Figure 3.5B displays the probability distribution of the population-averaged activity for different values of N. When $N \lesssim 150$ the distribution has only one maximum at low activity, $\bar{s} \approx 0.0497$, and when $N \gtrsim 150$ a second probability maximum at high activity, $\bar{s} \approx 0.9502$, appears. The probability at this second maximum increases sharply until $N \approx 200$ and thereafter maintains an approximately stable value, roughly 6000 times smaller than the low-activity maximum. The minimum between the two modes becomes deeper and deeper as we increase N above 200.

As mentioned in the previous section, exact studies with small samples and studies with large samples and a different reduced model, which takes into account the population-variance of the second moments, indicate that the high-activity maximum in the inhomogeneous case is larger (roughly 2000 times smaller than the low-activity one when $N = 1000$) and shifted towards lower activities ($\bar{s} \approx 0.25$ when $N = 1000$).

This can also be seen by adding a Gaussian jitter to the multipliers of the reduced case $\mu_i = \mu_\mathrm{r}$, $\Lambda_{ij} = \Lambda_\mathrm{r}$, thereby making the model inhomogeneous. The results for small and large jitter are shown in Figure 3.5C–D, respectively. The basin of attraction of the second metastable regime is shifted to lower activities, and transitions between the two metastable regimes become more likely for larger jitters. This means that inhomogeneity makes the minimum in between the two modes shallower.

The population-averaged activity and Pearson correlation of our data (violet "3 ms" point in Figure 3.5A) fall within the bimodality range.

3.3.4 Relevance of the bimodality problem for massively parallel data

The important question is whether our dataset is a typical representative leading to this bimodality problem, or an outlier. To answer this question we take as reference the data summarized in Table 1 of Cohen & Kohn [49], which reports firing rates and spike-count correlations r_SC. The reported firing rates correspond to population-averaged activities $\bar{m}$ ranging between 0.02 and 0.25, if we use 3 ms time bins, and thus are close to the values we found in out data. For being able to compare the spike-count correlations we need to estimate our Pearson correlation ρ from their spike-count correlation r_SC. Both are particular cases of the "cross-correlogram metric" r_CCG introduced by Bair et al. [15, App. A]:

$$\begin{aligned} r_{\mathrm{CCG}\,ij}(\tau) &\coloneqq \frac{\mathrm{E}(\nu_i(\tau)\,\nu_j(\tau)) - \mathrm{E}(\nu_i(\tau))\mathrm{E}(\nu_j(\tau))}{\sqrt{\left[\mathrm{E}(\nu_i(\tau)^2) - \mathrm{E}(\nu_i(\tau))^2\right]\left[\mathrm{E}(\nu_j(\tau)^2) - \mathrm{E}(\nu_j(\tau))^2\right]}}, \\ \text{with} \quad \nu_i(\tau) &\coloneqq \sum_{t=1}^{\tau} s_i(t), \end{aligned} \tag{3.21}$$

i.e. $\nu_i(\tau)$ is the number of spikes of neuron i during the (real-)time window $\tau\Delta$. This metric also equals the area between times $-\tau\Delta$ and $\tau\Delta$ under the cross-correlogram of neurons i and j (stationarity is assumed). The spike count correlation r_{SC} corresponds to $\tau = n \equiv T/\Delta$, and our Pearson correlation ρ to $\tau = 1$. Several studies [15, 16, 200, 241, 321] report either measured values of $r_{\mathrm{CCG}}(\tau)$ for different windows τ, or measured cross-correlograms. From their analysis we can approximately say that $\rho \lesssim r_{\mathrm{SC}}/20$, so we take $\overline{\rho} = \overline{r_{\mathrm{SC}}}/20$ as a safest-case value (i.e. as far away from bimodality as possible).

Under these approximations the largest part of the values summarized by Cohen & Kohn fall in the bimodality regions of Figure 3.5A if $N = 250$, and almost all of them if $N = 500$; see Figure 3.6. These data points have only an indicative value but suggest that our dataset is not an outlier for the bimodality problem. If those data had been recorded from a population of 500 neurons, they would have yielded a bimodal pairwise maximum-entropy model. Thus, the more neurons we are able to record, the more likely the bimodality occurs. Thus, the bimodality problem and its consequences need to be taken seriously. So our next question arises: Is there any way to eliminate the bimodality problem?

3.3.5 Eliminating the bimodality: an inhibited maximum-entropy model and Glauber dynamics

We briefly summarize our results so far and the reason why a maximum-entropy model yielding a bimodal distribution in the population-summed activity is problematic:

- For commonly observed statistics of neuronal data, the maximum-entropy model predicts the presence of two sharply distinct modes (bimodality). This prediction seems unrealistic in view of present neuroscientific data.

- As N increases, the second mode becomes more pronounced, and the minimum between the modes shallower: above a particular population size, the bimodality cannot be dismissed as a small mathematical quirk.

- The Boltzmann-learning procedure based on Glauber dynamics becomes practically non-ergodic and the Lagrange multipliers of the model are difficult or impossible to find.

- The Glauber dynamics based on the pairwise model jumps between two metastable regimes and cannot be used to generate realistic surrogate data.

- Finally, the fact that the position and height of the second mode depend on N (in the inhomogeneous case) goes against basic statistical expectations. If we consider the N neurons to be a sample, chosen in an unsystematic way, of a larger population, then the maxima in our probability assignments for the population averages of the sample

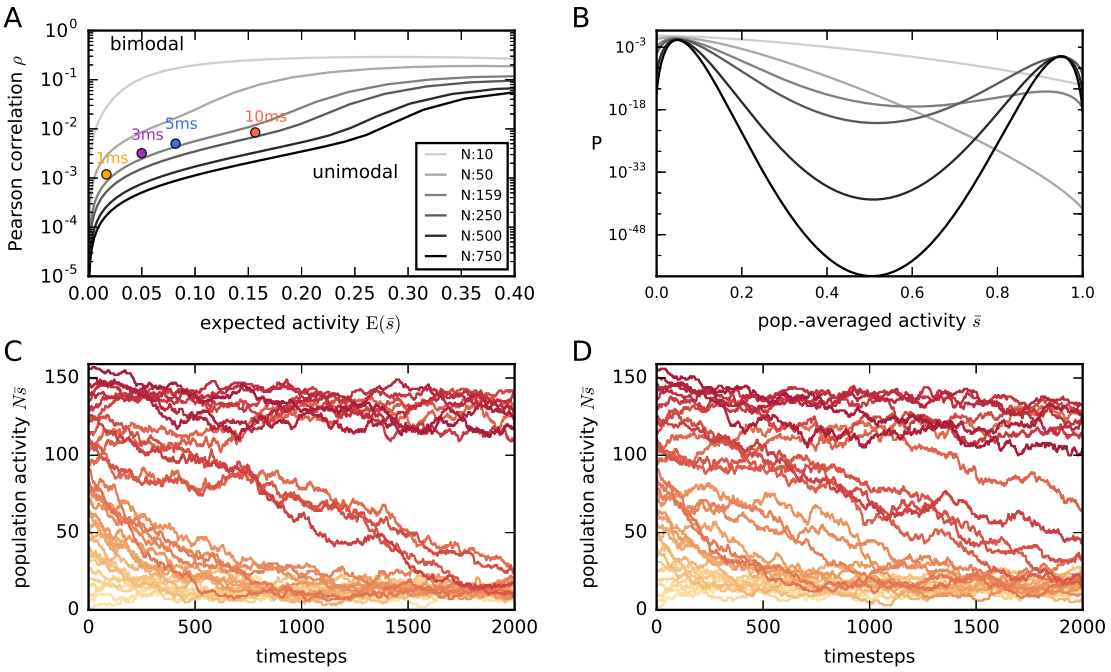

Figure 3.5: **Bimodality ranges for the reduced model and effects of inhomogeneity.** (**A**) The reduced maximum-entropy model (Eq. 3.9) yields a distribution $P_{\mathrm{r}}(\bar{s})$ that is either unimodal or bimodal, depending on the number of neurons N and the values of the experimental constraints $\big(\mathrm{E}_{\mathrm{r}}(\bar{s}), \overline{\rho}\big)$. Each curve in the plot corresponds to a particular N (see legend) and separates the values $\big(\mathrm{E}_{\mathrm{r}}(\bar{s}), \overline{\rho}\big)$ yielding a unimodal distribution (below the curve) from those yielding a bimodal one (above the curve). The curves are symmetric with respect to $\mathrm{E}_{\mathrm{r}}(\bar{s}) = 0.5$ (ranges $\mathrm{E}_{\mathrm{r}}(\bar{s}) > 0.4$ not shown). Note how the range of constraints yielding bimodality increases with N. Coloured dots show the experimental constraints for our dataset, for different time-binnings with widths $\Delta = 1\,\mathrm{ms}$, $\Delta = 3\,\mathrm{ms}$, $\Delta = 5\,\mathrm{ms}$, $\Delta = 10\,\mathrm{ms}$. (**B**) Probability distributions of the reduced model for the population-summed activity, $P_{\mathrm{r}}(S|\,N)$, obtained keeping the constraints (Eq. 3.15) fixed and using different N (same legend as panel A). (**C**) Population-summed activities $S(t)$ from several instances of Glauber dynamics, all with the same normally-distributed couplings Λ_{ij} and biases μ_i, with means as in (Eq. 3.17) and Figure 3.4B, and standard deviations $\sigma(\Lambda_{ij}) = 0.009$, $\sigma(\mu_i) = 0.8$. Each instance starts with a different initial population activity $\boldsymbol{s}(0)$, having different initial population sum $S(0)$, and is represented by a different red shade, from $S(0) = 0$ (light red) to $S(0) = N$ (dark red). Note how the basins of attraction of the two metastable regimes are wider than in the homogeneous case of Figure 3.4B. (**D**) The same as panel C, but with larger standard deviations $\sigma(\Lambda_{ij}) = 0.012$, $\sigma(\mu_i) = 1.08$; the jumps between the two metastable regimes become more frequent than in Figure 3.4B, indicating that the minimum between the modes becomes more shallow with increasing inhomogeneity.

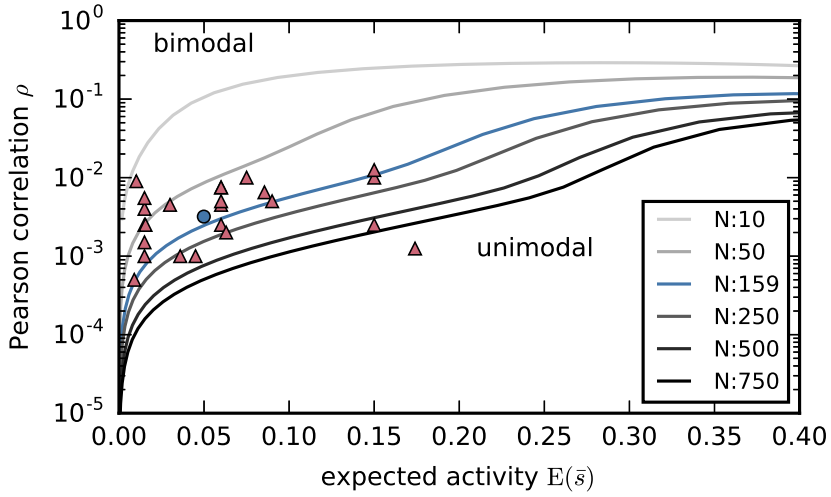

Figure 3.6: **Bimodality for experimental data from neuroscientific literature.** Mean activities and correlations $(\mathrm{E}_{\mathrm{r}}(\bar{s}), \overline{\rho})$ inferred from experimental data reported in Cohen & Kohn [49, Table 1], plotted upon the curves separating bimodal from unimodal maximum-entropy distributions of Figure 3.5A. The plot suggests that typical experimental neural recordings of 250 neurons and above are likely to lead to bimodal maximum-entropy pairwise distributions.

and of the larger population should roughly coincide (the former being obtained from the latter by convolution with a hypergeometric distribution).

We now propose a way to eliminate the bimodality and the above problems. Let us re-examine what happens with the Glauber dynamics first.

3.3.6 Importance of inhibition in neural networks: modified Glauber dynamics

As mentioned in Section 3.3.2, from the viewpoint of the Glauber dynamics the jumps to high activities happen because the couplings $\boldsymbol{\Lambda}$ are positive on average, making the network effectively excitatory.

The positivity of the couplings corresponds to an experimentally observed positive average correlation (Figure 3.1D). However, their symmetry is an assumed mathematical feature of the pairwise model. Thus, we wonder if we can break this symmetry to avoid the bimodality of the probability distribution e.g. by adding an asymmetric inhibitory feedback to the Glauber dynamics.

We approach the problem in the following way. By connecting all N neurons to a single inhibitory neuron that instantaneously activates whenever their average activity exceeds a threshold $\theta \in \{1/N, 2/N, \dots, (N-1)/N\}$ (the cases $\theta = 0$ or 1 are trivial). Upon activation, the inhibitory neuron inhibits (because $\Lambda_{\mathrm{I}} < 0$) all other N neurons

(see Figure 3.7A). The algorithm for this "inhibited" Glauber dynamics is explained in Appendix 5.2.1.

The results from simulations of the inhibited Glauber dynamics are shown in Figure 3.7; in all cases we set the inhibitory coupling to $\Lambda_{\mathrm{I}} = -24.7$ and the inhibition threshold to $\theta = 0.3$. It can clearly be seen that in contrast to the symetrical and on average excitatory networks ($\Lambda_{ij} = \Lambda_{\mathrm{r}}$ and $\mu_i = \mu_{\mathrm{r}}$, (Eq. 3.17)) the additional inhibitory neuron has eliminated the bistability, leaving only the stable low-activity regime . The homogeneous ($\Lambda_{ij} = \Lambda_{\mathrm{r}}$ and $\mu_i = \mu_{\mathrm{r}}$, (Eq. 3.17)) as well as the inhomogeneous (where Λ_{ij} and μ_i are normally distributed as in Figure 3.5C) model has a unimodal probability distribution and the activity converges over time to one activity solution. Furthermore, in the case of experimental data with the inhomogeneous couplings and biases (distributed as in Figure 3.2D) that caused the quasi-non-ergodic behaviour in our previous Boltzmann learning results, Figure 3.3, the addition of the inhibitory neuron (again with $\Lambda_{\mathrm{I}} = -24.7$, $\theta = 0.3$) eliminates the second metastable state: see Figure 3.7E.

In summary, the asymmetric coupling of an additional inhibitory neuron clearly eliminates the bistability of the Glauber dynamics. This works for any network size N with an appropriate choice of the inhibitory coupling $\Lambda_{\mathrm{I}} < 0$ and threshold θ. We discuss these two parameters in the next section.

We now show analytically how this idea also eliminates our original problem, i.e. the bimodality of the pairwise maximum-entropy model.

3.3.7 Inhibited maximum-entropy (IME) model

The pairwise maximum-entropy model is the stationary distribution of the Glauber dynamics with symmetric couplings. We have now modified the latter in an asymmetric way. We are wondering if the resulting distribution of the inhibited Glauber dynamics (Eq. 3.7) *is still a maximum-entropy model.* We consider

$$\begin{aligned} P_{\mathrm{i}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta) &= \frac{1}{Z_{\mathrm{i}}(\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta)}\times \\ &\quad \exp\big[\sum_i \mu_i s_i + \sum_{i>j} \Lambda_{ij} s_i s_j + \Lambda_{\mathrm{I}} N\, G(\bar{s}-\theta)\big], \\ Z_{\mathrm{i}}(\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta) &\coloneqq \sum_{\boldsymbol{s}} \exp\big[\sum_i \mu_i s_i + \sum_{i>j} \Lambda_{ij} s_i s_j + \Lambda_{\mathrm{I}} N\, G(\bar{s}-\theta)\big], \\ G(\bar{s}-\theta) &\coloneqq (\bar{s}-\theta)\,\mathrm{H}(\bar{s}-\theta), \end{aligned} \tag{3.22}$$

where Λ_{I} is the (negative, in our case) coupling strength from the inhibitory neuron to the other neurons, θ is the activation threshold of the inhibitory neuron, and H is the Heaviside step function. We call (Eq. 3.22) the *inhibited pairwise maximum-entropy model.*

The function $G(\bar{s}-\theta)$ (plotted in Figure 3.8 together with its exponential) can also be written as a linear combination of population-averaged K-tuple activities, $s_{i_1} s_{i_2} \cdots s_{i_K}$, for

K equal to $N\theta$ and larger (we leave the proof of this as a classic "exercise for the reader"):

$$N\,G(\bar{s}-\theta) = \sum_{K=N\theta}^{N} \binom{-N\theta}{-K+1} \left(\sum_{i_1<i_2<\cdots<i_K} s_{i_1} s_{i_2} \cdots s_{i_K} \right), \tag{3.23}$$

the linear coefficients being binomial coefficient functions [91], which have alternating signs. For example, if $N = 5$ and $\theta = 3/5$,

$$\begin{aligned} N\,G(\bar{s}-\theta) = \\ (s_2s_3s_4s_5 + s_1s_3s_4s_5 + s_1s_2s_4s_5 + s_1s_2s_3s_5 + s_1s_2s_3s_4) - \\ 3\,s_1s_2s_3s_4s_5. \end{aligned} \tag{3.24}$$

(This function differs from the additional function appearing in maximum-entropy model by Tkačik et al. [335–337], which consists in $N+1$ constraints enforcing the observed

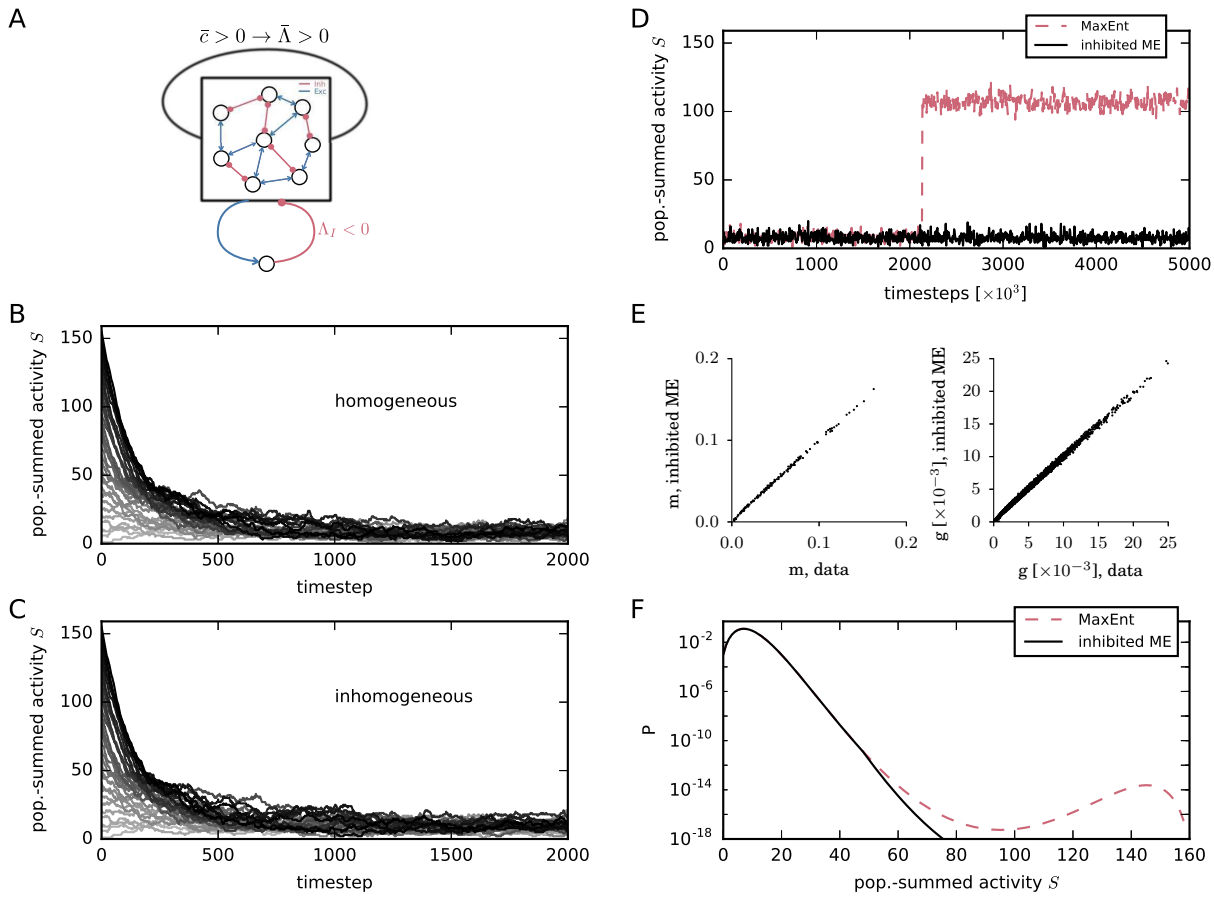

Figure 3.7: *(Caption on the next page)*

Figure 3.7 *(previous page)*: **Asymmetric inhibition and elimination of bimodality and non-ergodicity.** (**A**) Illustration of self-coupled network with additional asymmetric inhibitory feedback. Each neuron receives inhibitory input $\Lambda_{\mathrm{I}} < 0$ from the additional neuron whenever the population-average $\bar{s}$ becomes greater than the inhibition threshold θ. (**B**) Population-summed activities $S(t)$ from several instances of the inhibited Glauber dynamics, with $\Lambda_{\mathrm{I}} = -25$, $\theta = 0.3$, and same homogeneous $\Lambda_{ij} = \Lambda_{\mathrm{r}}$, $\mu_i = \mu_{\mathrm{r}}$ of (Eq. 3.17), as used for Figure 3.4B. Each instance starts with a different initial population activity $\boldsymbol{s}(0)$, having different initial population sum $S(0)$, and is represented by a different grey shade, from $S(0) = 0$ (light grey) to $S(0) = N$ (black). Note the disappearance, thanks to inhibition, of the bistability that was evident in the "unhinibited" case of Figure 3.4B. (**C**) Analogous to panel B, with $\Lambda_{\mathrm{I}} = -25$, $\theta = 0.3$, but inhomogeneous normally distributed couplings and biases as in the unhinibited case of Figure 3.5C. Note again the disappearance, thanks to inhibition, of the bistability that was evident in the activities $S(t)$ of that figure. (**D**) Comparison of a longer (5×10^6 timesteps) Glauber sampling with couplings and biases of Figure 3.2D obtained from our first Boltzmann learning, and inhibited-Glauber sampling with same couplings and biases and $\Lambda_{\mathrm{I}} = -25$, $\theta = 0.3$. The comparison confirms that inhibition eliminates the second metastable regime and makes the Glauber dynamics ergodic. (**E**) Time averages m_i and g_{ij} obtained from Boltzmann learning for the inhibited model P_{i}, versus experimental ones. (**F**) Probability distribution of the population-summed activity $P_{\mathrm{i}}(S)$ given by the inhibited model (Eq. 3.22) for our dataset (Eq. 3.15), compared with the one previously given by the reduced model $P_{\mathrm{r}}(S)$, Figure 3.4A. Asymmetric inhibition, expressed by the reference prior (Eq. 3.25), has eliminated the second mode.

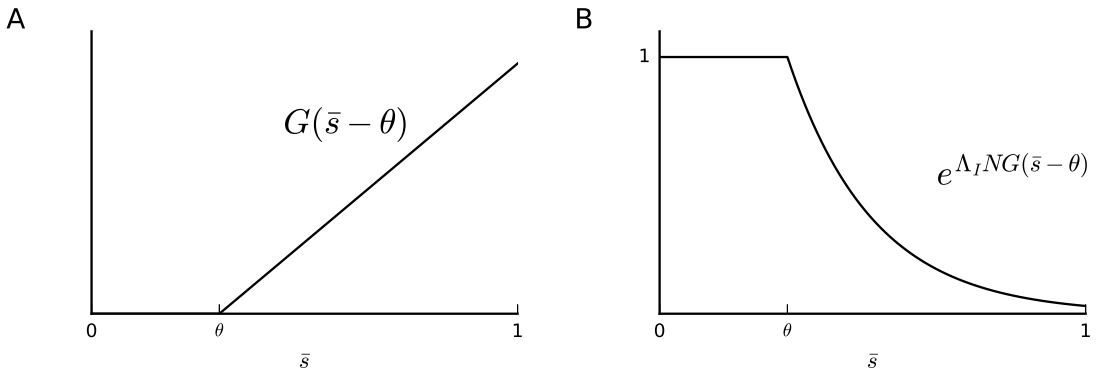

Figure 3.8: **Reference prior.** The function $G(\bar{s} - \theta)$ and its exponential

population-average distribution. For reasons discussed at the end of Section 3.2.1, the use of all those constraints may not be justified or meaningful.)

The stationarity of the distribution $P_{\mathrm{i}}(\boldsymbol{s})$ under the inhibited Glauber dynamics is proved in the Materials and Methods section. This distribution is a maximum-entropy model in two different ways:

1. As an application of the minimum-relative-entropy (minimum-discrimination-information) principle [5, 43, 48, 51, 143–145, 165, 176, 207–211, 277], with the pairwise constraints (Eq. 3.5), with respect to the reference (or prior) probability distribution

$$P_0(\boldsymbol{s}|\, \Lambda_{\mathrm{I}}, \theta) \propto \exp\bigl[\Lambda_{\mathrm{I}} N\, G(\bar{s} - \theta)\bigr], \qquad (N\theta \in \{1, 2, \dots, N-1\}). \tag{3.25}$$

   This distribution assigns decreasing probabilities to states with average activities above θ; see Figure 3.8B. This probability can be interpreted as arising from a more detailed model in which we know that external inhibitory units make activities above the threshold θ increasingly improbable. In this interpretation the parameters Λ_{I} and θ are chosen a priori.

2. As an application of the "bare" maximum-entropy principle, given the pairwise constraints (Eq. 3.5) and an additional constraint for the expectation of $N\, G(\bar{s} - \theta)$:

$$\begin{aligned} \mathrm{E}_{\mathrm{i}}(N\, G(\bar{s} - \theta)) &= \sum_{S=N\theta}^{N} (S - N\theta)\, P_{\mathrm{i}}(S) \\ &= \sum_{K=N\theta}^{N} \binom{-N\theta}{-K+1} \mathrm{E}_{\mathrm{i}}\left(\sum_{i_1 < i_2 < \dots < i_K} s_{i_1} s_{i_2} \cdots s_{i_K} \right). \end{aligned} \tag{3.26}$$

   This is a constraint of the "tail first moment", so to speak, of the probability for the population-averaged activity $P_{\mathrm{i}}(\bar{s})$: it determines whether the right tail of $P_{\mathrm{i}}(\bar{s})$ has a small ($\Lambda_{\mathrm{I}} < 0$) or heavy ($\Lambda_{\mathrm{I}} > 0$) probability. It can also be seen as a constraint on the $N\theta$th and higher moments, owing to (Eq. 3.23). In this interpretation the parameter Λ_{I} is the Lagrange multiplier associated with this constraint, hence it is determined by the data; the parameter θ is still chosen a priori. Note, however, that experimental data are likely to give a vanishing time average of $N\, G(\bar{s} - \theta)$, so that $\Lambda_{\mathrm{I}} = -\infty$. This interpretation has therefore to be used with care, for the reasons discussed at Section 3.2.1.

Several features of the inhibited maximum-entropy model (Eq. 3.22) are worth remarking upon:

1. The inhibited distribution P_{i} includes the pairwise one P_{p}, (Eq. 3.3), as the particular case $\Lambda_{\mathrm{I}} = 0$ (obviously, as this is equivalent to removing the inhibitory neuron).

2. Pairwise and inhibited distributions P_{p} and P_{i} having same Lagrange multipliers $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ are *equal* if restricted to states with population-averaged activity below the

threshold θ, because $G(\bar{s}-\theta)=0$ if $\bar{s}\leqslant\theta$:

$$P_{\mathrm{p}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\bar{s}\leqslant\theta)=P_{\mathrm{i}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta,\bar{s}\leqslant\theta). \tag{3.27}$$

Said otherwise, the pairwise and inhibited distributions have the same shape for $\bar{s}\leqslant\theta$, modulo rescaling by a constant factor:

$$P_{\mathrm{i}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta)=P_{\mathrm{p}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda})\times\frac{Z_{\mathrm{p}}(\boldsymbol{\mu},\boldsymbol{\Lambda})}{Z_{\mathrm{i}}(\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta)},\quad \bar{s}\leqslant\theta. \tag{3.28}$$

3. For states with average activity $\bar{s}$ above the threshold θ, the inhibited model is a "squashed" version of the pairwise one when $\Lambda_{\mathrm{I}}<0$:

$$P_{\mathrm{i}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta)\propto P_{\mathrm{p}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda})\times\exp\bigl[N\Lambda_{\mathrm{I}}\,(\bar{s}-\theta)\bigr],\quad \bar{s}>\theta. \tag{3.29}$$

4. If $\Lambda_{\mathrm{I}}\neq 0$, then inhibited and pairwise models with the same Lagrange multipliers $(\boldsymbol{\mu},\boldsymbol{\Lambda})$ have *different* expectations for single and coupled activities:

$$\begin{aligned}\mathrm{E}_{\mathrm{i}}(s_i|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta)&\neq\mathrm{E}_{\mathrm{p}}(s_i|\,\boldsymbol{\mu},\boldsymbol{\Lambda}),\\ \mathrm{E}_{\mathrm{i}}(s_is_j|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta)&\neq\mathrm{E}_{\mathrm{p}}(s_is_j|\,\boldsymbol{\mu},\boldsymbol{\Lambda}),\end{aligned} \tag{3.30}$$

and obviously also different covariances and correlations.

Remarks 2 and 3 above imply that if the inhibitory coupling Λ_{I} is negative and very large, so that $\exp\bigl[N\Lambda_{\mathrm{I}}(\bar{s}-\theta)\bigr]\approx 0$ when $\bar{s}>\theta$, then *the inhibited maximum-entropy distribution P_{i} is approximately equal to the truncated distribution P_{t}* – the incorrect one (Eq. 3.16) obtained via Boltzmann learning – having the same multipliers $(\boldsymbol{\mu},\boldsymbol{\Lambda})$ and threshold θ:

$$P_{\mathrm{i}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta)\approx P_{\mathrm{t}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\theta)\quad\text{if }\Lambda_{\mathrm{I}}\ll -1, \tag{3.31}$$

(mathematically speaking we have pointwise convergence as $\Lambda_{\mathrm{I}}\to-\infty$), and their expectations are also approximately equal. This suggests a way to reinterpret and keep the results of our first Boltzmann-learning algorithm, Figure 3.2.

Boltzmann learning for the inhibited maximum-entropy model

Our first Boltzmann-learning calculation, with results shown in Figure 3.2, returned a distribution that reproduced the desired constraints $(\boldsymbol{m},\boldsymbol{g})$. But that distribution turned out to be not the true pairwise maximum-entropy one, but a truncated version of it P_{t}, (Eq. 3.16), owing to the bimodality of the true pairwise distribution and the resulting non-ergodicity.

Applying the inhibited pairwise maximum-entropy model $P_{\mathrm{i}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\Lambda_{\mathrm{I}},\theta)$ (with $\Lambda_{\mathrm{I}}\ll -1$ and $0.3\lesssim\theta\lesssim 0.5$) to our data, instead of the pairwise one $P_{\mathrm{p}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda})$, and

findingits Lagrange multipliers via Boltzmann learning, we obtain the same results as in Figure 3.2. This happens because the inhibitory neuron does not allow jumps to higher activities, unlike Figure 3.3A (see Figure 3.7D); and at the same time it does not influence the dynamics below $\bar{s} \approx 0.3$ (i.e. $S \approx 50$). The sampling phase of our Boltzmann learning thus becomes sufficient. We confirm this by applying Boltzmann-learning procedure (with the inhibited Glauber dynamics) to find the multipliers of the inhibited model with $\Lambda_{\mathrm{I}} = -25$, $\theta = 0.3$. The resulting multipliers are close to those in Figure 3.2D, and the constraints are satisfied, see Figure 3.7E.

3.4 Summary

In this chapter we provided evidence that the pairwise model, applied to experimental recordings, predicts a bimodal distribution for the population-averaged activity, and for some population sizes the second mode peaks at high activities, with 90% of the neuron population active within time-windows of few milliseconds. This bimodality has several undesirable consequences: 1. The presence of two modes is unrealistic in view of observed neuronal activity. 2. The prediction of a high-activity mode is unrealistic on neurobiological grounds. 3. Boltzmann learning becomes non-ergodic, hence the pairwise model found by this method is not the maximum entropy distribution; similarly, solving the inverse problem by common variants of mean-field approximations has the same problem. 4. The Glauber dynamics associated with the model is either unrealistically bistable, or does not reflect the distribution of the pairwise model. This bimodality is first demonstrated for an experimental dataset comprising activity recorded from 159 neurons in the motor cortex of macaque monkey. Using a reduced maximum-entropy model, evidence is then provided that this bimodality affects analysis of typical neural recordings of population sizes of a couple of hundreds or more neurons. As a way to eliminate the bimodality and its ensuing problems, the inhibited maximum entropy (IME) was presented, which – most important – has an associated pairwise Glauber dynamics. This model avoids bimodality thanks to a minimal asymmetric inhibition. It can be interpreted as a minimum-relative-entropy model with a particular prior, or as a maximum-entropy model with an additional constraint.

CHAPTER 4

Population unitary events (PUE) analysis

The following chapter is prepared as a manuscript to be submitted soon to a peer-reviewed journal. Preliminary results have been presented in Rostami et al (2015a) [288].

Author contributions: Under supervision of Junji Ito and Sonja Grün, the author designed the statistical test and proposed the method, implemented the method in Python, calibrated the method, and analyzed the simulated as well as the experimental data. Prof. Dr. Laura Sacerdote provided useful feedback on the design of the statistical test.

4.1 Introduction

Precise, coordinated spike timing within an ensemble of neurons has been hypothesized to be a potential mechanism behind cortical information processing [1, 74, 156]. Consequently, different statistical analysis tools have been proposed to investigate this hypothesis on the recordings from cortical neurons. Nevertheless, most of these tools, if applicable to MPST, exploit the assumption of independence among neurons to test the significance of the occurrence of specific spike patterns or synchrony. The independence assumption in particular becomes an issue when studying MPST because the results of the statistical test (with the null-hypothesis based on independence) are inadequate to answer questions such as: How much of the observed significance synchrony comes simply because the neurons are connected and therefore exhibit correlation? How much of this is task, stimulus related? Appropriate approaches to those questions call for statistical tools which can incorporate the null hypothesis beyond independence. In this chapter we propose a novel statistical method which on the one hand can handle MPST and on the other hand is able to incorporate a null hypothesis beyond independence to take into account the baseline correlation.

4.2 Methods

Let us consider a set of N spike trains $s_i(t) = \sum_j \delta(t - t_{i,j})$ on a common time axis, with neurons $i = 1, 2, \cdots, N$ and the spike times $\tau \leq t_{i,1} < t_{i,2} < \cdots \leq \tau + T$ of the i-th neuron in a finite interval $[\tau, \tau + T]$. Since we are interested in spike synchrony on a physiologically relevant time scale, we discretize the time axis into n_{bins} exclusive bins of identical size $\triangle$ (typically on the order of a few ms), and define a variable $d_i^k = \int_{k\Delta}^{(k+1)\Delta} s_i(\tau + t)dt$ which represents the number of spikes of the i-th neuron within the k-th bin. If more than one spikes are found in a bin, i.e. $d_i^k > 1$, d_i^k is set to 1 ('clipping'). After this clipping d_i^k represents a binary process.

The initial step of the analysis is to detect spike synchrony events of a specific order in a parallel set of binary processes d_i^k , $i = 1, \ldots, N$ representing parallel spike trains. Here "order" of synchrony refers to the number of spikes which constitutes a synchronous spike event. Concretely, we call a pairwise spike coincidence if two neurons have a spike in the same time bin, and name it a spike synchrony event of order 2, likewise, a triple-wise spike coincidence of order 3, a quadruple coincidence of order 4, etc. To count the number of spike synchrony events of a specific order, we first consider the population spike count C^k defined as the sum of d_i^k across the neurons in each time bin k: $C^k = \sum_{i=1}^N d_i^k$ (Figure 4.1A). We then count the number of spike synchrony events of a given order ξ_c, which is derived by the binomial coefficient C^k-choose-ξ_c for the k-th bin. The obtained counts of a certain order

ξ_c are summed across the time bins to yield the population coincidence count n_{ξ_c}defined as:

$$n_{\xi_c} = \sum_{k=1}^{n_{bins}} \left(\begin{array}{c} C^k \\ \xi_c \end{array} \right) = \sum_{k=1}^{n_{bins}} \frac{C^k!}{\xi_c!(C^k - \xi_c)!}. \tag{4.1}$$

The population coincidence count n_{ξ_c} represents the total number of ξ_c-tuple-wise spike synchrony events in the given data. For instance, as illustrated in Figure 4.1B, n_2 is the total number of pair-wise spike synchrony events (marked in blue), n_3 is the total number of triple-wise synchrony events (marked in red), and so on.

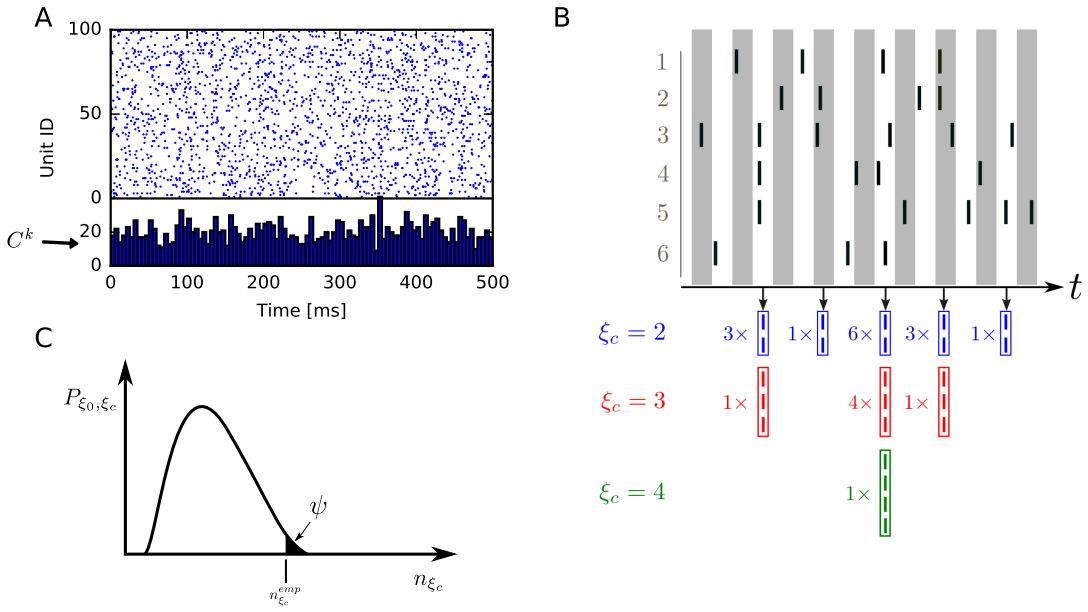

Figure 4.1: **Population coincidence count n_{ξ_c}, based on population count C^k.** (**A**) Raster plot of 100 simulated parallel independent Poisson spike trains with a stationary rate of 40 Hz per neuron. The panel below shows the population histogram that contains in each bin of $\Delta = 2$ ms the number of spikes across the neurons. (**B**) Illustration of the ξ_c-tuples of different order (blue of order 2, red of order 3 and green of order 4). Their respective number of occurrences per bin are shown next to each pattern type. The population coincindence count n_{ξ_c} is the sum of all counts across the bins for ξ_c. (**C**) The p-value ψ of n_{ξ_c} of a given order ξ_c is calculated as the area under the null distribution from the empirical population coincidences to infinity.

In the following we devise a framework for testing the significance of the population coincidence count derived from given empirical data, which we denote as $n_{\xi_c}^{emp}$. For doing this we assume a certain order ξ_0 of correlation among the given spike trains, referred to as "null-order", and define the null-hypothesis $\mathbb{H}_0$ for the test as "the order of correlation

among the given spike trains is up to the null-order ξ_0". This formulation includes the null-hypothesis of independence as a special case of $\xi_0 = 1$. Thus the alternative hypothesis $\mathbb{H}_1$ is "the order of correlation among the given spike trains is higher than ξ_0".

For testing if our empirical data deviate from the null-hypothesis assuming a maximal order ξ_0, we make use of a compound Poisson process (CPP) model [141, 327, 328] to realize the null-hypothesis. It would be optimal if we had the null-distribution $P_{\xi_{c,0}}$ at hand, however, it is analytically not tractable. Therefore we realize the test statistics $n_{\xi_{c,0}}$ and its distribution by the CPP model (see Section 2.4.1 and Appendix 5.1.1 for the detailed explanation of the CPP model). From multiple realizations of the null-model we derive from each the test statistics, and build up the distribution. Then we test if the empirical count $n_{\xi_c}^{emp}$ deviates significantly from the distribution of the $n_{\xi_{c,0}}$ values.

To use the CPP model to realize the null-distribution $P_{\xi_{c,0}}$, the parameters of the model, i.e. $f_A(\omega)$ and λ_{cp} (as explained in Section 2.4.1), need to be specified so that the highest order of correlation among the generated spike trains is ξ_0 and the statistics on the orders up to ξ_0 are consistent with the given empirical data. The model has N free parameters: one is λ_{cp} and the others are $f_A(\omega)$ for $2 \leq \omega \leq N$ ($f_A(1)$ is determined by the condition $\sum_\omega f_A(\omega) = 1$). As we assumed that the highest order is ξ_0 we naturally set $f_A(\omega)$ for $\omega > \xi_0$ to zero. This reduces the number of free parameters to ξ_0, i.e., λ_{cp} and $f_A(\omega)$ for $2 \leq \omega \leq \xi_0$. To further reduce the number of free parameters in f_A, we follow the approach taken in [327, 328] and use an amplitude distribution which has entries only at 1 and ξ_0, i.e., a double-peaked distribution. After these reductions we are left only with three parameters: 1) the assumed highest correlation order ξ_0 and 2) its strength η in the amplitude distribution, i.e. $\eta = f_A(\xi_0)$, and 3) the rate λ_{cp} of the carrier process. As explained in Appendix 5.3.1, the values of these parameters can be derived from the average firing rate and the average pairwise correlation of the given spike trains, once a null order ξ_0 is specified. The average firing rate defined as

$$\hat{\lambda} = \langle \bar{\lambda}_i \rangle = \frac{1}{N} \sum_{i=1}^{N} \bar{\lambda}_i \tag{4.2}$$

where $\bar{\lambda}_i$ is the firing rate of individual neurons i, defines the rate of the carrier process of the CPP model. The average pairwise correlation

$$\hat{\rho} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{i} \bar{\rho}_{ij}}{N(N-1)/2} \tag{4.3}$$

estimated by the Pearson correlation coefficient $\bar{\rho}_{ij}$ between pair i, j, averaged over all possible pairs.

Based on that parameters, we can numerically estimate the null-distribution by computing n_{ξ_c} for multiple realizations of the CPP model and constructing its distribution.

The empirical population coincidence count $n_{\xi_c}^{emp}$ is then tested for its significance. The p-value ψ of $n_{\xi_c}^{emp}$ is derived as the area under the null-distribution $P_{\xi_{c,0}}$ from $n_{\xi_c}^{emp}$ to infinity (Figure 4.1C). If the p-value is smaller or equal to a predefined significance level α, we reject the null-hypothesis and conclude that the data contain correlations of an order higher than ξ_0.

4.3 Results

4.3.1 Example application of the method to synthetic data

Throughout this study, the performance of our method is assessed by applying it to synthetic data with known ground truth. We also use the CPP model (with the double-peaked amplitude distribution) explained in the former section to generate such data. Hence the ground truth of the data is specified by the three parameters of CPP, i.e., the mean firing rate λ of the neurons, the average pairwise correlation coefficient ρ, and the order ξ of the synchrony the data exhibit (termed "true order"). For a given set of values for these parameters we generate $N = 100$ parallel spike trains to mimic a typical number of a parallel recordings by, e.g., a Utah array [279] and for a duration $T = 100\,\text{ms}$ to mimic a typical width of a sliding window used in a time resolved analysis [136]. Example raster plots of two such data sets are shown in Figure 4.2A. In both cases the firing rates of each neuron is 30 Hz, and their average pairwise correlation coefficient 0.002. The true order of inserted correlation is $\xi = 4$ (left panel) and $\xi = 8$ (right panel). From these spike trains we derive the population coincidence count $n_{\xi_c}^{emp}$ for a given test order ξ_c.

The null-distribution P_{ξ_0,ξ_c} is then realized for a model of a certain null-order ξ_0 and the test statistic $n_{\xi_c}^{test}$ of a certain order ξ_c is derived from the realizations of the model. It serves to derive the p-value of $n_{\xi_c}^{emp}$ of the ground truth data set. As illustrated in Figure 4.2B, the shape of the null-distribution depends on the null-order ξ_0 and hence does the p-value of $n_{\xi_c}^{emp}$. In this example, n_3^{emp} from the data of $\xi = 4$ (blue arrow) resulted in a significant p-value of 0.02 (for a chosen significance level $\alpha = 0.05$) for the null order of $\xi_0 = 2$ and test order of $\xi_c = 3$, and n_3^{emp} resulted for $\xi_0 = 6$ for the same test order in a non-significant value (0.11). Both reflect correctly the correlation in the data: since the true order was $\xi = 4$, the data were correctly detected as deviating from the null-hypothesis of order $\xi_0 = 2$, and not detected as significant in case of the null-order of $\xi_0 = 6$. In case of the data set with the true order $\xi = 8$ (green arrow), both tests led to significant results.

However, the shape of the null-distribution also depends on the test-order ξ_c. Figure 4.2C shows the null-distribution for one null-order $\xi_0 = 4$ and for two different test orders ($\xi_c = 2$ (gray) and 4 (black)), as well as the corresponding population coincidence counts $n_{\xi_c=2}^{emp}$, $n_{\xi_c=4}^{emp}$ for data with $\xi = 4$ and $\xi = 8$ together with their p-values. As expected, the data of true order $\xi = 4$ did not become significant for the test orders $\xi = 2, 4$ since the null-order was $\xi_0 = 4$ corresponded to the true order of the data set. For the data set

of true order $\xi = 8$ only the test with the test order $\xi_c = 4$ becomes significant, but not for the test order of $\xi_c = 2$, although in both cases the test order was smaller than the true order. Thus we conclude, that the test order is also influencing the significance test results. In the next section we will study this dependence more systematically in order to understand the reasons for its behavior.

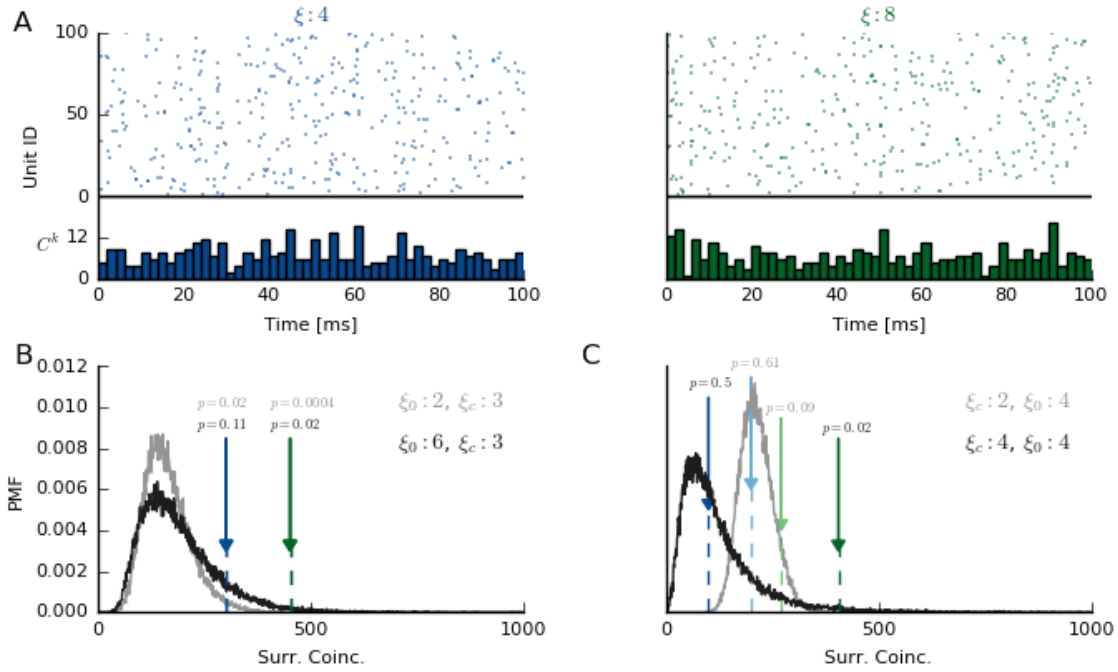

Figure 4.2: **Impact of true order, test order ξ_c and null order ξ_0 on the null distribution and the significance of empirical data.** (**A**) Raster plot and population histogram (below, bin width= 2ms) of synthetic data generated as two CPP models with true order of $\xi = 4$ (blue) and 8 (green), left and right, respectively. Each data set has a duration of 100 ms, and each neuron has a firing rate of 30 Hz. The processes have an average pairwise correlation of $\hat{\rho} = 0.002$ (blue, left) and 0.005 (green, right). (**B**) Null-distributions for different null orders $\xi_0 = 2, 6$ (gray and black, respectively) for a fixed test order $\xi_c = 3$ generated by $R = 10^4$ realizations of the respective CPP model. The blue and green arrows indicate the counts of the empirical spike synchrony events n_3^{emp} retrieved from the two data sets shown in panel A. (**C**) Null-distributions generated in the same way as in panel B for different test orders $\xi_c = 2, 4$ (gray and black, respectively) and a fixed null-order $\xi_0 = 4$. However, in contrast to panel B, the empirical spike synchrony events $n_{\xi_c}^{emp}$ are shown here for two test orders (light colors for $\xi_c = 2$ and darker colors for $\xi_c = 4$).

4.3.2 Calibration of the method

For a systematic assessment of the performance of our method, we evaluate it in terms of false positive (FP) and true positive (TP) rates and explore its dependence on different parameters. To observe these dependencies, we generate multiple realizations of ground truth data via a CPP model with a given set of parameters, apply the method and test in how many of the realizations - given a certain significance level α - the null-hypothesis is rejected. From this we derive *the rate of null-hypothesis rejection* as the ratio between the given count and the total number of realizations R. Note that this rate indicates the TP rate when $\xi_0 < \xi$ and the FP rate when $\xi_0 \geq \xi$.

This procedure is illustrated in Figure 4.3A. We generated $R = 10^4$ realizations of the CPP model with parameters $T = 100\,\text{ms}$, $N = 100$, $\lambda = 10\,\text{Hz}$, $\rho = 0.01$ and a correlation order $\xi = 6$, and derived the p-value for a test statistic for $\xi_c = 1$ (corresponds to the population rate) of $n_{6,1}^{emp}$ for each realization under the null-hypothesis of $\xi_0 = 2$. The dark blue curve in Figure 4.3A shows the distribution of the obtained p-values in the form of a cumulative distribution function (CDF). The rate of the null-hypothesis rejection (here TP since $\xi = 6 > \xi_0 = 2$) is derived as the value of the CDF curve at which the p-value equals the significance level α (here 0.05, gray line). It's value is marked by a dark blue asterisk in Figure 4.3A at the crossing of the latter with the CDF. We extracted the same for different test statistic orders $\xi_c = 2$ to 5 (shown in Figure 4.3B-E, respectively) for identical other parameters. Figure 4.3F summarizes the TP results as a function of ξ_c. Obviously, the performance of the method depends on the choice of ξ_c. For $\xi_c = 1$ the TP rate was very low ($\sim 6\%$), which may be due to the fact that information about the higher-order correlation was attempted to be extracted only based on the population spike count (see (Eq. 4.1)). By increasing ξ_c, the method considers the population coincidence counts of increasing order which enhances the performance up to $\sim 80\%$ at $\xi_c = 3$, and slowly decreased with higher ξ_c. Next, we systematically explore the dependence of the rate of the null-hypothesis rejection on combinations of parameters ξ, ξ_0 and ξ_c. For different true orders $\xi = 4, 6, 8, 10$ (Figure 4.3G-J, respectively) we show the rate of the null-hypothesis rejection (significance level $\alpha = 0.05$) for all combinations of $\xi_0 = 1, 2, \ldots, 8$ and $\xi_c = 1, 2, \ldots, 9$, displayed in matrix form. This matrix of rates is named as *TP/FP rate spectrum* since, as mentioned above, the rate of null-hypothesis rejection represents TP rate when $\xi_0 < \xi$ and FP rate when $\xi_0 \geq \xi$. The border between the representations of TP and FP rates is indicated by a gray line in each matrix: below is the TP region, above the FP region. The white dots indicate the test order ξ_c at which the largest TP rate is taken for each null-order ξ_0.

For all values of ξ the FP rates are at most equal to the significance level ($\alpha = 0.05$), meaning that the FP rate is as expected given the significance level and is not enhanced. The TP rate gets larger the larger the difference between ξ_0 and ξ , meaning that the

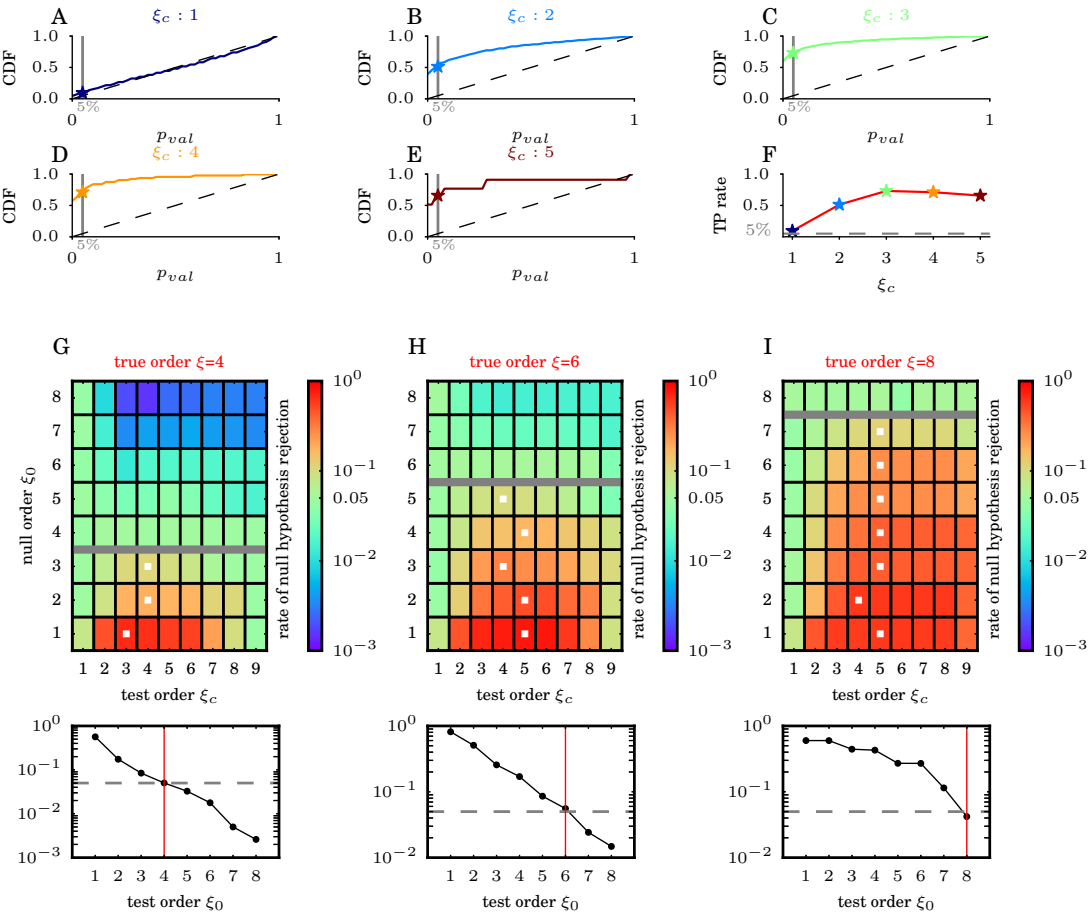

Figure 4.3: **Parameter dependence of p-values.** (**A-E**) Probability distribution of p−values in form of a cumulative distribution function (CDF) over $R = 10^4$ realizations of the CPP model with parameters $T = 100\,\text{ms}$, $N = 100$, $\lambda = 10\,\text{Hz}$, $\rho = 0.01$ and $\xi = 6$. The test parameters are $\xi_0 = 2$ and $\xi_c = 1$,..., 5 in A-E, respectively. The significance level $\alpha = 0.05$ is indicated by the gray line. Its intersection with the CDF marks the null-hypothesis rejection rate, here TP rate (marked by a star). The diagonal dashed lines indicate the dependence of the CDF and the p-values for independent data ($\xi_c = 1$). (**F**) Summarizes the results shown in A-E: TP rate as a function of ξ_c**.** (**G-J**) TP/FP rate spectra for synthetic data of a given true order $\xi = 4, 6, 8, 10$ (other parameters are as A-F), each for varying ξ_0 (horizontal axis, $\xi_0 = 1, 2, \ldots, 8$) and ξ_c (vertical axis, $\xi_c = 1, 2, \ldots, 9$). The border between the representations of the TP (when $\xi_0 < \xi$) and FP (when $\xi_0 \geq \xi$) rates are shown by a gray line. The highest TP rate is indicated by a white dot for each ξ_0 as a function of ξ_c. Below of each TP/FP rate spectra the changes of rate of null hypothesis rejection is shown for a fix ξ_c and different ξ_0. The vertical red line indicates the null order which is equivalent to the true order.

power of the method becomes larger when the data contain higher order of synchrony than assumed in the null-hypothesis.

As seen before, for an increasing test order ξ_c and a fixed ξ_0 the TP rate increases until a certain ξ_c (shown by white dots in Figure 4.3G-J) and then decreases again. In the following we carefully investigate this dependence of the TP rate on ξ_c.

Dependence of the TP and FP rates on the test order ξ_c

To illustrate how the choice of ξ_c affects the performance of the method, we investigate the dependence of the empirical population coincidences $n^{emp}_{\xi_c}$ and the null distribution $P_{\xi_{0,c}}$ on ξ_c. To investigate the dependence of $n^{emp}_{\xi_c}$ on ξ_c, we generated again $R = 10^4$ realizations of the CPP model with parameters $T = 100\,\text{ms}$, $N = 100$, $\lambda = 10\,\text{Hz}$, $\rho = 0.01$ and $\xi = 6$, and computed the distributions $P_{\xi_{0,c}}$ from $n^{test}_{\xi_c}$ for different values of ξ_c. As illustrated in Figure 4.4A, the distribution becomes broader as ξ_c gets larger. The peak of the distribution shifts to the right until $\xi_c = 3$ or 4 and then shifts back to the left with its peak getting higher again. These changes can be explained in terms of the dependence of the binomial coefficient $\begin{pmatrix} C^k \\ \xi_c \end{pmatrix}$ on ξ_c, because $n^{test}_{\xi_c}$ is defined as a sum of $\begin{pmatrix} C^k \\ \xi_c \end{pmatrix}$ in (Eq. 4.1).

How the distribution shape changes as a function of ξ_c depends on the true order ξ of the data. To illustrate this dependence, the median of the distribution is plotted against ξ_c for several values of ξ in Figure 4.4B. The rightward shift of the distribution, which is represented by the growth of the curve for small values of ξ_c in Figure 4.4B, is faster for larger values of ξ. This reflects the fact that, for a given ξ_c, $n^{emp}_{\xi_c}$ takes a larger value for larger ξ because the data contain more higher order coincidence events.

Now, to understand how this behavior relates to the performance of the method, we study the relation between one of these distributions with $\xi = 6$ and the null distribution P_{ξ_0} for different ξ_0 values (here $\xi_0 = 2, 6$ and 8 are chosen). For this purpose we characterize the distribution of $n^{emp}_{\xi_c}$ by several percentiles, i.e. the $25, 50, 75, 95$ and 99-th percentiles and plot them together with the 95th percentile of P_{ξ_0}(see Figure 4.4C). Let us first focus on $\xi_0 = 2$ (blue dashed curve in Figure 4.4C) and the 25th percentile of the distribution of $n^{emp}_{\xi_c}$ (the darkest gray curve in Figure 4.4C) and see how their relation changes depending on ξ_c. At $\xi_c = 1$ the darkest gray curve is below the blue dashed curve, indicating that the rate of null hypothesis rejection is less than $1 - 0.25$. As ξ_c increases from 1, the 25th percentile value increases and the darkest gray curve goes above the blue dashed curve at $\xi_c = 3$, meaning that the rate of null hypothesis rejection exceeds $1 - 0.25$. Beyond $\xi_c = 3$ the 25th percentile value decreases and the darkest gray curve goes below the blue dashed curve again at $\xi_c = 5$. In this way the rate of null hypothesis rejection can be derived from this plot by reading the crossings of the gray curves and blue dashed curve.

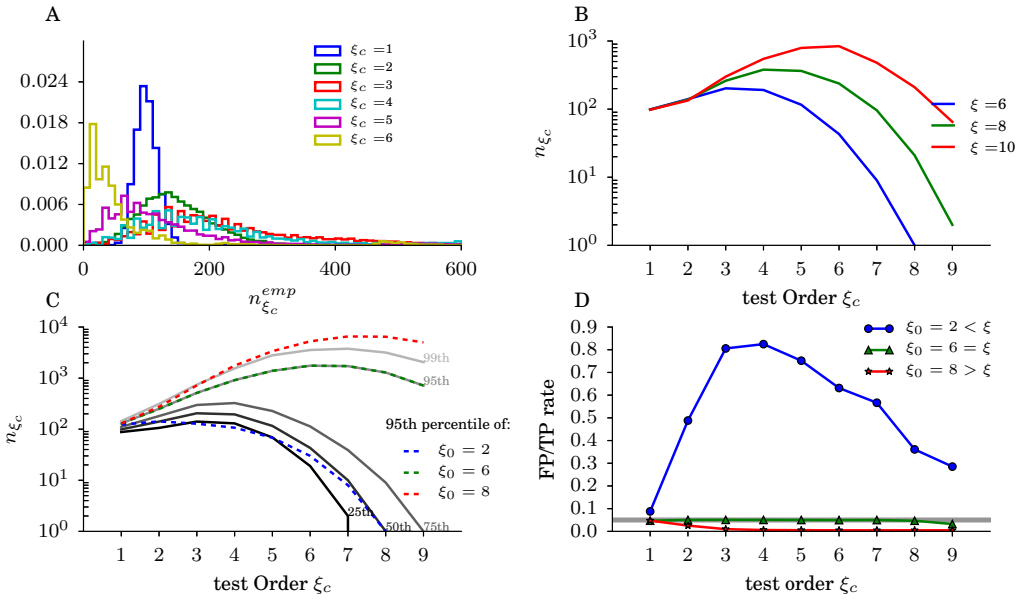

Figure 4.4: **Relation between test order and TP/FP rates.** (**A**) Distribution of $n^{emp}_{\xi_c}$ calculated over many realizations of the CPP model, with fixed parameters $N = 100$, $T = 100\,\text{ms}$, $\lambda = 10\,\text{Hz}$, $\xi = 6$ and different test orders $\xi_c = 1, 2, \ldots, 6$. (**B**) 50th percentile of the $n^{emp}_{\xi_c}$ distribution versus different test order ξ_c for true orders $\xi = 6, 8$ and 10 in blue, green and red, respectively. The other parameters are as A. (**C**) Distribution of $n^{emp}_{\xi_c}$ characterized by $25, 50, 75, 95, 99$-th percentiles versus different test order ξ_c for $\xi = 6$ in shades of gray. Dashed lines show the 95th percentile of the null distribution for different null order of $\xi_0 = 2, 6, 8$ in blue, red and green respectively. (**D**) TP (in blue) and FP (in green and red) rates at 0.05 significance level for the data with true order $\xi = 6$ are shown versus the test order ξ_c.

Figure 4.4D shows the rate of null hypothesis rejection obtained in such a way for $\xi_0 = 2, 6$ and 8. In the case of $\xi_0 = 2$, the growth of the gray curves is faster than that of the blue dashed curve for small ξ_c values, and hence the rate of null hypothesis rejection, which is TP rate in this case, becomes higher as ξ_c takes larger values up to $\xi_c = 4$. Beyond that, the gray curves drop fast and hence the TP rate goes down. In the cases of $\xi_0 = 6$ and 8, the gray curves grow at most as fast as green and red dashed curve and hence the rate of null hypothesis rejection, which is FP rate in these cases, never exceeds the predefined significance level (here $\alpha = 0.05$).

Dependence of TP and FP rates on firing rate, number of neurons and data length

We further study how the performance of the method depends on other properties of the data, such as the firing rate, the number of neurons and the length of the data. For these dependencies we also compute TP/FP rate spectra.

First, we vary the firing rate λ of the neurons while we fix the $\xi = 6$, the number of neurons $N = 100$, and the data length $T = 100\,\text{ms}$. Figure 4.5A-C show the TP/FP rate spectra for $\lambda = 20\,\text{Hz}$, $50\,\text{Hz}$ and $100\,\text{Hz}$, respectively. The spectra show that there is no increase in the FP rate with an increase of the firing rate of the neurons. The TP rates seem also almost not affected, but the peak of the highest TP rates are at higher ξ_c values as the firing rate increases. This means that the optimal ξ_c value depends on the firing rate in the data set.

Next, we vary the number of neurons N while keeping $\xi = 6$, $\lambda = 10\,\text{Hz}$ and $T = 100\,\text{ms}$ fixed. The corresponding TP/FP rate spectra are shown in Figure 4.5D-F for $N = 200, 400$ and 600, respectively. The spectra show a very similar behavior as for the increase in firing rate. Also the optimal ξ_c value increases in a similar fashion to higher values with N as for the rate.

Lastly we computed the TP/FP rate spectra for various data length T for fixed $\xi = 6$, $\lambda = 10\,\text{Hz}$ and $N = 100$ (see Figure 4.5G-I). As for the cases before, the FP rate does not increase with the length of the data set, but rather decreases for an intermediate test order ξ_c and for high ξ_0, but generally is at most equal to the predefined significance level (here $\alpha = 0.05$). The TP rate is overall higher. The optimal ξ_c for maximum TP rate varies stronger with ξ_0 as for the former parameters: the higher ξ_0, the lower the optimal ξ_c.

In summary we observed the following parameter dependencies: generally the TP rate decreases with ξ_0 getting closer to ξ. The optimal ξ_c is not constant for different ξ_0. The more data samples are available (as for longer data sets) the higher the TP rate and the lower the FP rate. Still, the FP rate is here lower than the significance level, which is not expected. An increase of the number of neurons N and an increase of the firing rate leads to a decreased sensitivity (lower TP rates and higher FP rates), which may be due to the increase of *"noise"* data by these parameters, i.e. more spikes not involved in correlations.

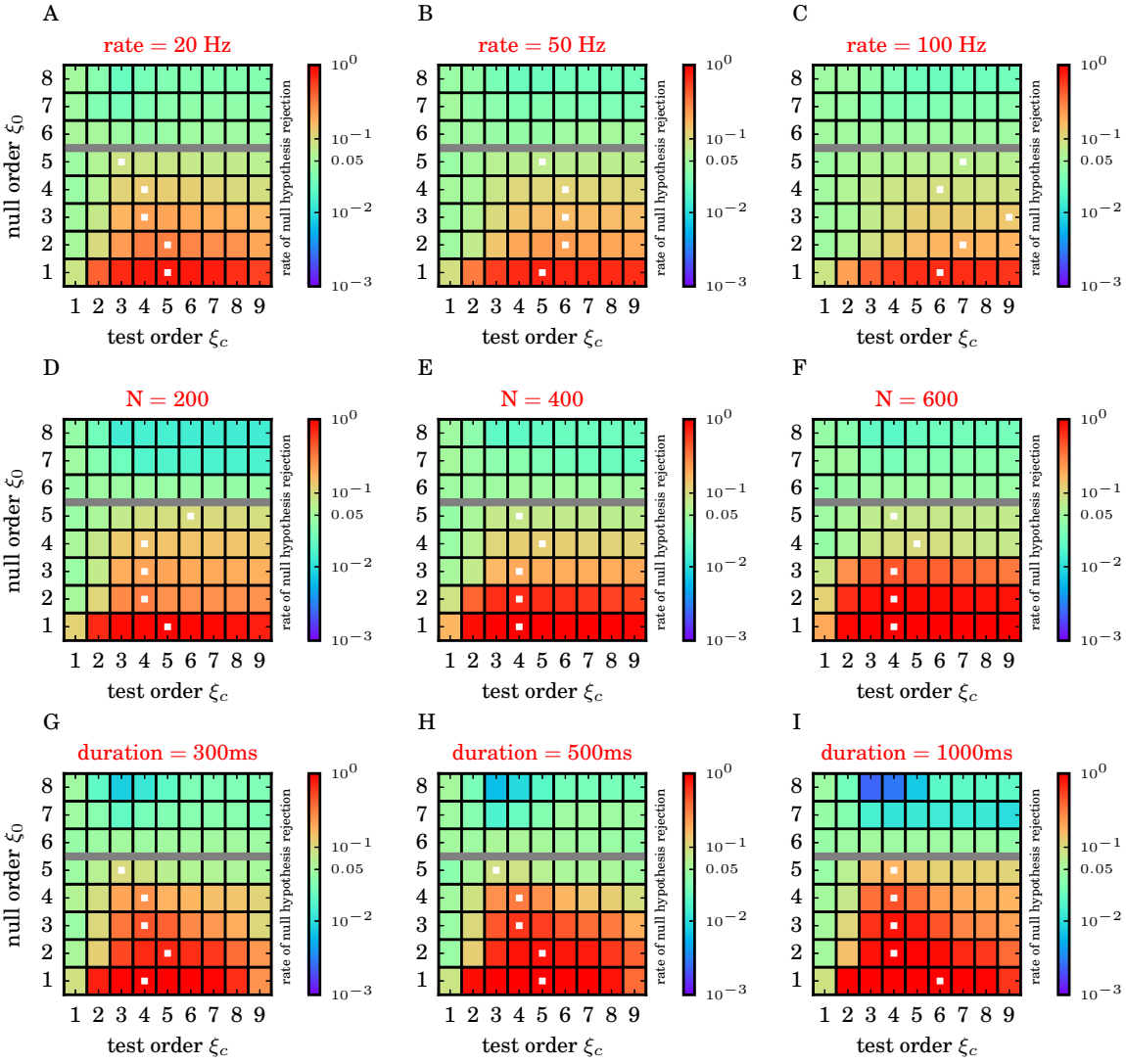

Figure 4.5: **TP and FP rates for different firing rates, number of neurons and data lengths.** (**A-C**) TP/FP rate spectrum at $\alpha = 0.05$ on $R = 10^4$ realizations of the CPP model with parameters $\xi = 6$, $N = 100$, and $T = 100\,\text{ms}$ shown for different firing rates $\lambda = 20\,\text{Hz}$, $50\,\text{Hz}$ and $100\,\text{Hz}$ from A to C. (**D-F**) TP/FP rate spectrum for parameters $\xi = 6$, $\lambda = 10\,\text{Hz}$ and $T = 100\,\text{ms}$ for different number of neurons $N = 200, 400$ and 600 from D to F. (**G-I**) TP/FP rate spectrum for $\xi = 6$, $\lambda = 10\,\text{Hz}$ and $N = 100$ equal to Figure 4.4B for different data lengths $T = 300\,\text{ms}, 500\,\text{ms}$ and $1000\,\text{ms}$ from G to I.

4.3.3 Changing perspective: how to approach unknown data

The question is now how to approach unknown, typically experimental data, of which we do not know the ground truth. From such data we would like to know a) if there is (higher-order) synchrony contained, b) if these change as a function of time, and c) of what the order of synchrony is. In this section we approach these questions.

Estimation of firing rate and pairwise correlation

When applying the PUE method to an unknown data set, we first require to estimate the average firing rate and average pairwise correlation of the data to parametrize the CPP model of our null-hypothesis. Here the question arises to what extent a deviation of the estimated values from the true values, due to finite data size, affects the performance of the PUE method. To answer that we use again data of known ground truth. First we quantify the deviation of estimated values from the true values with respect to the amount of available data. Therefore, we generate 400 realizations of a CPP model of $N = 100$, $\lambda = 10\,\mathrm{Hz}$, $\rho = 0.005$ and $\xi = 6$ and calculate from the data the firing rate $\hat{\lambda}$ and the pairwise correlation $\hat{\rho}$ using (Eq. 4.2) and (Eq. 4.3) for each realization. Throughout this chapter we use a circumflex $\hat{\cdot}$ to denote estimated values. We repeat the same procedure for data sets of different durations ($T = 100, 300, 500, 1000, 3000$ and $10000\,\mathrm{ms}$). For each duration T, we characterize the distributions of the obtained values of $\hat{\lambda}$ and $\hat{\rho}$ by a Box-and-whisker plot (see Figure 4.6A). As expected, the deviations of the estimates $\hat{\lambda}$ and $\hat{\rho}$ from the ground truth values λ and ρ, respectively, are large for small durations (up to about $100\,\mathrm{ms}$) and they decrease as T is increased. Furthermore, for short duration of the data ($100\,\mathrm{ms}$), $\hat{\rho}$ is generally underestimated (compare the red line in the Box-and-whisker plot with the true value of ρ) and it has negative values. The underestimation arises because of the small number of spikes.

Next we investigate the impact of the deviation of $\hat{\lambda}$ and $\hat{\rho}$ from the underlying true values λ and ρ on the resulting p-value of the PUE analysis. For doing that we first compute the p-values of the ground truth data, i.e. realizations of the CPP model with $N = 100$, $\lambda = 10\,\mathrm{Hz}$, $\rho = 0.005$ and $\xi = 6$ and compute the p-value based on the null hypothesis $\mathbb{H}^0(\lambda, \rho, \xi_0)$, i.e. using the theoretically given rate λ and the pairwise correlation coefficient ρ. We plot for each of the realization the estimated $\hat{\lambda}$ and $\hat{\rho}$ and the p-value of the evaluation of the higher-order correlation (Figure 4.6B, red dots). The distribution of the p-values (red histogram on the right) is - as expected - approx. homogeneously distributed over the range of 0 and 1. The variations of the p-values result from the stochastic variations of the process realizations for given λ and ρ. We notice that the p-value becomes the lower the lower the resulting $\hat{\lambda}$. The same holds for the correlation $\hat{\rho}$, or in other words $\hat{\lambda}$ and $\hat{\rho}$ correlate positively. The estimated pairwise correlations show even negative values.

Now we compute the p-values as we would do for unknown data. For the same data as before we now use the estimated $\hat{\lambda}$ and $\hat{\rho}$ from each realization for the implementation of the null-hypothesis, i.e. $\mathbb{H}^0(\hat{\lambda}, \hat{\rho}, \xi_0)$ (Figure 4.6B, blue dots). For $T = 100\,\text{ms}$ data duration about 15% of the realizations led to negative correlation coefficients. Since the CPP is not defined for that case we could not perform the PUE analysis and do also not show the estimated values for $\hat{\lambda}$ and $\hat{\rho}$. The distribution of the p-values for the resulting data is much more clustered than compared to the results based on $\mathbb{H}^0(\lambda, \rho, \xi_0)$. Here the p-values range from slightly above zero to about 0.65. The p-value is higher (i.e. less significant) the higher the estimated $\hat{\rho}$, and there seems no correlation to the estimated rates $\hat{\lambda}$.

Figure 4.6B-E show the scattergrams of the p-values versus $\hat{\lambda}$ and $\hat{\rho}$ (blue dots) as well as the marginal histograms of each of these measures (blue) for different durations of the data. For comparison, we also calculate the p-values resulted from the test which uses the ground truth values λ and ρ for implementing the null hypothesis $\mathbb{H}^0(\lambda, \rho, \xi_0)$.

Let us first focus on Figure 4.6B, which shows the result for the data duration of 100 ms. The obtained p-values from the null hypothesis $\mathbb{H}^0(\lambda, \rho, \xi_0)$ are distributed uniformly (Figure 4.6B, the right most histogram, red). This is as expected when data correspond to the null hypothesis. The p-values show a negative correlation for $\hat{\lambda}$ and $\hat{\rho}$. For $\hat{\rho}$, this is because realizations with larger pairwise correlations have larger empirical test statistics $n_{\xi_c}^{emp}$ as well. For $\hat{\lambda}$, this is simply because there is a correlation between $\hat{\lambda}$ and $\hat{\rho}$, i.e. those realizations with larger $\hat{\lambda}$ (hence larger number of spikes) have larger $\hat{\rho}$. Comparing the p-values resulted from $\mathbb{H}^0(\hat{\lambda}, \hat{\rho}, \xi_0)$ with those from $\mathbb{H}^0(\lambda, \rho, \xi_0)$ shows a bias of the former p-values toward lower values when the estimated pairwise correlation $\hat{\rho}$ is very small and close to zero (Figure 4.6B the right scattergram). This bias in the p-values arises because of the small number of spikes in the synthetic data. As discussed above, for data duration of 100 ms, the small number of spikes causes the underestimation of $\hat{\rho}$ (see Figure 4.6A). This underestimation leads to a null hypothesis with smaller correlations than actual data and , hence, smaller p-values.

Next, we increase the data duration and study the behavior of p-values for $T = 300, 500$ and $1000\,\text{ms}$. As it is shown in Figure 4.6C-E the bias observed in the results for the data duration of 100 ms vanishes for longer data durations and p-values are concentrated around 0.5, demonstrating that the test with $\mathbb{H}^0(\hat{\lambda}, \hat{\rho}, \xi_0)$ is conservative compared to the test with $\mathbb{H}^0(\lambda, \rho, \xi_0)$.

We further quantify the performance of the method by the FP and TP rates for different data durations and various ξ_c and ξ_0. We use the same data as in Figure 4.6B-E and implement the null hypothesis with $\hat{\lambda}$, $\hat{\rho}$ calculated for each realization of the data to generate the TP/FP rate spectra, as shown in Figure 4.7A-D. For $T = 100\,\text{ms}$, comparing Figure 4.7A to Figure 4.3G shows a decrease of rate of null hypothesis rejection for $\xi_c = 1$ and 2. This is because the empirical population coincidences n_1^{emp} and n_2^{emp} (see

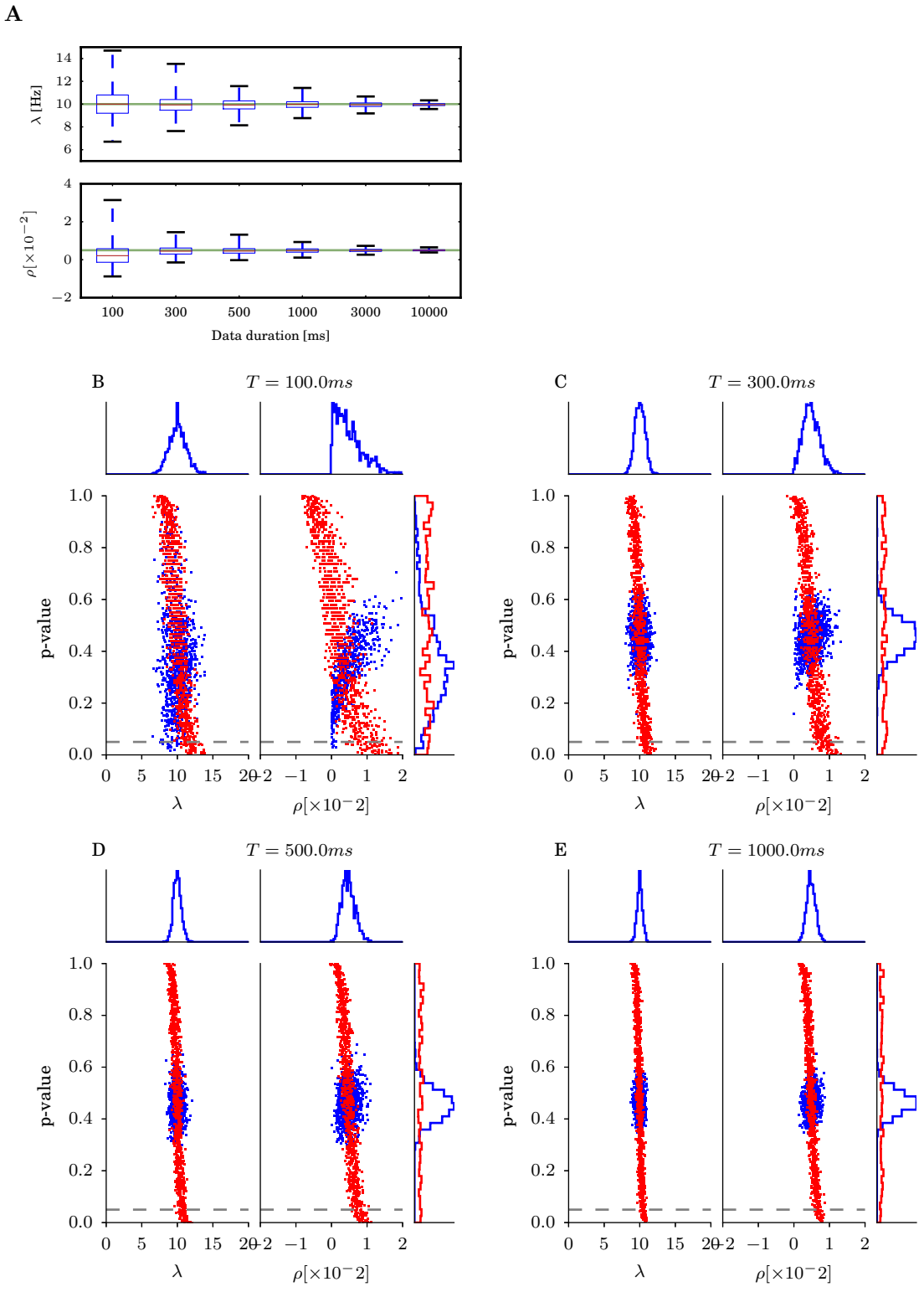

Figure 4.6: *(Caption on the next page)*

Figure 4.6 *(previous page)*: **Deviation of estimated average firing rate and pairwise correlation of the CPP data from the ground truth.** (**A**) Estimated $\hat{\lambda}$ and$\hat{\rho}$ (on top and bottom respectively) over 400 realizations of the CPP model with parameters: $N = 100$, $\lambda = 10Hz$, $\rho = 0.005$, $\xi = 6$ versus duration of the data ($T = 100, 300, 500, 1000, 3000, 10000$ ms). (**B**) Blue: Each dot represents a p-value computed on one realization of the CPP data (with identical parameters as the data A) with $T = 100$ ms and using $\hat{\lambda}$, $\hat{\rho}$ and $\xi_0 = 6$ for implementing the null hypothesis. The p-values are plotted versus $\hat{\lambda}$ and $\hat{\rho}$ (in left and right columns respectively) correspond to each realization of the data. Note that those realizations with $\hat{\rho} < 0$ are discarded as the CPP can only generate data with independence or positive correlations. The histograms of $\hat{\lambda}$ and $\hat{\rho}$ values are plotted on top and for the p-values on the right. Red: The same as blue but using the ground truth values of the realized data (true firing rate and pairwise correlation λ and ρ) for implementing the null hypothesis. (**C-E**) The same as B for different data durations $T = 300, 500, 1000$ ms

(Eq. 4.1)), which count the total number of spikes (i.e. firing rate $\hat{\lambda}$ times the number of bins) and the total number of pairwise coincidences, respectively, are accounted for by the estimated firing rate $\hat{\lambda}$ and pairwise correlation $\hat{\rho}$. For $\xi_c > 2$, Figure 4.7A in comparison to Figure 4.3G shows a slight increase of the FP rate. This increase can be explained by the bias in p-values toward lower values observed in Figure 4.6B. For $T > 100$ ms (compare Figure 4.7B-D to Figure 4.5G-I), the rate of null hypothesis rejection decreases. This means that the PUE method is more conservative when the firing rate and pairwise correlation estimated on a single realization are used to construct the null hypothesis. As for the TP rate we see that a larger TP rate is obtained by a better estimation of the firing rate and pairwise correlation (longer data have larger TP rate).

The observed bias in the p-values can be avoided by increasing T. But the question now is: how long the data should be? or in the other word what is the lower bound of the number of spikes that assures un-biased p-values? Based on further analysis (not shown) for various firing rates λ, pairwise correlations ρ, number of neurons N and data durations T, we propose that the total number of spikes across all neurons in the selected data should equal or larger than number of neurons N. This lower limit identifies an appropriate T given the firing rate of neurons. For example in a case where neurons have an average firing rate of 5 Hz the duration of data we need for an un-biased p-value is about $T = 300$ ms.

In summary, estimation of firing rates and pairwise correlations from a limited data set deviates from the ground truth values. For a short data duration (up to about 100 ms), this deviation might result in higher FP rates than expected (roughly 1% more for α =0.05 in the example data set shown here). This problem arises because of not having enough spikes for he analysis. The TP rates of the method increases by having longer data.

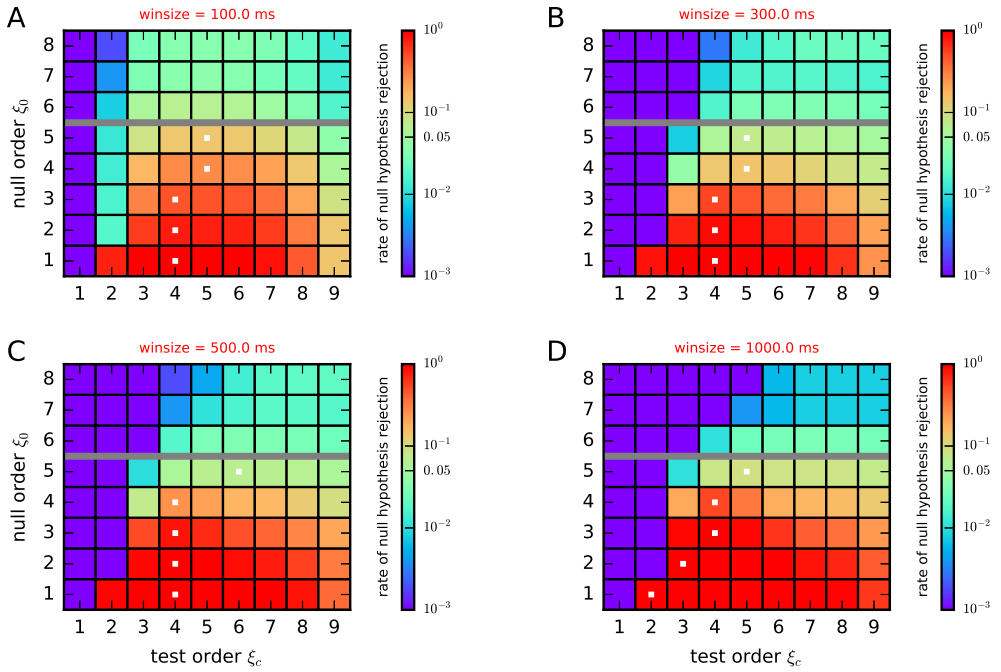

Figure 4.7: **FP and TP rates with estimated firing rate and correlation.** (**A**) TP/FP rate spectra of the data similar to Figure 4.6A for various null orders ($\xi_0 = 1, 2, \ldots, 8$) and test orders ($\xi_c = 1, 2, \ldots, 9$) with estimated firing rate and correlation in small window of 100 ms. (**B-D**) Similar to A for different data durations of $T = 300, 500, 1000$ ms, respectively.

Time resolved analysis

An important aspect of the proposed method is its ability to obtain a reliable statistic for short data segment (in a range of 100 ms). This allows us to apply the method in a sliding window fashion to longer data sets to examine a potential temporal modulation of the higher-order synchrony. We demonstrate this application using synthetic data with time dependent higher-order synchrony. These are $T = 2000$ ms CPP data where the order of synchrony changes from $\xi = 2$ to $\xi = 6$ for a time interval of 400 ms, referred to as period of enhanced synchrony hereafter. The other parameters (such as the firing rate and the pairwise correlation coefficient) stay constant. Figure 4.8A, top panel displays the time course of the order of synchrony.

Then we choose an analysis window of duration $T_w = 300$ ms and estimate the average firing rate and the average pairwise correlation coefficient $\hat{\lambda}$,$\hat{\rho}$ of the data therein. With these estimated parameters and a null order $\xi_0 = 2$, we calculate the p-value within a window. We slide the window along the data in steps of 10 ms, and at each window position do the same analysis. Thus in the end we have a time dependent p-value. To pronounce the changes in small p-values (ψ), we illustrate the significance instead by the surprise S which is a logarithmic transformation of p-value ψ: $S = \log\frac{1-\psi}{\psi}$.

Figure 4.8A, second panel shows the raster plot and the population histogram, calculated with a bin width of 1 ms, of the synthetic data set. The average pairwise correlation coefficient and the average firing rate estimated at each window position are plotted in Figure 4.8A, third and fourth panels, respectively. The bottom panel in Figure 4.8A displays the surprise, computed using a range of test orders ξ_c (see color bar). The surprise fluctuates as a function of time. The largest surprise values are attained for $\xi_c = 4$ within the period of enhanced synchrony. This corresponds to our observation (in Section 4.3.2) that the TP rate increases for larger ξ_c until an optimal ξ_c^* and then decreases due to the paucity of samples. Thus the surprise indicates indeed that there is an increase of the synchrony order in the period of enhanced synchrony (gray area). The presence of the synchrony of order $\xi = 2$ before and after the period of enhanced synchrony is not detected as significant, because this synchrony order is accounted for by the null hypothesis ($\xi_0 = 2$).

Concatenated trials Analysis

In realistic scenarios of analysis of experimental data we often have multiple realizations of the same stimulus or behavioral condition, i.e., multiple trials. We propose an approach, which we name *"Concatenated trials Analysis"*, to make use of trials to enhance the performance of the method.

Given M trials of N parallel spike trains in a time window of width T_w, we concatenate all M trials along the time axis to form a new set of N spike trains of duration $T_w \times M$ (see Figure 1 of [136] for an illustration of the concatenation procedure). Now we apply the PUE method on the concatenated data and calculate the surprise for the window. We can obtain the time modulation of surprise by applying the concatenation for each window position and repeat the same procedure. An important remark is that this approach works under the assumption of stationary firing rates and pairwise correlations across trials.

Figure 4.8B displays the results of the concatenated trials analysis for various number of realizations ($M = 20, 50$), mimicking data from multiple trials, of the same CPP as Figure 4.8A with various window sizes ($T_w = 50, 100, 300$ ms) in steps of 20 ms and $\xi_0 = 2$. The surprise increases in the period of enhanced synchrony order for $\xi_c > 2$. This increase starts when the analysis window enters the enhanced synchrony period and reaches its maximum when the window is completely within this period. Therefore, the time onset of the increase in the surprise is different for different window sizes. For some ξ_c, e.g. $\xi_c = 4$, the surprise is extremely significant (with p-value< 0.0001). Outside this period, where the true order ($\xi = 2$) equals null order ($\xi_0 = 2$), the surprise remains below 5% significance level (gray dash line). Comparing the results of Figure 4.8A (single trial) to Figure 4.8B (multiple trials) shows that the concatenated trials analysis is a suitable approach to enhance the power of the method and including more trials leads to a larger power.

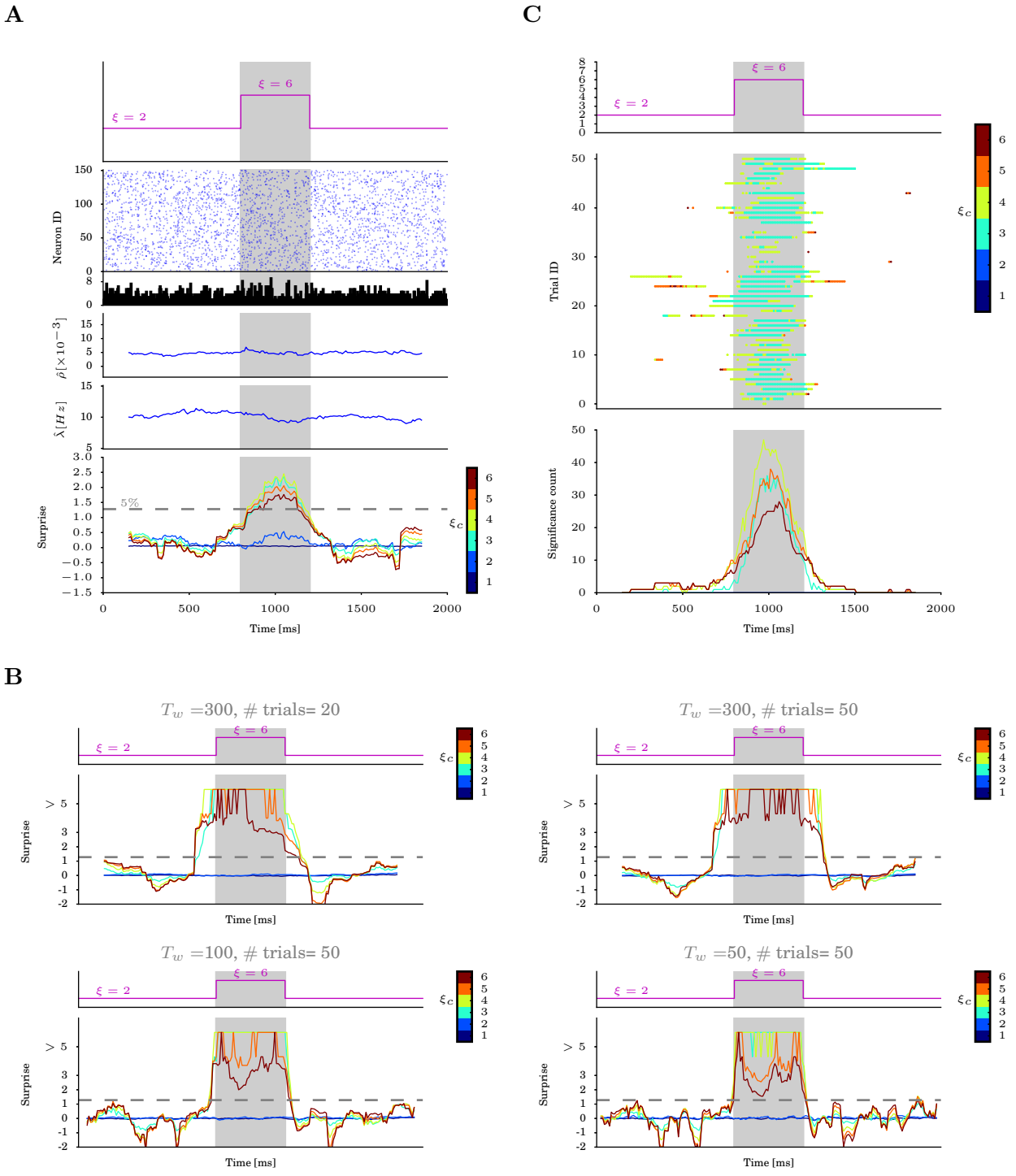

Figure 4.8: *(Caption on the next page)*

Figure 4.8 *(previous page)*: **(Simulated data) Detection of significant higher order correlation in a time resolved manner.** (**A**) *Top panel:* Simulated data using CPP model with stepwise changes in the order of correlation, from $\xi = 2$ to $\xi = 6$ and then back to $\xi = 2$ while the pairwise correlation is constant. The amplitude distribution of the CPP has a double peaked shape at 1 and ξ. To keep the firing rate constant, the rate of mother process λ_{mp} is adjusted during this time. *Second panel:* Dot display and population spike counts of the simulated data. Population spike counts are calculated with a bin width of $\Delta = 1\,\text{ms}$. *Third and fourth panels:* Estimated pairwise correlation coefficient $\hat{\rho}$ and firing rate $\hat{\lambda}$ in a sliding window of 300 ms in step of 20 ms. *Bottom panel:* Surprise calculated in a time resolved manner with the estimated firing rate and pairwise correlations $\hat{\lambda}$, $\hat{\rho}$ in the same windows as shown in the third and fourth panels. Different colors correspond to different test orders used to derive the surprise values. (**B**) Time modulation of surprise obtained from concatenated trial analysis is shown for various trial numbers and window sizes of the same CPP as A. The top panel shows the modulation of surprise for a fixed window size of 300 ms and different trial numbers of $M = 20, 50$ (left and right, respectively). In the bottom panel the number of trial is fixed ($M = 50$) and window size has been changed from 100 ms to 50 ms, from left to right respectively. (**C**) *Top panel:* Time course of correlation order (same as the top panel in A). *Middle panel:* Windows with significant (5%) surprise indicated by dots. In each row surprise is calculated on one realization of simulated data (the same parameters as A). Each realization is referred to one trial. Color indicates the least coincidence order on which surprise is significant i.e. for a given window, if the surprise is always significant for coincidence order more than e.g. 4, the dot represents only $\xi_c = 4$. *Bottom panel:* Results of cross-trial analysis on 100 trials shown as number of trials with significant surprise, summed across trials for each window position $W_{sum}(t_w)$.

Cross-trial Analysis

A limitation of concatenated trials analysis is that it requires the assumption of stationarity across trials. This stationarity assumption may not hold in given experimental data. Here we propose another approach for the circumstances where data violates the stationarity assumption across trials.

In this approach, which we name *"Cross-trial analysis"*, given M trials of duration T, we apply the PUE method in a time resolved manner to one trial j. We indicate the window position in time by t_w ($t_w = T_w/2 + nT_{step}$, where n is an integer number and T_{step} is the step of sliding window), which corresponds to the center of the window at each window position. The time resolved analysis yields the surprise at each window position in trial j, i.e., $S^j(t_w)$. With a significance level α, we identify the window positions in which the surprise is significant $S^j(t_w) > S_\alpha$, where $S_\alpha = log\frac{1-\alpha}{\alpha}$. Then, we construct a binary vector $\vec{W}^j(t_w)$ (of the same length as $S^j(t_w)$) where "1" indicates window positions with significant surprise and "0" otherwise. After obtaining the vectors of all trials, we compute

their sum $\vec{W}_{sum}$ over trials:

$$\vec{W}_{sum}(t_w) = \sum_{j=1}^{M} \vec{W}^j(t_w). \tag{4.4}$$

$\vec{W}_{sum}(t_w)$ is now a vector which contains the number of significant trials at each window position.

To demonstrate the application of the proposed method, we apply it with $\xi_0 = 2$, $T_w = 300\,\text{ms}$ and $T_{step} = 10\,\text{ms}$ to the same data set as for Figure 4.8B with significance level $\alpha = 0.05$. Figure 4.8C, second panel displays the window positions where the surprise is significant by colored dots. The color of each dot indicates the smallest ξ_c value that gives a significant result at the corresponding window position (see color bar). Note that if, for instance, the surprise is significant for $\xi_c = 4$, it usually remains significant for $\xi_c > 4$ (until an optimal ξ_c^* which was discussed in Section 4.3.2).

Figure 4.8C, bottom panel displays $W_{sum}(t_w)$ for different test orders ξ_c. $W_{sum}(t_w)$ is small ($\leq M \times \alpha$) for windows within the periods of $\xi = \xi_0 = 2$ and increases for the windows within the period of enhanced synchrony $\xi = 6 > \xi_0 = 2$. Thus, the proposed cross-trial analysis is an approach to increase the power of the method in terms of true positive (TP) detection. Importantly, this approach does not presume any stationarity in the firing rates and the pairwise correlations across trials and hence makes the proposed method applicable to data sets which do not fulfill the assumption of statistical stationarity across trials.

Which order of synchrony?

In Section 4.3.2 we discussed the parameter dependencies of the rate of null hypothesis rejection (or equivalently the TP and FP rates). We learned that for a fixed ξ_c, and for increasing ξ_0, the rate of null hypothesis rejection decays, until it reaches the percentage equivalent to the significance level, and then further decays (Figure 4.3G-I). From the above observation we learned that the percentage equivalent to the significance level is reached when ξ_0 is at or close to ξ. Thus the changes in the rate of null hypothesis rejections as a function of null order ξ_0 would provide us a way to estimate the underlying synchrony order ξ. As we also observed in Figure 4.5, this behavior is independent from other parameters, such as the firing rate, the number of neurons or the length of the data. Only the steepness and the modulation depth vary depending on these parameters.

Here we propose a procedure to extract the synchrony order from unknown data. We apply this procedure to the synthetic data used in the previous section and infer what the order of synchrony is at window positions where the null hypothesis (with $\xi_0 = 2$) is rejected. We name these window positions window of interest (WOI). Note that in the other window positions the order of synchrony is at most 2, i.e. there is no higher order synchrony outside WOIs.

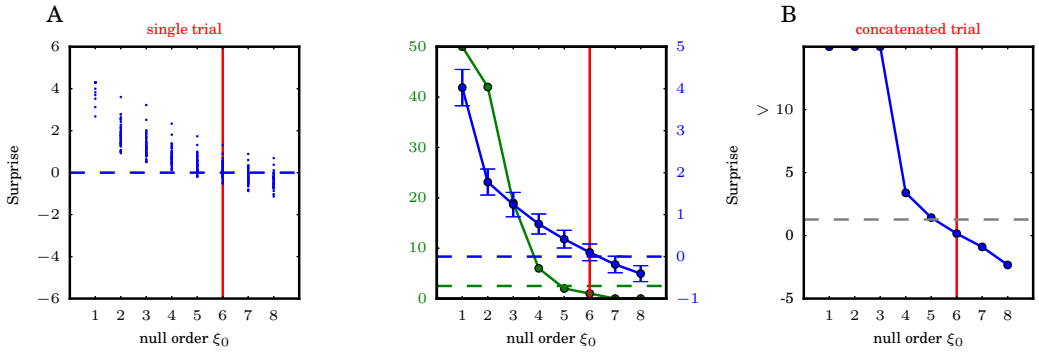

Figure 4.9: **(Simulated data) Extracting the order of synchrony.** (**A**) *Left panel:* Changes of the surprises versus the null order ξ_0: for a given null order each dot represents the value of surprise calculated for different trials ($M = 50$ trials) of the window centered at $t = 1000$ ms in Figure 4.8B. Here the $\hat{\lambda}$ and $\hat{\rho}$ are estimated in each window. The red horizontal line indicates the true order ξ in the data. *Right panel:* Green curve: percentage of significant surprise of A with $\alpha = 0.05$ versus null orders. The green dashed line indicates the 5%. Blue curve: 95% confidence interval of the mean of surprises from A for different ξ_0. The blue dashed curve is at 0 and the red horizontal line indicates the true order ξ in the data. (**B**) Surprise calculated by concatenated trial analysis on the same data as A for different null orders $\xi_0 = 1, 2, \ldots, 8$.

For the data used in the previous section, the WOIs are within the period of enhanced synchrony (see Figure 4.8B-C, gray area). We choose one of the WOIs and explain the procedure for inferring the order of synchrony in the following for each of the two approaches to multi-trial data proposed in previous sections: 1) Concatenated trials analysis, and 2) Cross-trial analysis.

In the concatenated trials analysis, we apply the test on the concatenated data (over trials) of the selected WOI with different null orders ξ_0. We infer the true order of synchrony in the given data from the bahavior of surprise as a function of ξ_0 . Figure 4.9B shows the result for the WOI at 1000 ms in Figure 4.8B. For the test order we choose $\xi_c = 4$ because it leads to the maximum significant count in the selected WOI. The surprise decreases as the null order increases and becomes non-significant when the true order and the null order are equal ($\xi_0 = \xi$).

For the cross-trials analysis, starting with $\xi_0 = 1$, we apply the PUE on each trial separately. Then we change the null order consecutively and repeat the same analysis. The scatter plot of the obtained surprise values over trials are shown in Figure 4.9A for different $\xi_0 (= 1, 2, 3, \cdots, 8)$. For small ξ_0, the surprises have overall large values and by increasing ξ_0 they decrease. To quantify the behavior of the surprises, we employ two measures:

1. Significant count: For each null order ξ_0 we count the number of trials with significant surprises and plot the count as a function of ξ_0 (see Figure 4.9B, green curve). This measure, then, indicates the true order of synchrony at the null order where

the number of significant surprises is below significant level (green dashed line in Figure 4.9B). This inference depends on the number of trials and data durations and it might vary from the underlying true synchrony order, e.g. here it infers synchrony order of 5 whereas the underlying order is 6.

2. Avergae surprise: For a given null order ξ_0, we calculate the 95% confidence interval of the mean surprises over trials (see Figure 4.9B, blue curve). Then, the least ξ_0 of which confidence interval of the mean surprises contains zero indicates the true order of synchrony.

4.3.4 Impact of non-homogeneous firing rates and pairwise correlations across neurons

So far, to implement the null hypothesis, we have implicitly assumed that all neurons have similar marginal statistics; in particular same firing rate across all neurons and same pairwise correlations across all neuron pairs. However, electrophysiological data often show heterogeneous firing rates and pairwise correlations between neurons. The aim of this section is to investigate the effect of violation of homogeneous assumption on the FP rates of the PUE method. We consider two scenarios: S1) Neuron pairs have homogeneous pairwise correlations but firing rate is heterogeneous across neurons, S2) Firing rate is homogeneous across neurons but neuron pairs have heterogeneous pairwise correlations.

We start with the S1: To mimic the firing rate heterogeneity of the electrophysiological data, we assume a log-normal shape for the firing rate distribution across neurons. This type of distribution has been often considered to explain the observed firing rate in experimental data. To generate a set of spike trains which has a log-normal firing rate distribution across neurons and also incorporates a predefined order of synchrony among them, we use a modified CPP model which is explained in Compound Poisson Process (CPP). In short, it is similar to the conventional CPP model but here we copy the spikes from carrier process with amplitude $\omega = 1$ into a chosen neuron according to a log-normal distribution (in the case of conventional CPP model this distribution is uniform), meaning that a neuron with a higher firing rate will receives more spikes.

Using the modified CPP model we first generate K realizations of a set of spike trains containing a synchrony order $\xi = 4$ and a log-normal firing rate distribution across neurons with a mean $\mathbb{E}[\lambda]$ and a standard deviation $\mathrm{SD}[\lambda]$. The value of SD quantifies the strength of heterogeneity among neurons where $\mathrm{SD}[\lambda] = 0$ resembles the homogeneous case and the larger the $\mathrm{SD}[\lambda]$ the stronger the heterogeneity. Then, we apply the PUE method with as assumed ξ_0 equals to ξ and estimated $\hat{\lambda}$ and $\hat{\rho}$ on each realization and obtain a p-value for each realization. Defining the significance level $\alpha = 0.05$ we compute the FP rate: total number of null hypotheses rejection in proportion to the total number of tests ($= K$). We varies the $\mathrm{SD}[\lambda]$ while keeping the $\mathbb{E}[\lambda]$ constant and repeat the above procedures. Here

it should be noted that even though the empirical data have heterogeneous firing rate distribution, we assume a homogeneous firing rate across neurons in implementing the null hypothesis. The obtained results are shown in Figure 4.10A together with the firing rate distribution for different SD[λ]s. As follows from Figure 4.10A the FP rate is at most equal to the predefined significance level (here $\alpha = 0.05$) and puts forward the claim that the PUE method is robust against heterogeneous firing rate across neurons.

Turning now to the impact of heterogeneous pairwise correlations across neuron pairs on the FP rate of the PUE method, we consider the S2 for the ground truth of the empirical data. To this end we generate a set of N spike trains where a subset of them with n neurons is correlated and the rest $m = N - n$ neurons are independent. By keeping the average pairwise correlation ρ over all pairs (including correlated and independent subsets) constant, we change the n. Note that for $n = N$ the S2 resembles the homogeneous case and for $n < N$ the heterogeneous case, where smaller n corresponds to stronger

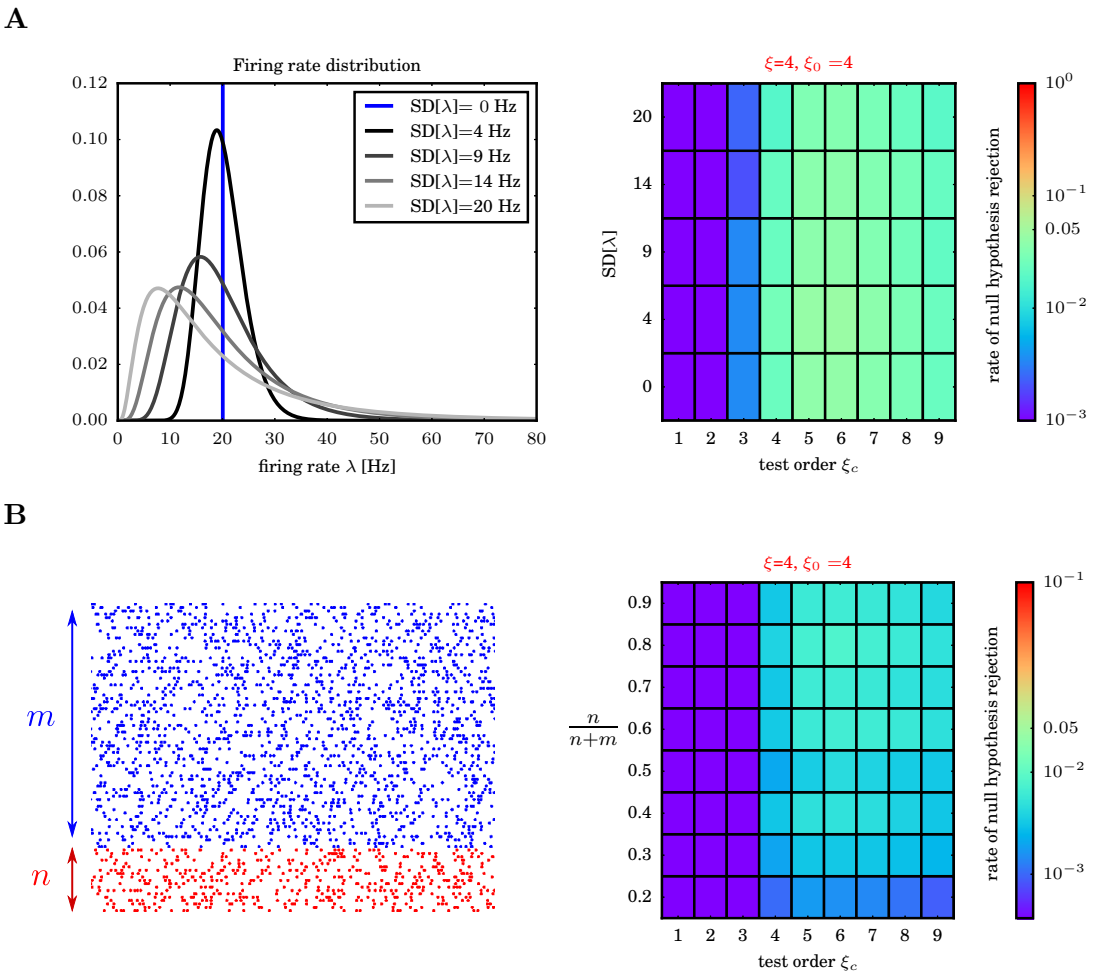

Figure 4.10: *(Caption on the next page)*

Figure 4.10 *(previous page)*: **Non-homogeneous firing rates and pairwise correlations**. (**A**) FP rate of the PUE method applied to synthetic data with inhomogeneous firing rate distribution across neurons (pairwise correlation is homogeneous across neuron pairs with $\rho = 0.005$). The firing rate distribution has a log-normal shape $\ln \mathcal{N}(\mu, \sigma^2)$ which can be characterized by two parameters: location μ and scale σ, where the arithmetic mean is $\mathbb{E}[\lambda] = \exp(\mu + \frac{1}{2}\sigma^2)$ and the arithmetic standard deviation $\mathrm{SD}[\lambda] = \mathbb{E}[\lambda]\sqrt{\exp(\sigma^2 - 1)}$. By changing the SD from 0 (corresponds to homogeneous case; blue line of the left panel) to 20 Hz (corresponds to extremely heterogeneous case; the widest distribution on the left panel), the FP rate (right panel) is calculated for the simulated data (with $\xi_0 = \xi = 4$) following the corresponding firing rate distribution, while the null hypothesis of the PUE is based on the homogeneous assumption of firing rate across neurons (with the mean rate $\hat{\lambda}$ estimated on each realization of the data). (**B**) FP rate of the PUE method applied to synthetic data with inhomogeneous pairwise correlation across neuron pairs (firing rate is homogeneous across neurons). To mimic the inhomogeneous pairwise correlations across neuron pair we simulated a set of N spike trains where a subset of them with n neurons is correlated and the rest $m = N - n$ neurons are independent. By keeping the average pairwise correlation ρ over all pairs (including correlated and independent subsets) constant, we change n to increase the inhomogeneity: $n = N$ mimic homogeneous case. For different n, then, we simulate 1000 realizations of data following the desired pairwise correlation distribution and compuet the FP rate (shown in the right panel).

heterogeneity. The FP rates obtained for different n similar to the S1 case. The results are shown in Figure 4.10B. The FP rates are at most equals to α and with decreasing the n i.e. increasing the heterogeneity, the FP rates decrease. This shows that the PUE method becomes more conservative for the data with heterogeneous pairwise correlation distribution.

4.3.5 Application to experimental data

In this section we demonstrate the application of our method to a massively parallel recording of spike trains from an awake, behaving monkey. The monkey was trained to perform a reach-to-grasp task. The behavioral task protocol is shown in Figure 4.11A. The monkey's task is to reach to an object located in front of the animal, grasp it with either of two alternative grip types, i.e., side grip (SG) or precision grip (PG) (upper left pictures in Figure 4.11A) and hold it with either a high or a low force. A trial is initiated when the monkey holds a switch placed at waist-level (event label: TS). 800 ms after TS, the monkey gets the first cue about the grip type (event label: CUE-ON), which is turned off after 300 ms (event label: CUE-OFF). Then after a 1000 ms delay, the monkey gets the second cue about the force the monkey needs to exert on object, which is also a Go signal to initiate the arm movement (event label: GO). The movement initiation is detected by a

switch release (event label: SR) and varies from trial to trial depending on the reaction time of the animal. Thus, each trial is composed of four behavioral epochs (see Figure 4.11A): pre-cue period (from TS to CUE-ON), cue period (CUE-ON to CUE-OFF), preparatory delay period (from CUE-OFF to GO) and movement period (after GO).

Spiking activities of the primary motor and premotor cortices during the task were recorded by using a 10×10 electrode grid (Plexon Utah array) [245, 279, 340]. The data set that we selected for illustrating the application of our method contains spike trains of 156 well-sorted single neurons (SUA) (*cf.* [279] for details of the experiment and recordings). Data preprocessing has been done as described in [340].

The spiking activities of $N = 156$ neurons in an example trial and its population histogram, computed with a bin width of $\Delta = 5$ ms, are shown in the top panel of Figure 4.11B. We apply our analysis on these data in a time resolved manner using a time window of 300 ms width slid in steps of 20 ms. At each window position we estimate the average pairwise correlation coefficient $\hat{\rho}$ and the average firing rate of all neurons $\hat{\lambda}$ within the window. The temporal modulations of $\hat{\rho}$ and $\hat{\lambda}$ are shown in the second and third panels of Figure 4.11B, respectively. The experimental events that mark the borders of the behavioral epochs are shown with red vertical lines and the delay period is indicated by a gray shade. We apply our test and calculate the surprise with a null order $\xi_0 = 2$ for different ξ_c. The bottom panel of Figure 4.11B shows the temporal modulations of the surprises. In this example trial, the surprises for ξ_c greater than 4 exhibit large values above the 5% significant level during the delay period shortly after the offset of the first cue.

To explore the results shown above across trials, we first employ the cross-trial analysis (explained in Section 4.3.3). Before applying this analysis, we separate the trials into two groups according to the different grip types (SG and PG). Then, separately for the SG and PG trial groups, we align the time axis of each trial to GO for analysis of the periods before GO, or to SR for analysis of the movement period, because a previous study has shown that the temporal modulation of firing rates in the movement period is locked to SR rather than to GO.

The second panel of Figure 4.11C and Figure 4.11D display the significance windows in each trial (same as explained in Figure 4.11B) for the SG and PG trial groups, respectively. The significance count $W_{sum}(t_w, \xi_c)$ (see (Eq. 4.4)) is shown in the bottom panel of Figure 4.11C-D. W_{sum} tends to take larger values during the delay period in both SG and PG trials.

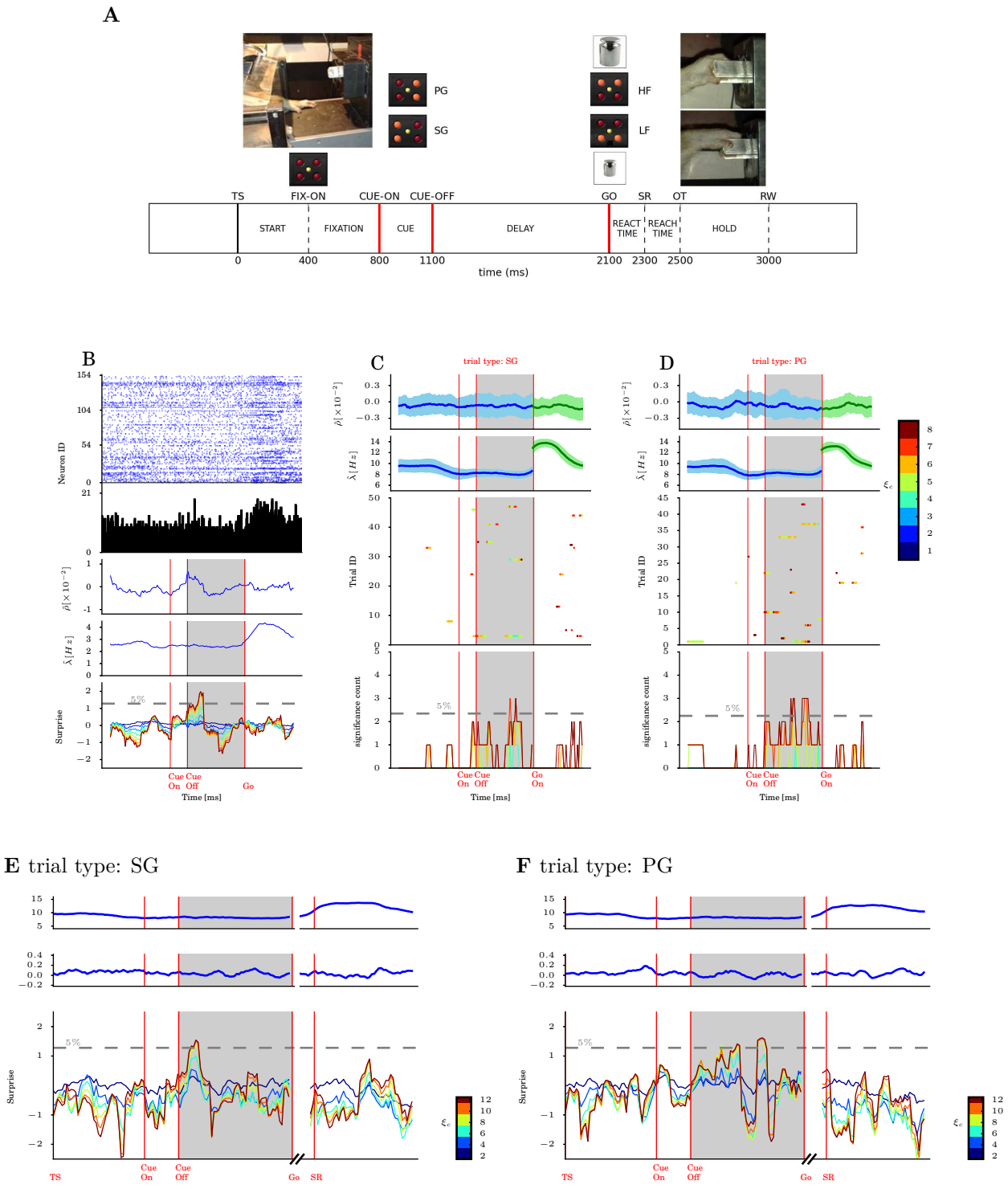

Figure 4.11: *(Caption on the next page)*

Figure 4.11 *(previous page)*: **(Experimental data) Detection of significance higher order correlation in a time resolved manner.** (**A**) Experimental protocol (see text for details)**.** (**B**) *Top panel:* Dot display and population histogram of the one trial in reach-grasp experiment. Population histogram is calculated in a bin size of $\Delta = 5\,\text{ms}$. *Second and third panel:* Estimated pairwise correlation coefficient $\hat{\rho}$ and firing rate, $\hat{\lambda}$ in a sliding window of 300 ms with step of 20 ms. *Bottom panel:* Surprise calculated in a time resolved manner similar to panels two and three. Different colors correspond to different test ordes. Red lines indicate events of the experiment (see A). (**C-D**) Results of cross trials analysis: *First and second panel:* Average pairwise correlation coefficient and average firing rate across all neurons and trials. The period from TS GO signal is aligned on TS and the movement period (i.e. after GO) is aligned on the switch release (SR). *Third panel:* Windows with significance (5%) surprise indicated by dots. In each row surprise is calculated on one trial of the experimental data (the same parameters as A). Different colors correspond to different at least test order where surprise becomes significant. In the other word for a given window, if the surprise is always significance for test order more than e.g. 4, the dot represents only $\xi_c = 4$. *Bottom panel:* Sum over number of significance surprise over all realizations (trials) in a given window. The results of the SG trial type is shown in C and PG is shown in D. (**E-F**) Results of concatenated trial analysis: *First and second panel:* Time modulation of average pairwise correlation and firing rate calculated on the concatenated trials with window size of 100 ms in step of 20 ms for SG in E and PG in F. The trial alignment is TS for the before GO and SR after GO. *Bottom panel:* Modulation of surprise calculated on the concatenated data (similar parameters as E and F). The gray line indicates the 5% significance level.

Next, we employ the concatenated trials analysis as explained in Section 4.3.3. Using a window of 100 ms width slid in steps of 20 ms, we apply the PUE method on the concatenated data at each window position. Figure 4.11E-F show the temporal modulations of the surprises for the SG and PG trial groups, respectively. In the delay period, the surprises for both the SG and PG trial groups reach significant (with $\alpha = 0.05$) values while they remain non-significant in the other periods. For the SG trial group, the significant surprise appears shortly after CUE-OFF and only at one window position, whereas for the PG trial group it appears in the middle of the delay period and at more than one window positions.

Finally we infer the order of synchrony in the WOIs as explained in Section 4.3.3. Figure 4.12 shows the results of the inferences, similar to Figure 4.9, for the SG and PG trial groups. This analysis suggests the inferred true synchrony order of 6 for the SG trial group and 7 for the PG trial group (Figure 4.12B and D). For the cross-trial analysis, the average surprise and the significance count, which were introduced in Section 4.3.3, are shown in Figure 4.12A for the SG trial group and in Figure 4.12C for the PG trial group. The average surprise infers the synchrony order of 2 for both of the SG and PG trial groups. The significance count infers the synchrony order of 5 for the SG trial group and 7 for the PG trial group.

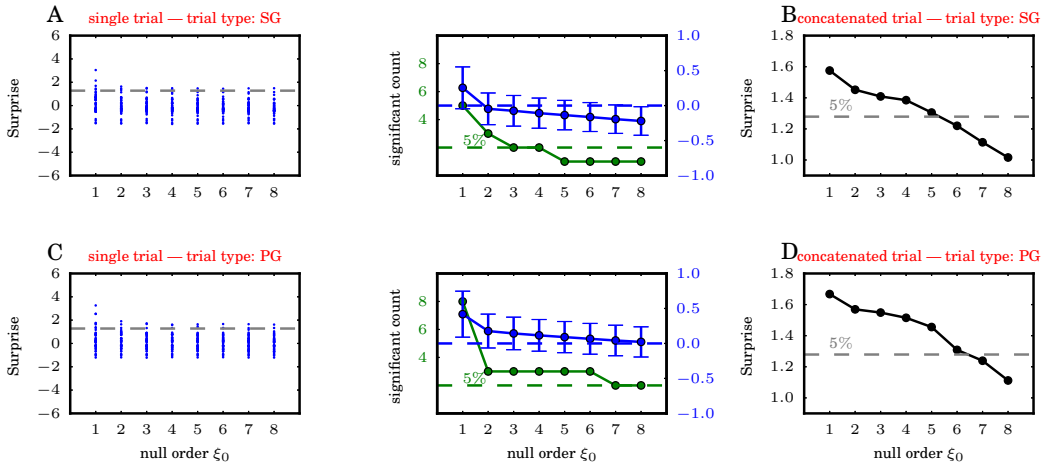

Figure 4.12: **(Experimental data) Extracting the order of synchrony.** (**A, C**) Similar to Figure 4.9A for experimental data with SG trial type in A and PG trial type in C. The WOI for each trial type has been chosen from Figure 4.11E-F where the surprise becomes significant (the chosen window is the one with largest value of the surprise). (**B, D**) Similar to Figure 4.9B for experimental data with SG trial type in B and PG trial type in D. The WOI is the same as A for B and the same as C for D.

4.4 Summary

In this chapter, we have presented the population unitary events (PUE) method as a novel statistical tool to assess higher order synchrony in MPST. The method was carefully designed to construct, on the one hand, a null hypothesis beyond independence by incorporating the correlations (pairwise and higher order) among MPST, and on the other hand to be applicable to small-sampled data (e.g. 100 ms duration) with a reasonably high test's power. In an extensive calibration of the PUE method with synthetic data which mimicked the real case scenarios of the electrophysiological recordings, we demonstrated the good performance of the method in terms of false positive (FP) and true positive (TP) rates. Furthermore, we showed that the method is able to capture the temporal modulation of higher-order synchrony in MPST, and can infer the underlying order of correlation. Finally, we demonstrated the application of the PUE to MPST recorded from the motor cortex of monkey during a delayed reach to grasp task. As a result we detected the functional higher-order synchrony locked to delay-period when monkey is expecting a GO signal to appear.

III

Discussion

In this thesis, we propounded the importance of segregating correlations which arise merely as a natural consequence of network connectivity (baseline correlations) from those with a functional role (functional correlations) when studying the relevance of correlations in cortical information processing. We demonstrated that this segregation becomes more and more crucial in increasingly larger parallel recordings of neurons, which are available nowadays through modern electrophysiological recordings. We discussed the indispensable prerequisite of advanced statistical tools, which on the one hand enable to distinguish baseline and functional correlations, and on the other hand can cope with massively parallel spike data and the complexity thereof. Toward addressing these needs, we studied in detail the aptness of two commonly used methods in statistical neuroscience, i.e. the Unitary Events method [135, 136, 139] and the maximum entropy model [175, 297, 312], and discovered that these methods are prone to become impracticable and might yield a wrong interpretation of the data when applied to massively parallel spike trains (MPST). To overcome the limitation of these methods and handle the complexity of high-dimensional data, we presented two novel statistical methods for analyzing MPST and studying the functional role of correlations in cortical information processing.

The first method is Inhibitory Maximum Entropy (IME) which incorporates in a simple way the effects of neural inhibition into the classical pairwise maximum entropy model. This method can therefore be quite useful as a realistic hypothesis to model the baseline correlations among the activity of an ensemble of neurons.

The second method is Population Unitary Events (PUE) which was developed to handle MPST in search for higher-order synchrony in a time-resolved manner, and also to incorporate a null-hypothesis beyond independence. This method was extensively calibrated through application to test data covering the typical features of the electrophysiological recordings such as inhomogeneous firing rate across neurons and trials, or inhomogeneous pairwise correlations across neuron pairs. Applied to experimental data recorded from the motor cortex of monkey, this method proved to be very useful in the search for the functional role of correlations, thanks to its accounting for baseline correlations.

The following sections will summarize our results, and outline the future projects that can be initiated based on our studies.

Reproducibility of Riehle et al (1997)

In Chapter 2, we started by inspecting the UE method by replicating the results of [278] for two purposes: i) to provide an open source implementation of this method, and ii) to ensure that the original research is reproducible. By providing a new reimplementation of the UE method in Python and applying this method to the original data of [278], we demonstrated that we are able to reproduce the original results of [278].

The UE method involves a number of numerical computations and is very sensitive, as we showed in Chapter 2 by the comparison of Figure 2.1 and Figure 2.2, which differed in the events to which trials were aligned. This difference in the alignment would not make any difference in the results if the time differences between the two events (PS and RS) were identical across the trials. But this was not the case, due to hardware features of the recording setup (as we learned from the first author of the original publication); therefore the binning of the data started at a slightly different time points in different trials. This likely led to a loss or an addition of a spike in a bin and thus to a small difference of the number of spike synchrony events (see also the discussion on the issues of exclusive binning in [137]). Nevertheless, this sensitivity of the method is a strong indication that our reproductions of Figure 2 and Figure 4A of the original publication are correct and that our new implementation of the analysis is faithful to the UE method.

The event to which the data were aligned, and the cut time which defined the start of the (exclusive) binning, are not documented in the original publication. The original scripts for the analysis could have revealed this information, even without having the original UE software code at hand, but they are not available anymore either. Thus due to the lack of documentation we were only able to reproduce the results of [278] by communicating with some of the authors of the original publication.

The reproduction of Figure 4A of the original publication is a further and important test of whether our reimplementation is also correct for $N > 2$ neurons. This is relevant since the implementation for $N > 2$ requires a more generic, complex algorithm for the analysis than the $N = 2$ case. The reason is that in the case of two neurons there is only one pattern type which has to be analyzed (i.e. [1,1]), but in the case of e.g. 3 neurons there are already 4 different spike patterns to analyze ([1,1,0], [0,1,1], [1,0,1], and [1,1,1]), and exponentially more for more neurons ($2^N - N - 1$). Thus the statistics of each of the patterns is performed separately and the bookkeeping needs to be carefully done.

There are also other reasons why the reproduction of Figure 4A is more complicated than that of Figure 2. First of all, the data were not available to us. After requesting them from the first author of the original publication we received some data but were not able to reproduce the result - both in terms of the UE results but also the data seemed slightly different. After further consultation with the original authors we learned that the original Figure 4A was not generated by the Matlab implementation used for Figure 2 but by another implementation in IDL. Neither is available to us. However, we were told that the third author of the original publication performed a thorough comparison of the two implementations at the time and the final report on that was made available to us. This enables us to define the correct workflow that reproduces the original result, given we have the data in the correct resolution at hand.

The reimplemented UE analysis software contains extensions for improving the statistics that were developed after the original publication. First, it contains the option to

adjust the statistics to take into account cross-trial inhomogeneity, by calculating the number of expected spike synchrony events based on the firing rates on a trial-by-trial fashion (option: *"analytic_TrialByTrial"*), as suggested in [139], in contrast to using trial averages of the firing rates. Second, our reimplementation offers the possibility to calculate the significance of the empirical number of spike synchrony events based on a Monte Carlo approach (option: *"surrogate_TrialByTrial"*). Instead of computing the significance using a parametric distribution based on the estimate of the firing rates, the null-hypothesis of independence is implemented by surrogate data [140, 141, 223]. By repeated intentional manipulation of the original data, potential spike synchrony is deleted. Each of these surrogate data created by this procedure are then searched - as the original data - for spike synchrony, and these numbers create the distribution underlying the significance test of the method. Obviously this version is much more computationally expensive than the parametric approach used here for the reproduction of [278] and we are currently working on an HPC implementation to make use of parallelization. The Python implementation of the UE method is publicly available in the open source software package Elephant at http://neuralensemble.org/elephant/.

If the authors of the original paper had not been accessible, we would not have been able to reproduce the results. Nevertheless, here our final validation of the reproduction is based only on a detailed visual comparison of the original and reproduced figures. In an optimal scenario, we would be able to exactly validate the results based on a numerical comparison. To do so, all of the following pieces of information ought to be available:

1. the original primary data
2. metadata describing the primary data in detail
3. the original statistics software package (e.g. Unitary Events)
4. the loading routine for the data
5. all specific code required to produce each figure of the original publication
6. detailed documentation of all code
7. the original software environment (with programs available in the original versions used), including, e.g. the interpreter/compiler (here: Matlab) and operating system
8. unique identifiers of the data records that are used to unambiguously identify data from within the analysis code

In the analysis presented in this work, not even the original primary data (1), recorded more than 20 years ago, is available, but only a slightly preprocessed version of it. However, even today many of the pieces of information listed above are often not made available by scientists. This is due in part to the enormous complexity of the task to record all

information in fine detail, leading from the experiment to a final analysis. Moreover, there is still a lack of software tools to support researchers in the process of acquiring, storing, and organizing this information. Currently, there are emerging approaches suggested for metadata annotation (2) of electrophysiological data, such as the odML[1] framework (see, e.g. [131, 362]) for storing hierarchical collections of metadata or the NIX[2] data format [7] for linking data and metadata. In our concrete example, the information about the hardware limitations in storing the event times, would have been essential information contained in the metadata. Using modern tools for version control, points 3-6 can be easily addressed. There are emerging approaches to keep software environment, i.e., the original Matlab version and the operating system (7), e.g., by freezing the environment in a virtual machine. Point 8 is still challenging, because it requires the data to be addressed in an unambiguous manner from within the analysis scripts. Including data within the code repositories is typically prohibitive due to the size of the data. A solution would be to deposit data in public or private databases that allow them to be tagged by a unique identifier in combination with a tool to generate a detailed provenance track of the analysis process, but the implementation of tools and services for the workflows used in data analysis of electrophysiological data is still an ongoing endeavor [13, 63]. In summary, there are still components missing such that researchers are put into a position to build complex data acquisition and analysis workflows that enable optimal reproducibility in neuroscience.

Correlations between cortical neurons

Having implemented the UE method, to test for systematic occurrences of UEs across data sets and neuronal sampling, in Chapter 2 we applied the extended UE method, proposed in [196], to data recorded from motor cortex of monkey during a center-out reaching task. We observed that: i) the obtained surprise exhibits a slight constant positive offset, and ii) the modulation depth of the surprise is the larger, the more data (neuron pairs) are included in the analysis. The natural question was whether these observations were features of network processes, or the result of false positives due to a not yet identified flaw in the UE analysis. The latter explanation was excluded by testing the UE analysis with synthetic data: the analysis of an increasing number of pairs does not lead to a larger surprise (i.e. false discoveries) when pairs are sampled from independent spike trains. However, when the spike trains are correlated, the UE analysis results in larger surprise. This has a clear explanation: the more pairs are considered, the more evidence of the correlation enter the analysis, resulting in a larger surprise, which is a correct and desired property of the UE analysis.

[1] `https://github.com/G-Node/python-odml`
[2] `https://github.com/G-Node/nix/wiki`

To explain the positive offset we referred to theoretical studies where networks of connected neurons are shown to exhibit correlations owing to the shared input or recurrent connectivity. By simulating such a network (a random balanced network without any function) and applying the extended UE method to the obtained spike trains, we demonstrated the presence of the positive offset. Thus, we interpret such a positive offset as baseline correlations, since they occur independent of correlations with functional role in the network (functional correlations).

To summarize, the co-occurrence of action potentials on a fine time scale (range of milliseconds) has been shown in cortical-neuron activity time-locked to a behavioral task [196, 278]. On the one hand, such synchrony is thought to result from information processing in the network related to function. On the other hand, low spike correlations between pairs of neurons have been shown to arise as a direct consequence of shared input in network models with inhibitory feedback [159, 332]. We refer to these types of activity as functional and baseline correlations, respectively.

- **Functional Correlation:** Time-dependent correlation due to task-related network function.
- **Baseline Correlation:** Time-independent correlation due only to network structure (connectivity).

Baseline correlations are extensively studied in theoretical works. In experimental data there has been no evidence of such correlations yet, since until recently parallel recordings have been limited to only a small number of neurons. Recent technical advances now enable massively parallel recordings of cortical neurons. Only here baseline correlations become visible, due to the large sample size.

Maximum entropy model and bimodality problem

In Chapter 2 we provided evidence for baseline correlations among cortical neurons recorded from motor cortex of monkey. This, in line with theoretical studies, emphasizes the importance of baseline correlations in MPST and calls for statistical tools which take them into account in the construction of null hypotheses for the search of functional relevance of correlations. One of the available tools which can serve this purpose is maximum-entropy models. They predict the activity of neuronal populations by taking into account of the observed mean firing rates and correlations (pairwise and higher-order). Thus, in Chapter 3, we employed some of these models for constructing a null hypothesis beyond independence and examined their aptness in application to MSPT.

The obtained results showed that the pairwise model, applied to experimental neuronal data of populations of hundreds of neurons, is very likely to give a *bimodal* probability distribution for the population-averaged activity. We have provided evidence for this claim

starting from an experimental data set and then looking at summarized data from the literature. The first mode of the bimodal distribution is the one observed in the data. The second mode (unobserved) can appear at very high activities (even 90% of the population simultaneously active) and its height increases with population size. The mechanistic origin of the second mode is simple and generic: Positive pairwise correlations imply positive Lagrange multipliers that couple pairs of units. This mutual positive feedback, given sufficiently large numbers of units, may become strong enough to sustain a second mode of high activity. The mechanism in hence identical to the transition to the ferromagnetic state of the Ising model.

The presence of a second mode is problematic for several reasons:

- As far as we know, a second mode has never been observed in experimental recordings, and surely not at high activity – data in which 180 out of 200 neurons spike simultaneously are unheard of. So it is an unrealistic prediction of the pairwise model.

- Above certain population sizes the second mode cannot be dismissed as too small to be recorded, because it becomes more pronounced as N increases, and the minimum that separates it from the main mode becomes shallower.

- The Boltzmann-learning [4, 37, 161] procedure based on asynchronous Glauber dynamics [111, 225, chap. 29], used to numerically construct this model, becomes practically non-ergodic – it can already be so for population sizes of roughly 50 neurons – so that the Lagrange multipliers of the pairwise model are difficult or impossible to find. Approximate methods like mean-field [153, 252, 257], Thouless-Anderson-Palmer [257, 334], Sessak-Monasson [300, 301] also seem to break down in this case.

- The Glauber dynamics based on the pairwise model jumps between two metastable regimes, remaining in each for long times (owing to its asynchronous update) and cannot be used to generate realistic surrogate data.

- The fact that the position and height of the second mode vary with N contradicts the natural assumption that the recorded N neurons are a "random sample" of a larger population (see Appendix 5.2.4). From the probability calculus we know that the population-average distributions of a full population and a "random sample" from it should have maxima at roughly the same relative heights and locations, since they are connected by convolution with a hypergeometric distribution [83, ch. II][286, ch. 4][182, ch. 3][107, 358–360].

Eliminating the second mode also eliminates all these problems.

We gave an intuitive explanation of why the second mode appears: because the pairwise model, given positive pairwise correlations, corresponds to a network that is

excitatory on average and symmetric. And symmetric connectivity is incompatible with the presence of a subset of neurons that have an inhibitory effect, but receive excitatory input (see Section 3.3.2). This explanation also suggested a way to eliminate the second mode: by adding a minimal asymmetric inhibition to the network, in the guise of an additional, asymmetrically coupled inhibitory neuron (Figure 3.7A).

This idea led to the construction of an "inhibited" pairwise maximum-entropy (IME) model $P_{\mathrm{i}}(\boldsymbol{s})$, (Eq. 3.22), with the important properties:

- It is a maximum-entropy or minimum-relative-entropy model.
- It is the stationary distribution of a particular asynchronous Glauber dynamics with pairwise couplings.
- Its Lagrange multipliers can be found via Boltzmann learning.
- Its parameters can be chosen to have the main mode only.
- It is numerically equal to the distribution one would obtain from a non-ergodic Boltzmann learning.

Inhibited maximum entropy (IME) model

In Chapter 3, we have shown that bimodality is a problem in the application of the pairwise maximum-entropy model. We wish to stress, though, that the presence of bimodality and non-ergodicity can easily go unnoticed. Sampling from a bimodal distribution, the probability to switch to the second mode may be so small that is practically does not occur. If the high mode is not visited during neither Boltzmann learning, nor while surrogate generation, we face the most subtle situation. The obtained distribution is then formally not the assumed maximum entropy distribution (Eq. 3.3), but rather the inhibited maximum entropy distribution (Eq. 3.22). But otherwise all results seem consistent. Results obtained with non-ergodic Boltzmann learning therefore do not correspond to the pairwise maximum-entropy model. Still, they can be reinterpreted as correct results for the IME model, with appropriately chosen Λ_{I} and θ. This is relevant for any work in the literature that may unknowingly have been affected by non-ergodicity.

Obvious inconsistencies appear as soon as the system visits the second mode during either of these phases. This breakdown may appear at random points in time and is hence not predictable. Using the original pairwise model (Eq. 3.3) therefore requires a check for the presence of bimodality and non-ergodicity in a case by case way. For example, by starting the sampling from different initial conditions, at low and high activities, looking out for bistable regimes [212, § 2.1.3]. Another way out of this problem is to use other sampling techniques or Markov chains different from the Glauber one [30, 212, 225]. Alternatively, one may use the IME model (Eq. 3.22) with the standard approaches.

In the presence of inhomogeneous and randomly chosen parameters, the obtained distribution is mathematically similar to the Boltzmann distribution of the Sherrington & Kirkpatrick infinite-range spin-glass [197, 308] (see Appendix 5.2.5). A more systematic analysis of the effect of inhomogeneity on the appearance of the second mode could therefore employ methods developed for spin glasses [87]. This way, one could also arrive at approximate expressions for the inverse problem, the determination of Lagrange multipliers form the data. One may think of modifying the Thouless-Anderson-Palmer (TAP) mean-field approach [290, 334], generalizations of which exist form the asymmetric non-equilibrium case [291] appearing here due to the inhibitory unit. An approximation based on the ideas of Sessak and Monasson [300, 301] may be an alternative. Another possibility is the use of cumulant expansions [39, 110] have the advantage of being valid also in regimes of strong coupling, as opposed to TAP-based approaches, and recent extensions exist that allow us to obtain the statistics on the level of individual units [54].

Different readers will draw different conclusions from the presence of bimodality. Some may dismiss or abandon the whole pairwise model as flawed. Some may still want to use it, bimodality notwithstanding. Some may look for other maximum-entropy-inspired alternatives. We have presented *one* (as opposed to *the*) such alternative: the IME model P_{i}, (Eq. 3.22). It is an interesting alternative for at least two reasons.

First, the IME model P_{i} is stationary under a Glauber dynamics with *pairwise* couplings. Consider that pairwise models with additional constraints are stationary under Gibbs samplers with higher-order couplings – and thus lose some of their analogies with real neuronal networks.

Second, the inhibited distribution P_{i} incorporates the effects of neural inhibition in a simple way. These effects are represented by the reference or prior probability $P_0(\boldsymbol{s}|\, \Lambda_{\mathrm{I}}, \theta)$, (Eq. 3.25).

The IME model P_{i} includes Shimazaki's model [310] and its "simultaneous silence" constraint as the limit $\Lambda_{\mathrm{I}} \to -\infty$, $\theta = 1/N$. Because of this limit, Shimazaki's model has a sharp jump in probability when $\bar{s} = 1/N$. Their constraint uniformly removes probability from $P(\bar{s} > 1/N)$ and assigns it to $P(\bar{s} = 0)$. In contrast, the IME model P_{i} only presents a kink when $\bar{s} = \theta$, with a discontinuity in the derivative proportional to Λ_{I}.

Some readers may actually object to the usefulness of the inhibited distribution P_{i} exactly because it is derived from a particular prior via minimum-relative-entropy, and may thus appear less "non-committal" or less "unstructured" than a "bare" maximum-entropy one. We briefly countered this argument by pointing out that bare maximum-entropy can be quite "committal", and that reference priors can correct that.

At first sight one may be tempted to interpret the maximum-entropy method to give the maximally non-committal probability distribution consistent with the given information. But there are many qualifications behind this statement, especially behind the terms "non-committal" and "information". The term "information" does not mean only "experimental

data": it also means knowledge of the specific problem and resulting *assumptions* on the variables describing the problem. The way we set up a maximum-entropy problem implies many underlying assumptions, already before experimental data are taken into account [172, 180].

A concrete assumption underlying the bare maximum-entropy principle applied to neuronal activity is that *the recorded neurons should not be a sample from a larger population.* To be more precise: consider N neurons from a population of $R > N$ neurons. If the N activities are modelled by a maximum-entropy distribution, then the R activities *cannot* be modelled by a maximum-entropy distribution. Vice versa, if if the R activities are modelled by a maximum-entropy distribution, the marginal distribution for the N activities cannot be a maximum-entropy one (see Appendix 5.2.4). This fact is known in statistical mechanics [84, 232, 298]. This N-dependence implies a lack of robustness of our inferences under augmentation of data. If we acquire data for and additional set of M neurons, besides the N we already have, and reapply maximum-entropy to the $N + M$ activities, the marginal distribution for the original N will change, and some statistical features of it – e.g., surprising moments or correlations, distribution tails, "critical points", "heat capacities" – will disappear or be different. This N-dependence demands a lot of care in assessing the meaningfulness of statistical features deduced via maximum-entropy on a sample; a similar remark was already made by Good in the 1960s [114, p. 379]. This N-dependence shows that the bare maximum-entropy principle is far from "non-committal" or "unstructured".

The "committal" nature of bare maximum-entropy also appears in its derivation from the probability calculus. This derivation requires a particular prior [51, 52, 172, 180, 270], but one could use other, quite natural priors (e.g., the "flat prior over probability distributions" considered by Bayes [23, Scholium] and Laplace [213, p. xvii]) and the result would *not* be a bare maximum-entropy distribution.

Reference priors can in some cases correct such implicit assumptions. For example, consider a pair of neurons with binary states s_1 and s_2. Without "experimental data", the maximum-entropy principle assigns a uniform probability of 1/4 to each of the four possible joint states (s_1, s_2). The probability assigned to the total activity, $S \coloneqq s_1 + s_2$, is therefore *not uniform* (2/4 probability to $S = 1$ and 1/4 probability to the remaining two values). If we apply the maximum-entropy principle to the total activity S directly, instead, it gives a *uniform* probability of 1/3. Both applications of the principle are consistent, but they use different assumptions about the structure of the biophysical problem. The information implicit in the first application can be specified in the second by using the minimum-relative-entropy method with a non-uniform prior distribution assigning 2/4 probability to $S = 1$. (Something analogous happens in statistical mechanics with the probability distribution for energy, in which a "density of states" term multiplies the

Boltzmann factor). Some implicit assumptions, however, like the sampling assumption previously discussed, cannot be corrected by reference priors.

The necessity of reference priors, reflecting deeper assumptions, is well-known in maximum-entropy image reconstruction [318, 354], for example of astronomical sources [319, 354]: as Skilling remarked, "bare maximum-entropy is surprised to find isolated stars, but astronomers are not" [180].

An analogous remark can be made in our case: bare maximum-entropy is surprised to find so many inactive neurons, and it tries to make some more active ones by creating a second maximum, if that does not break the constraints. But neuroscientists are not surprised at inactive neurons. Bare maximum-entropy assumes that we have abstract "units" whose states are symmetrically exchangeable. But neuroscientists know that these units are *neurons,* whose individual and collective properties are *asymmetric* with respect to state exchanges, for biophysical reasons. The prior of the IME model P_{i} reflects this asymmetry. It is fortunate that we can partially correct the symmetry assumption of bare maximum-entropy by using a prior, without having to overturn our whole space of variables.

The long argument above shows, we hope, that the IME model P_{i} and its reference prior do not break the "non-committal" nature of the maximum-entropy principle; rather, they prevent maximum-entropy from committing to unrealistic assumptions. The IME can therefore be quite useful in all applications of the maximum-entropy model mentioned in Section 3.1. For example, as a realistic hypothesis against which to check or measure the prominence of correlations in simulated or recorded neural activities, to separate the baseline correlation from the potentially behaviourally relevant departures thereof (functional correlation). The surprise measure to effect such separation would, according to the IME model, take into account the presence of inhibition and the overall baseline correlations that are natural in the cortex. As a second example, for the generation of surrogate data, which would also include the natural effect of inhibition besides the observed level of pairwise activity. As a third example, in the study of the predictive sufficiency of pairwise correlations as opposed to higher-order moments, for example for distribution tails [21, 33, 49, 96, 105, 126, 227, 228, 230, 235, 293, 297, 310, 312, 338]. And as a final example, in the characterization of dynamical regimes of neuronal activity [247, 337–339]. Some of the maximum-entropy applications above could suffer from the N-dependence problem for small samples, thus requiring the use of considerable sample sizes. The IME model becomes therefore increasingly more useful, since the problematic bimodality of the bare pairwise model is more pronounced at large sample sizes.

Population Unitary Events (PUE): a novel method to study baseline and functional correlations

The major difficulty of detecting task, stimulus related significant synchrony (what we referred to as functional correlation) among MPST is that the statistical test should carefully account for the chance coincidences which arise as the results of phenomena unrelated to tasks and stimuli, e.g. contribution from firing rate of individual neurons or consequence of connectivity between neurons (what we referred to as baseline correlation). These phenomena become more and more visible in larger populations of neurons and call for a null hypothesis beyond independence in the statistical test. For example, the IME model introduced in Chapter 3 is one of these methods to implement a null hypothesis beyond independence. However, the IME method requires sufficiently large samples of data in order to make reliable inferences about baseline correlations. As a consequence, IME cannot be employed to assess the time modulation of synchrony (in the "sliding window" fashion), and is therefore unable to relate transient, short-lasting synchrony to the time course of a behavioral task. Such relation, however, is a signature of cell assemblies (see Section 1.4.1). To overcome this limitation we introduced the Population Unitary Event (PUE) analysis in Chapter 4.

A powerful aspect of the statistical test designed for the PUE method – besides the null hypothesis beyond independence – is a novel test statistic which allows the application of the method on small-sample data. To explain how this is achieved, consider the cell assembly picture. In this picture synchronous activity of neurons is considered to have a functional role related to the behavioral task, whereas the rest of the recorded spikes are the results of so-called "ongoing" activity, which we refer to as "background noise" (see Figure 1.5). The test statistic (Eq. 4.1) of our method includes a free parameter ξ_c which can reduce the background noise. This is achieved by a systematic increase of ξ_c: For $\xi_c = 2$ the test statistic, being the sum over the binomial coefficient $\binom{C^k}{\xi_c}$ (with C^k the total number of spikes at bin k), excludes those bins with less than two spikes, because $\binom{C^k}{\xi_c} = 0$ for $\xi_c > C^k$. For $\xi_c = 3$ bins with less than three spikes are excluded and so on. This systematically decreases the contribution of bins with lower-order synchrony in the analysis. Note that the bins with lower-order synchrony which are excluded from the analysis might contain spikes from cell assemblies, but it has been shown (see e.g. [340]) that the occurrences of cell assemblies in MPST are rare in comparison to all the recorded spikes (background noise). Therefore, the majority of excluded bins contains the background noise.

The performance of the PUE method was examined in terms of false positive (FP) and true positive (TP) rates by applying it to synthetic MPST with known ground truth of the correlation order. We showed that the method is able to detect the underlying order of correlation by varying the null order for a given test statistic. Further, the

method has a high degree of specificity and sensitivity, as well as high tolerance against relatively small sample sizes, which enables its time-resolved application to capture the temporal modulation of the correlations. This is particularly important, because to detect the signature of cell assembly one needs to test the detailed structure of the transient higher-order correlations in time and in relation to the behavioral paradigm.

The null hypothesis of our method explicitly assumes the CPP as a model behind observed parallel spike trains. Such model-driven approaches have a serious consequence on the test results: the rejection of null-hypothesis could come from the disagreement of the predefined model rather than the actual parameters in the null hypothesis. Therefore, each of the critical assumptions behind the model should be carefully considered when interpreting the test results. In our framework we have the following assumptions: First, we assume a stationary CPP model where each neuron exhibits constant firing rate in time. This assumption may appear to be a strong constraint, given that the activities of cortical neurons have been reported to show time modulation, owing to transient input related to internal cognition or stimulus. However, with an extensive calibration of our method we have shown that PUE yields reliable results even for small sample sizes, of e.g. 100 ms. We can therefore apply PUE to non-stationary data using the sliding-window approach suggested in [136], where we first cut the data in quasi-stationary segments and then apply the method on those segments. Second, the CPP model assumes Poisson statistics for the spikes of carrier process. This implies that we explicitly model the population activity of neurons as a Poisson process. It has been shown that the spiking activity of cortical neurons is less variable than for a Poisson process [8, 249, 251]; the latter has therefore been criticized as a model of single-neuron spiking. This criticism, however, might not hold for the population activity of neurons, achieved by pooled spike trains across neurons. From a mathematical point of view, the pooling of sparse point processes has Poisson limits under sufficient conditions, e.g. large number of processes [55, 132]. Although we are not aware of experimental studies on realistic distribution of the pools of spike trains in large populations of cortical neurons (hundreds or more), we expect that the Poisson assumption of pools of spike trains should be valid on MPST, and therefore is not a limitation of our method. Third, the CPP model can only generate independent or positively correlated spike trains. Consequently, the null hypothesis of the PUE method cannot incorporate negative correlations among parallel spike trains, which makes the method impractical for spike data with average negative pairwise correlations. Nevertheless, reported correlations among cortical neurons from diverse brain areas (see [49] for a review) are often positive, which makes the mentioned limitation of our method unimportant.

Finally, we demonstrated our method by its application on MSPT recorded from motor cortex of a behaving monkey in a reach to grasp task. The method showed the occurrence of higher order synchrony during the delay period in which the monkey is expecting a GO signal for the task. This finding from MPST is consistent with aprevious

study of time-locked synchrony related to behavior [278], which was carried out on a limited number of neurons. Of course, further analysis is required to check the consistency of this finding across different recordings as well as different animals.

Future work and outlook

In this thesis, we presented two novel statistical methods which can be employed in studying the functional relevance of correlations in MPST beyond baseline correlations, which are ubiquitous in the brain owing to the connection between neurons. Rapid advances in electrophysiological recordings call for these methods to shed light on the long-standing questions about the mechanisms behind cortical information processing. Nevertheless, this thesis is a part of continuous endeavors to provide appropriate statistical tools to make sense of a progressively growing amount of available data. In the following we sketch an outline of possible future projects that arise from this thesis.

The proposed methods, PUE and IME, can be readily applied to the huge amount of spike data which are today being recorded from diverse brain areas and diverse species. To make this method accessible to the whole community from different countries, one natural next step would be to provide a publicly available implementation of these methods.

As a follow up study of IME, we are continuing to explore maximum-entropy methodologies, but in the broader context of Bayesian theory (as it happened in astrophysics). This allows for more flexibility and understanding, and also enables us to ask and answer in a precise way model-comparison questions, like "does knowledge of higher-order correlations lead to better predictions?", or "which prior leads to best predictions?", as explained in [224].

For example, a pairwise maximum-entropy distribution corresponds to two possible large-sample limits in Bayesian theory: a limit model by *sufficient statistics*, i.e. statistics of the sample which are hypothesized to be sufficient for making inferences about a particular parameter, or a limit *exchangeable* model with a particular prior. The more general Bayesian point of view allows us to extend the pairwise model in two ways: "modulating" it with a prior distribution, and considering the non-limit case of small data samples. It also allows us to use this method to make predictions or test the hypotheses behind the model. In the broader context of Bayesian theory it is also immediately evident how to use the model in time-dependent cases. Our next project in fact considers a time-dependent "model by sufficient statistics".

The PUE method has been calibrated extensively on artificial test data to be robust against the typical non-stationarity observed in the activity of cortical neurons, such as firing rate non-stationarity in time and across trials. In the application of this method to MPST recorded from motor cortex of monkey in a reach-to-grasp task, we have observed higher order synchrony, time-locked to the delay period when monkey is preparing his

movement and expecting a GO signal to appear. This finding is in line with previous observations of functional synchrony in [196, 278] which were studied on small number of neurons. However, the next step will be to verify our results by applying the PUE method to different data sets recorded in different sessions and from different monkeys.

IV

Appendices

CHAPTER 5

Supplementary materials

This chapter contains the supplementary materials related to each chapter of this thesis.

5.1 Appendix to Chapter 2 “Baseline and functional correlations”

5.1.1 Compound Poisson Process (CPP)

A compound Poisson process (CPP), also known as marked Poisson, is a time-continuous doubly stochastic process with amplitudes (or jumps). It is composed of N simultaneous Poisson processes $x(t) = (x_1(t), ..., x_N(t))$, all having the same marginal rate parameter λ that is constant over time, and coupled with each other by a given correlation structure. It is parameterized by two parameters: firing rate of Poissonian carrier process λ_{cp} and amplitude distribution $f_A(\omega)$. Using this parameterization, we can generate parallel spike trains with the CPP. To this end we need two steps: first, generate a Poisson process with rate λ_{cp} and draw for each of its events at time t_j a corresponding amplitude ω_i from the amplitude distribution. In a second step, each event at t_j is copied to ω_i individual process, where the precise process IDs is determined by another distribution P_λ (see Figure 5.1). The distribution P_λ can be, in the simplest case, a uniform distribution which will results in a homogeneous rate across all processes, or it can have any arbitrary shapes to mimic inhomogeneous rate across population.

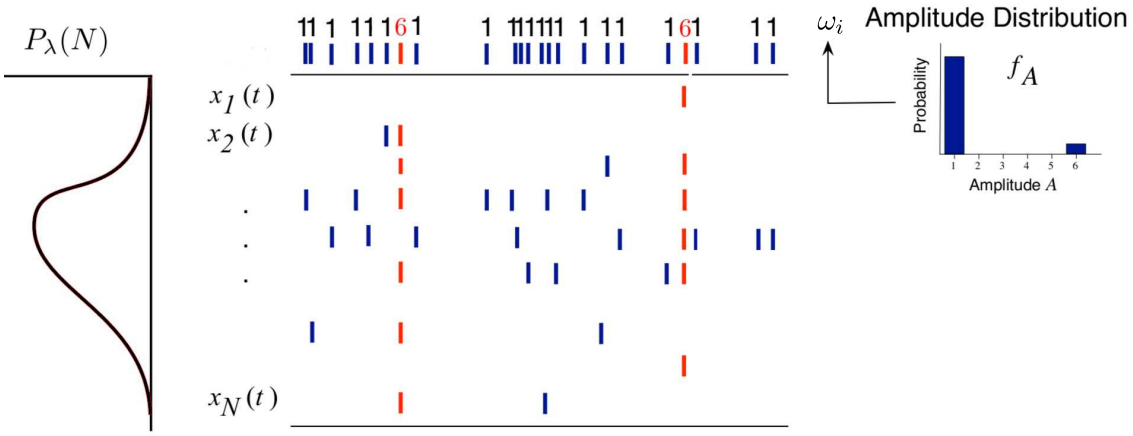

Figure 5.1: Scheme of the compound Poisson process with an arbitrary firing rate distribution across population.

5.2 Appendix to Chapter 3 "Inhibited maximum entropy (IME) model"

5.2.1 Definition of Glauber dynamics

In this section we show that there is a temporal process that is able to sample from the the distribution $P_{\mathrm{p}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda})$ (Eq. 3.3). This temporal dynamics is called *Glauber dynamics*. It is an example of a Markov chain on the space of binary spins $\{0,1\}^N$ [111]. At each time step a spin s_i is chosen randomly and updated with the update rule

$$s_i \leftarrow 1 \text{ with probability } F_i(s) = g(\sum_j \Lambda_{ij} s_j + \mu_i) \text{ and 0 else} \tag{5.1}$$

$$g(x) = \frac{1}{1+\exp(-x)}, \tag{5.2}$$

where the coupling is assumed to be symmetric, $\Lambda_{ij} = \Lambda_{ji}$, and self-coupling is absent, $\Lambda_{ii} = 0$. The transition operator of the Markov chain, κ, only connects states that differ by at most one spin, so for the transition of spin i we can write, if $s^{i+} = (s_1, \ldots, \underbrace{1}_{i-th}, \ldots, s_N)$ and $s^{i-} = (s_1, \ldots, \underbrace{0}_{i-th}, \ldots, s_N)$,

$$\begin{aligned} \kappa(s^{i+}|s^{i-}) &= F_i(s^{i-}) \\ \kappa(s^{i-}|s^{i+}) &= 1 - F_i(s^{i+}). \end{aligned} \tag{5.3}$$

The pairwise maximum-entropy distribution $P_{\mathrm{p}}(\boldsymbol{s}|\,\boldsymbol{\mu},\boldsymbol{\Lambda})$ is stationary under the Markov dynamics above. The proof can be obtained as the $\Lambda_{\mathrm{I}} = 0$ case of the proof, given below, for the inhibited pairwise maximum-entropy model.

5.2.2 Inhibited Glauber dynamics and its stationary maximum-entropy distribution

Inhibited Glauber dynamics.

In the "inhibited" Glauber dynamics, the network of N neurons with states $s_i(t)$ has an additional neuron with state $s_{\mathrm{I}}(t)$. The dynamics is determined by the following algorithm starting at time step t with states $\boldsymbol{s} = \boldsymbol{s}(t)$, $s_{\mathrm{I}} = s_{\mathrm{I}}(t)$:

1. One of the N units is chosen, each unit having probability $1/N$ of being the chosen one. Suppose i is the selected unit.

2. The chosen unit i is updated to the state $s_i' \coloneqq s_i(t+1)$ with probability

$$\begin{aligned} p(s_i'|\, \boldsymbol{s}, s_\mathrm{I}) &= \bigl(1+\exp[(1-2s_i')F_i(\boldsymbol{s}, s_\mathrm{I})]\bigr)^{-1} \\ &= \begin{cases} \bigl[1+\mathrm{e}^{F_i(\boldsymbol{s},s_\mathrm{I})}\bigr]^{-1}, & \text{for } s_i'=0, \\ \bigl[1+\mathrm{e}^{-F_i(\boldsymbol{s},s_\mathrm{I})}\bigr]^{-1}, & \text{for } s_i'=1, \end{cases} \\ \text{with} \quad F_i(\boldsymbol{s}, s_\mathrm{I}) &\coloneqq \mu_i + \sum_k^{k\neq i} \mathit{\Lambda}_{ik} s_k/2 + \mathit{\Lambda}_\mathrm{I} s_\mathrm{I}. \end{aligned}$$

   Note the additional coupling from the neuron s_I, with strength $\mathit{\Lambda}_\mathrm{I}$. This strength can have any sign, but we are interested in the $\mathit{\Lambda}_\mathrm{I} \leqslant 0$ case; we therefore call s_I the “inhibitory neuron”.

3. The inhibitory neuron is deterministically updated to the state $s_\mathrm{I}' \coloneqq s_\mathrm{I}(t+1)$ given by

$$s_\mathrm{I}' = \mathrm{H}\Bigl(\sum_k s_k/N - \theta\Bigr), \tag{5.4}$$

   corresponding to a Kronecker-delta conditional probability

$$p(s_\mathrm{I}'|\, \boldsymbol{s}, s_\mathrm{I}) = p(s_\mathrm{I}'|\, \boldsymbol{s}) = \delta\bigl[s_\mathrm{I}' - \mathrm{H}\bigl(\textstyle\sum_k s_k/N - \theta\bigr)\bigr]. \tag{5.5}$$

   In other words, the inhibitory neuron becomes active if the population-averaged activity of the other neurons is equal to or exceeds the threshold θ.

4. The time is stepped forward, $t+1 \to t$, and the process repeats from step 1.

The original Glauber dynamics, described in the previous section, is recovered when $\mathit{\Lambda}_\mathrm{I} = 0$, which corresponds to decoupling the inhibitory neuron s_I.

The total transition probability can be written as

$$\begin{aligned} p(\boldsymbol{s}', s_\mathrm{I}'|\, \boldsymbol{s}, s_\mathrm{I}) = \frac{1}{N}\, \delta\bigl[s_\mathrm{I}' - \mathrm{H}\bigl(\textstyle\sum_k s_k/N - \theta\bigr)\bigr] \times \\ \sum_i \Bigl[\bigl(1+\exp[(1-2s_i')F_i(\boldsymbol{s}, s_\mathrm{I})]\bigr)^{-1} \prod_k^{k\neq i} \delta(s_k' - s_k)\Bigr]; \end{aligned} \tag{5.6}$$

the product of Kronecker deltas in the last term ensures that at most one of the N neurons changes state at each timestep.

The transition probabilities for the chosen neuron s_i and the inhibitory neuron s_I are independent, conditional on the state of the network at the previous timestep:

$$p(\boldsymbol{s}', s_\mathrm{I}'|\, \boldsymbol{s}, s_\mathrm{I}) = p(\boldsymbol{s}'|\, \boldsymbol{s})\, p(s_\mathrm{I}'|\, \boldsymbol{s}),$$

so the transition probability for the N neurons only can be written as

$$p(\boldsymbol{s}'|\,\boldsymbol{s}) = \frac{1}{N}\sum_i \Big[\big(1+\exp[(1-2s'_i)F_i(\boldsymbol{s})]\big)^{-1}\prod_k^{k\neq i}\delta(s'_k - s_k)\Big], \tag{5.7}$$

$$\text{with}\quad F_i(\boldsymbol{s}) \coloneqq \mu_i + \sum_k^{k\neq i} \Lambda_{ik}s_k/2 + \Lambda_{\mathrm{I}}\,\mathrm{H}\big(\textstyle\sum_k s_k/N - \theta\big)\big]. \tag{5.8}$$

Proof that the inhibited maximum-entropy model is the stationary distribution of the inhibited Glauber dynamics.

The modified maximum-entropy distribution P_{i}, (Eq. 3.22), is the stationary distribution of a slightly modified version of the above dynamics, with the update rule

$$s'_{\mathrm{I}} = \mathrm{H}\Big(\sum_k^{k\neq i} s_k/N - \theta\Big), \tag{5.9}$$

and the use of N inhibitory neurons, one for each of the original N units. This dynamics has a slightly different transition probability, with activation function

$$F_i(\boldsymbol{s}) \coloneqq \mu_i + \sum_k^{k\neq i} \Lambda_{ik}s_k/2 + \Lambda_{\mathrm{I}}\,\mathrm{H}\big(\textstyle\sum_k^{k\neq i} s_k/N - \theta\big)\big] \tag{5.10}$$

instead of (Eq. 5.8). Note that the two dynamics are very similar for large enough N. To prove the stationarity of inhibited maximum-entropy distribution P_{i}, we show that P_{i} satisfies the detailed-balance equality

$$p(\boldsymbol{s}'|\,\boldsymbol{s})\,P_{\mathrm{i}}(\boldsymbol{s}) = p(\boldsymbol{s}|\,\boldsymbol{s}')\,P_{\mathrm{i}}(\boldsymbol{s}') \quad \text{or} \quad \frac{p(\boldsymbol{s}'|\,\boldsymbol{s})}{p(\boldsymbol{s}|\,\boldsymbol{s}')} = \frac{P_{\mathrm{i}}(\boldsymbol{s}')}{P_{\mathrm{i}}(\boldsymbol{s})}, \quad \forall \boldsymbol{s},\boldsymbol{s}', \tag{5.11}$$

which is a sufficient condition for stationarity [98, 194, 346].

First note that if $\boldsymbol{s}'$ and $\boldsymbol{s}$ differ in the state of more than one neuron, the transition probability $p(\boldsymbol{s}'|\,\boldsymbol{s})$ vanishes and the detailed-balance above is trivially satisfied. Also the case $\boldsymbol{s}' = \boldsymbol{s}$ is trivially satisfied. Only the case in which $\boldsymbol{s}'$ and $\boldsymbol{s}$ differ in the state of one unit, say s_i, remains to be proven. Assume then that

$$s'_i = 1, \quad s_i = 0, \quad \forall k \neq i,\ s'_k = s_k; \tag{5.12}$$

by symmetry, if the detailed balance is satisfied in the case above it will also be satisfied with the values 0 and 1 interchanged.

Substituting the transition probability (Eq. 5.7) and (Eq. 5.10) in the left-hand side of the fraction form of the detailed balance (Eq. 5.11), and noting that $F_i(\boldsymbol{s}') = F_i(\boldsymbol{s})$, we

have

$$\begin{aligned}\frac{p(\boldsymbol{s}'|\,\boldsymbol{s})}{p(\boldsymbol{s}|\,\boldsymbol{s}')} &= \exp[-F_i(\boldsymbol{s})]^{-1} \\ &= \exp\Big[\mu_i + \sum_k^{k\neq i} \Lambda_{ik}s_k/2 + \Lambda_{\mathrm{I}}\,\mathrm{H}\big(\textstyle\sum_k^{k\neq i} s_k/N - \theta\big)\Big].\end{aligned} \tag{5.13}$$

Using the expression for the inhibited model P_{I}, (Eq. 3.22), in the right-hand side of the fraction form of the detailed balance (Eq. 5.11), we have

$$\begin{aligned}\frac{P(\boldsymbol{s}')}{P(\boldsymbol{s})} = \exp\Big[\mu_i + \sum_k^{k\neq i}\mu_k s_k + \tfrac{1}{2}\sum_k^{k\neq i}\Lambda_{ik}s_k + \\ \tfrac{1}{2}\sum_{k,m\neq i}^{k<m}\Lambda_{mk}s_m s_k + \Lambda_{\mathrm{I}}N\,G\Big(\sum_k^{k\neq i}\frac{s_k}{N} + \frac{1}{N} - \theta\Big)\Big] \\ = \exp\Big[\mu_i + \tfrac{1}{2}\sum_k^{k\neq i}\Lambda_{ik}s_k + \Lambda_{\mathrm{I}}\,\mathrm{H}\big(\sum_k^{k\neq i} s_k/N - \theta\big)\Big],\end{aligned} \tag{5.14}$$

where we have used the equality $NG(x+1/N) - NG(x) = \mathrm{H}(x)$, valid if $x = \sum_k^{k\neq i} s_k/N - \theta$ and $N\theta \in \mathbf{Z}$. Comparison of formulae (Eq. 5.13) and (Eq. 5.14) finally proves that the detailed balance is satisfied also in the case (Eq. 5.12).

5.2.3 Simulation of Glauber dynamics with NEST

The neuron model `ginzburg_neuron` in NEST implements the Glauber dynamics, if the parameters of the gain function are chosen appropriately. The gain function has the form

$$g_{\mathrm{ginzburg}}(h) = c_1 h + \frac{c_2}{2}(1 + \tanh(c_3(h-\theta)). \tag{5.15}$$

With $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, setting $x = c_3(h-\theta)$, $c_1 = 0$, $c_2 = 1$, $c_3 = \frac{1}{2}$ it takes the form

$$\begin{aligned}g_{\mathrm{ginzburg}}(h) &= \frac{1}{2}\frac{e^x + e^{-x} + e^x - e^{-x}}{e^x + e^{-x}}, \\ &= \frac{1}{1+e^{-2x}} = \frac{1}{1+e^{-(h-\theta)}},\end{aligned} \tag{5.16}$$

which is identical to (Eq. 5.2).

5.2.4 Inferences from a network to a subnetwork and vice versa

This section summarizes some mathematical relations between the probability distributions for the states of a network of binary units and a subnetwork thereof, under an assumption of symmetry. These relations are standard results of probability theory, but seem to be rarely used in neuroscience. Some of their consequences for inferences between network and subnetwork, especially in connection with the maximum-entropy principle, are briefly discussed. The meanings and applicability of the assumption of symmetry are also discussed.

Uncertainties about networks and subnetworks

If we are uncertain about the state of a network of neurons, what is our uncertainty about the state of a subnetwork? And vice versa, if we are uncertain about the state of a subnetwork, what is our uncertainty about the state of the whole network?

If our uncertainties are expressed by two probability distributions for the states of network and subnetwork, and these distributions satisfy a particular symmetry property, then they are related by precise and relatively simple mathematical formulae. These formulae are of essential importance when we want to make inferences about the whole network given data about the subnetwork and vice versa.

Formulae connecting network and subnetwork probabilities

Setup. Consider a network of N binary neurons with states $(X_1, \dots, X_N)$ having fixed but unknown binary values $(R_1, \dots, R_N)$, with R_i in $\{0, 1\}$, vectorially written $\boldsymbol{X} = \boldsymbol{R}$. For example, $\boldsymbol{X}$ can represent the state of a network at a particular time. We call the neurons "units" to lend some generality to our discussion. We shall make statements about the whole network of N units and about a subnetwork of n units; the word "network" will always refer to the *whole* network. The subnetwork states and their values are denoted by lowercase letters: $(x_1, \dots, x_n) \equiv \boldsymbol{x}$ and $(r_1, \dots, r_n) \equiv \boldsymbol{r}$; but note that $x_i \equiv X_{j_i}$ and $r_i \equiv R_{j_i}$ for some distinct $j_1, \dots, j_n$. We shall also make statements about the network-averaged state, or *network average*:

$$\overline{\boldsymbol{X}} \coloneqq (X_1 + \dots + X_N)/N, \tag{5.17}$$

and the subnetwork-averaged state, or *subnetwork average*:

$$\widehat{\boldsymbol{x}} \coloneqq (x_1 + \ldots + x_n)/n. \tag{5.18}$$

The quantities $N\overline{\boldsymbol{X}}$ and $n\widehat{\boldsymbol{x}}$ represent the total number of active units in the network and the subnetwork. Quantities like $\overline{\boldsymbol{R}}$ and $\widehat{\boldsymbol{r}}$ are defined analogously. The averaging operators

$\widehat{\cdot}$ and $\overline{\cdot}$ are also extended to averages of $\binom{n}{m}$ or $\binom{N}{m}$ products of m states; e.g.,

$$\overline{\boldsymbol{X}\boldsymbol{X}} \coloneqq \tbinom{N}{2}^{-1}(X_1X_2 + X_1X_3 + \cdots + X_{N-1}X_N), \tag{5.19}$$

$$\widehat{\boldsymbol{x}\boldsymbol{x}\boldsymbol{x}} \coloneqq \tbinom{n}{3}^{-1}(x_1x_2x_3 + x_1x_2x_4 + \cdots + x_{n-2}x_{n-1}x_n), \tag{5.20}$$

and so on.

Assumptions. Our uncertainty about the network state is represented by the joint probability distribution of the individual states, from which we can derive all other probabilities of interest. We denote it by

$$\mathrm{P}(X_1 = R_1, X_2 = R_2, \dots, X_N = R_N | \, H) \quad \text{or} \quad \mathrm{P}(\boldsymbol{X} = \boldsymbol{R} | \, H). \tag{5.21}$$

Such probability is conditional on our state of knowledge, i.e. the evidence and assumptions backing our probability assignments, denoted by the proposition H.

In the present discussion, H is a state of knowledge that leads to two specific properties in our probability assignments:

$H1$. *Permutation symmetry*, expressed as the invariance of the joint distribution (Eq. 5.21) under arbitrary permutations of the units's labels:

$$\begin{aligned}\mathrm{P}(X_1 = R_1, X_2 = R_2, \dots, X_N = R_N | \, H) = \\ \mathrm{P}(X_1 = R_{\pi(1)}, X_2 = R_{\pi(2)}, \dots, X_N = R_{\pi(N)} | \, H) \\ \text{for any permutation } \pi.\end{aligned} \tag{5.22}$$

This property can reflect two very different states of knowledge: physical homogeneity of the network, or symmetry in our ignorance about the network. This property is called *finite exchangeability* in the Bayesian literature and its basis, consequences, and alternatives to it are discussed in Section 5.2.4.

$H2$. The network average $\overline{\boldsymbol{X}}$ has a particular distribution Q:

$$\mathrm{P}(\overline{\boldsymbol{X}} = A | \, H) = Q(A), \qquad A \in \left\{0, \tfrac{1}{N}, \tfrac{2}{N}, \dots, 1\right\}. \tag{5.23}$$

For the moment we are not concerned about the specific form of Q and about how it was assigned: it could, e.g., arise from maximum-entropy arguments [e.g.: 5, 36, 82, 115, 174–176, 187, 315, 345] used with data on the network.

Formulae. The state of knowledge H has the following six (not independent) main consequences for our probability assignments:

I. Probability for the network state:

$$\mathrm{P}(\boldsymbol{X} = \boldsymbol{R}|\, H) = \binom{N}{N\overline{\boldsymbol{R}}}^{-1} Q(\overline{\boldsymbol{R}}). \tag{5.24}$$

II. Probability for the state $\boldsymbol{x}$ of any subnetwork of n units:

$$\mathrm{P}(\boldsymbol{x} = \boldsymbol{r}|\, H) = \sum_{NA=0}^{N} \binom{N-n}{NA-n\widehat{\boldsymbol{r}}} \binom{N}{NA}^{-1} Q(A). \tag{5.25}$$

Note that the only summands contributing to this sum are those for which $n\widehat{\boldsymbol{r}} \leq NA \leq N$; the others are zero because by definition $\binom{M}{y} = 0$ if $y < 0$. This remark applies to all the sums of this kind in the rest of this Note.

III. Probability for the subnetwork state conditional on a network state:

$$\mathrm{P}(\boldsymbol{x} = \boldsymbol{r}|\, \boldsymbol{X} = \boldsymbol{R}, H) = \binom{N-n}{N\overline{\boldsymbol{R}} - n\widehat{\boldsymbol{r}}}. \tag{5.26}$$

IV. Probability for the subnetwork average $\widehat{\boldsymbol{x}}$:

$$\mathrm{P}(\widehat{\boldsymbol{x}} = a|\, H) = \binom{n}{na} \sum_{NA=0}^{N} \binom{N-n}{NA-na} \binom{N}{NA}^{-1} Q(A), \qquad a \in \left\{0, \tfrac{1}{n}, \tfrac{2}{n}, \dots, 1\right\}. \tag{5.27}$$

V. Probability for the subnetwork average conditional on the network average:

$$\mathrm{P}(\widehat{\boldsymbol{x}} = a|\, \overline{\boldsymbol{X}} = A, H) = \binom{n}{na} \binom{N-n}{NA-na} \binom{N}{NA}^{-1}. \tag{5.28}$$

VI. The product of the states of any m distinct units from a given subnetwork,

$$x_{i_1} x_{i_2} \cdots x_{i_m}, \qquad 1 \leq i_1 < i_2 < \cdots < i_m \leq n$$

has an expectation equal to that of the subnetwork average of such products, is independent of the subnetwork size n:

$$\mathrm{E}(x_{i_1} \cdots x_{i_m}|\, H) = \mathrm{E}(\underbrace{\widehat{\boldsymbol{x} \cdots \boldsymbol{x}}}_{m \text{ factors}}|\, H) = \mathrm{E}(\underbrace{\overline{\boldsymbol{X} \cdots \boldsymbol{X}}}_{m \text{ factors}}|\, H), \tag{5.29a}$$

and has an explicit expression in terms of Q:

$$\begin{aligned} \mathrm{E}(x_{i_1}\cdots x_{i_m}|\,H) &= \binom{N}{m}^{-1}\sum_{NA=0}^{N}\binom{NA}{m}Q(A) \\ &\equiv \sum_{NA=0}^{N}\binom{N-m}{NA-m}\binom{N}{NA}^{-1}Q(A). \end{aligned} \tag{5.29b}$$

A useful relation connects the expectation of a product (Eq. 5.29) and the mth *factorial moment* [272] of the probability distributions for the averages. The mth factorial moment of the subnetwork average $\widehat{\boldsymbol{x}}$ is defined by

$$\mathrm{E}(*)\underbrace{n\widehat{\boldsymbol{x}}\,(n\widehat{\boldsymbol{x}}-1)\cdots(n\widehat{\boldsymbol{x}}-(m-1))}_{m\text{ factors}}|\,H \equiv \mathrm{E}(*)\frac{(n\widehat{\boldsymbol{x}})!}{(n\widehat{\boldsymbol{x}}-m)!}|\,H, \tag{5.30}$$

an analogous definition holding for $\overline{\boldsymbol{X}}$. We have that

$$\mathrm{E}(x_{i_1}\cdots x_{i_m}|\,H) = \tfrac{(n-m)!}{n!}\mathrm{E}(*)\frac{(n\widehat{\boldsymbol{x}})!}{(n\widehat{\boldsymbol{x}}-m)!}|\,H = \tfrac{(N-m)!}{N!}\mathrm{E}(*)\frac{(N\overline{\boldsymbol{X}})!}{(N\overline{\boldsymbol{X}}-m)!}|\,H. \tag{5.31}$$

As a consequence of the above relation, the first three moments of the probability distributions $\mathrm{P}(\widehat{\boldsymbol{x}}=a|\,H)$ and $\mathrm{P}(\overline{\boldsymbol{X}}=A|\,H)$, are related by

$$\mathrm{E}(\widehat{\boldsymbol{x}}|\,H) = \mathrm{E}(\overline{\boldsymbol{X}}|\,H), \tag{5.32a}$$

$$\begin{aligned} \mathrm{E}(\widehat{\boldsymbol{x}}^2|\,H) = \mathrm{E}(\overline{\boldsymbol{X}}|\,H)\frac{N-n}{(N-1)n} + \\ \mathrm{E}(\overline{\boldsymbol{X}}^2|\,H)\frac{N\,(n-1)}{(N-1)n}, \end{aligned} \tag{5.32b}$$

$$\begin{aligned} \mathrm{E}(\widehat{\boldsymbol{x}}^3|\,H) = \mathrm{E}(\overline{\boldsymbol{X}}|\,H)\frac{(N-n)(N-2n)}{(N-1)(N-2)n^2} + \\ \mathrm{E}(\overline{\boldsymbol{X}}^2|\,H)\frac{3N\,(N-n)(n-1)}{(N-1)(N-2)n^2} + \\ \mathrm{E}(\overline{\boldsymbol{X}}^3|\,H)\frac{N^2\,(n-1)(n-2)}{(N-1)(N-2)n^2}. \end{aligned} \tag{5.32c}$$

Relations for higher moments can be obtained recursively from (Eq. 5.31). In general, this means that the two sets of first m moments are related by a homogeneous linear transformation,

$$\mathrm{E}(\widehat{\boldsymbol{x}}^m|\,H) = \sum_{l=1}^{m} M_{ml}(n,N)\mathrm{E}(\overline{\boldsymbol{X}}^l|\,H), \tag{5.33}$$

with a universal, lower-triangular transformation matrix $M_{ml}(n,N)$ that depends only on n, N, and the condition of symmetry (Eq. 5.22).

As intuition suggests, we have

$$\mathrm{E}(\widehat{\boldsymbol{x}}^m|\,H) \xrightarrow{n\to N} \mathrm{E}(\overline{\boldsymbol{X}}^m|\,H), \qquad \mathrm{E}(\widehat{\boldsymbol{x}}^m|\,H) \xrightarrow{n\to 1} \mathrm{E}(\overline{\boldsymbol{X}}|\,H), \tag{5.34}$$

the latter because ${x_i}^m = x_i$, since states are $\{0,1\}$-valued.

The core of the six mathematical relations above are (Eq. 5.25) and (Eq. 5.27). The latter expresses the probability for the subnetwork average as a mixture of hypergeometric distributions [182, ch. 3, 286, § 4.8.3, 83, § II.6], with parameters $N, N\overline{\boldsymbol{X}}, n$, weighted by the probabilities $\mathrm{P}(\overline{\boldsymbol{X}} = A|\,H)$ [cf. 195, § 4, esp. eq. (22)]. The connection between this mixture representation and the condition of symmetry (Eq. 5.22) is well-known in the Bayesian literature [61, 67, 68, 155, 183, 195].

Asymptotic approximations and generalizations. Recall the definition of the Shannon and Burg [41, 81] entropies for a binary probability distribution $(p, 1-p)$ with p in $[0,1]$

$$H(p) \coloneqq -p\ln p - (1-p)\ln(1-p), \qquad H_{\mathrm{B}}(p) \coloneqq \ln p + \ln(1-p). \tag{5.35}$$

The binomial coefficient has an asymptotic form that involves the two entropies above [35, 62, 244, 261, 280, 349, 355]:

$$\ln \binom{M}{Mp} = M\,H(p) - \tfrac{1}{2}\ln(2\pi M) - \tfrac{1}{2}H_{\mathrm{B}}(p) + \mathrm{O}\Big(\tfrac{1}{\sqrt{M}}\Big), \qquad M \text{ large}, \tag{5.36}$$

and this expression can be used to obtain asymptotic forms of the mathematical formulae (Eq. 5.24)–(Eq. 5.31) for large N, depending on the magnitudes of n and n/N. For example, if N and n are large and their ratio $k \coloneqq n/N$ finite, the probability distributions for the averages can be approximated (provided some regularity conditions) by continuum densities,

$$\begin{gathered} n\,\mathrm{P}(\widehat{\boldsymbol{x}} = a|\,H) \approx f(a), \qquad N\,\mathrm{P}(\overline{\boldsymbol{X}} = A|\,H) = N\,Q(A) \approx F(A), \\ f, F\colon [0,1] \to [0,+\infty[, \end{gathered} \tag{5.37}$$

and sums by integrals, and (Eq. 5.25) takes on the approximate form

$$f(a) = \int_0^1 \exp\Big\{N\Big[k\,H(a) - H(A) + (1-k)\,H\Big(\frac{A-ka}{1-k}\Big)\Big]\Big\}\,F(A)\,\mathrm{d}A, \tag{5.38}$$

which shows an interplay of three entropies that depends on the ratio n/N. In some cases this allows us to approximate the integral by the value at a mode determined by a generalized entropy principle. (In fact, the standard maximum-entropy procedure can be derived from the probability calculus via a similar approximation [270].)

The relations above can also be generalized to K-valued states X_i in $\{0, \dots, K-1\}$, leading to the appearance of the generalized hypergeometric distribution [182, ch. 3, 286, § 4.8.3, 83, § II.6], or to real-valued states X_i in $\mathcal{R}$.

We do not pursue any of these approximations or generalizations in this Note.

Examples of inferential use of the formulae

From network to subnetwork. Let us illustrate with an example how the probability distribution for the subnetwork average $\widehat{\boldsymbol{x}}$, determined by (Eq. 5.27), changes with the subnetwork size n. Choose a network-average distribution $\mathrm{P}(\overline{\boldsymbol{X}} = A|\, H)$ belonging to the exponential family [28, § 4.5.3, see also 89]:

$$\mathrm{P}(\overline{\boldsymbol{X}} = A|\, H) = Q(A) \propto \binom{N}{NA} \exp[\lambda_2\, NA\,(NA-1)/2 + \lambda_1\, NA]. \tag{5.39}$$

This is the form obtained from the principle of maximum relative entropy [e.g.: 5, 36, 82, 115, 174–176, 187, 315, 345] with first and second moments as constraints and the reference distribution Q_0 defined by $Q_0(A) = 2^{-N}\binom{N}{NA}$, corresponding to a uniform probability distribution for the network state $\boldsymbol{X}$.

The probability distribution of (Eq. 5.39) is plotted in Figure 5.2, together with the resulting subnetwork-average distributions $\mathrm{P}(\widehat{\boldsymbol{x}} = a|\, H)$, for the case in which $N = 1000$ units, $\lambda_1 = -2.55$, $\lambda_2 = 0.005$, and $n = 10, 50, 100, 250$. The distributions become broader as n decreases, and the minimum of the original distribution disappears; at the same time the finite-difference

$$\frac{\mathrm{P}(\widehat{\boldsymbol{x}} = a + 1/n|\, H) - \mathrm{P}(\widehat{\boldsymbol{x}} = a|\, H)}{1/n}$$

presents a sharp jump at this minimum when $n \approx 100$.

To the eye familiar with maximum-entropy distributions, the subnetwork-average distributions of Figure 5.2 do not look like maximum-entropy ones with second-moment constraints. In fact, they are not and *cannot* be:

$$\mathrm{P}(\widehat{\boldsymbol{x}} = a|\, H) \neq \kappa \binom{n}{na} \exp[\kappa_2\, na\,(na-1)/2 + \kappa_1\, na] \tag{5.40}$$

for any $\kappa, \kappa_1, \kappa_2$, unless $n = 2$. This impossibility holds more generally for any number of constraints m and subnetwork size n such that $m < n$. The reason is simple: suppose we have assigned a maximum-entropy distribution with m moment constraints as the distribution for the network average. If we want the same kind of distribution for a subnetwork of size n, we are free to play with $m+1$ parameters (normalization included), but we must also satisfy the $n+1$ equations corresponding to the marginalization (Eq. 5.27). This is generally impossible unless $m \geq n$. (Impossibilities of a similar kind appear in statistical mechanics, see e.g. ref. [232].)

This fact can be significant for recent works [e.g., 96, 234, 297, 309, 310, 312, 336, 338, 339] in which a maximum-entropy probability distribution with second- or third-moment constraints is assigned to relatively small subnetworks ($n < 200$) of neurons. If we assume that such subnetwork is part of a larger network, and assume the condition of symmetry (Eq. 5.22), then the larger network *cannot* be assigned a maximum-entropy distribution with the same number of constraints. Vice versa, if we assign such a maximum-entropy distribution to the larger network, then none of its subnetwork of enough large size n can be assigned a similar maximum-entropy distribution.

The dependence of the first four moments $\mathrm{E}(\hat{\boldsymbol{x}}^m | H)$ as a function of size n is shown in Figure 5.3. The moments become practically constant when $n \approx 100$ or larger. The expectations of m-tuple products of states $\mathrm{E}(x_{i_1} \cdots x_{i_m} | H)$, proportional to the factorial moments, are not shown as they do not depend on n.

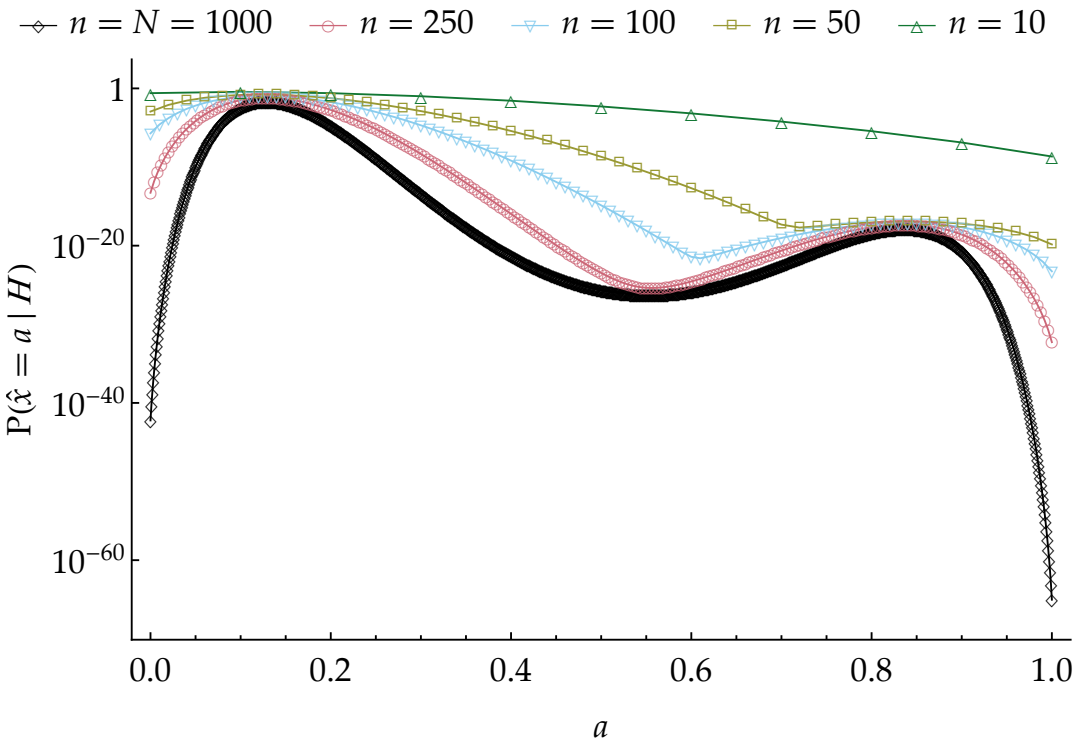

Figure 5.2: Probability distributions $\mathrm{P}(\hat{\boldsymbol{x}} = a | H)$ for different subnetwork sizes n, obtained from a network probability distribution $\mathrm{P}(\overline{\boldsymbol{X}} = A | H)$ having the maximum-etropy form (Eq. 5.39).

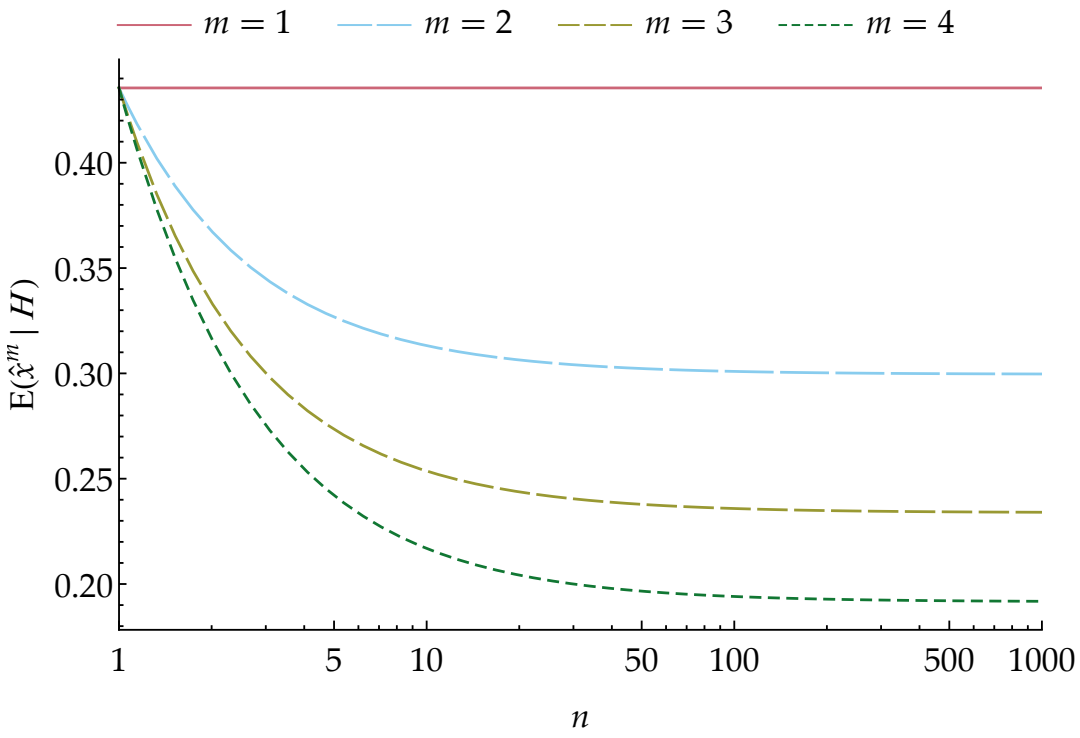

Figure 5.3: Moments of the probability distributions $\mathrm{P}(\widehat{\boldsymbol{x}} = a|\, H)$ as functions of the subnetwork size n.

From subnetwork to network. We have seen that, given the condition of symmetry (Eq. 5.22), the probability $\mathrm{P}(\overline{\boldsymbol{X}} = A|\, H)$ for the network average determines that of each subnetwork average, $\mathrm{P}(\widehat{\boldsymbol{x}}|\, H)$, by the marginalization (Eq. 5.27). The reverse is trivially not true, since (Eq. 5.27), as a linear mapping from $\mathcal{R}^{N+1}$ to $\mathcal{R}^{n+1}$, with N larger than n, is onto but not into. Assigning a probability distribution $\mathrm{P}(\widehat{\boldsymbol{x}} = a|\, H)$ to a subnetwork average $\widehat{\boldsymbol{x}}$ does not determine a network distribution $\mathrm{P}(\overline{\boldsymbol{X}} = A|\, H)$: it only restricts the set of possible ones; this set can in principle be determined via linear-programming methods [146–150].

Analogous situations appear in the truth-valued logical calculus: if the composite proposition $A \Rightarrow B$ is assigned the truth-value “true”, then assigning A the value “true” also determines the value of B, whereas assigning B the value “true” leaves the value of A undetermined.

The same linear-programming methods show that any inference from subnetwork properties to network ones must necessarily start from some assumptions I that assign a probability distribution $\mathrm{P}(\boldsymbol{X} = \boldsymbol{R}|\, I)$ for the network states. The approaches to this task and reformulations of it have become uncountable: they include exchangeable models, parametric and non-parametric models, hierarchical models, general linear models, models via sufficiency, maximum-entropy models, and whatnot [e.g.: 28, 56, 58, 85, 100, 107, 129, 182, 188, 189, 191]. We now show two examples, based on a maximum-entropy approach, that to our knowledge have not yet been explored in the neuroscientific literature.

First example: moment constraints for the network. Consider a state of knowledge H' leading to the following properties:

$H'1$. the expectations of the single and pair averages $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{x}\boldsymbol{x}}$ of a particular subnetwork have given values

$$\mathrm{E}(\widehat{\boldsymbol{x}}|\, H') = c_1, \qquad \mathrm{E}(\widehat{x_i x_j}|\, H') = c_2; \tag{5.41}$$

$H'2$. the network probability distribution $\mathrm{P}(\boldsymbol{X} = \boldsymbol{R}|\, H')$ has maximum relative entropy with respect to the uniform one, given the constraints above.

Then the probability distribution for the network conditional on H' is completely determined: it satisfies the symmetry property (Eq. 5.22) and is defined by

$$\mathrm{p}(\boldsymbol{X} = \boldsymbol{R}|\, H') = K \exp[\Lambda_2 N\overline{\boldsymbol{R}}\,(N\overline{\boldsymbol{R}} - 1)/2 + \Lambda_1 N\overline{\boldsymbol{R}}]$$

with K, Λ_m, such that the distribution is normalized and

$$K \sum_{NA=0}^{N} \binom{N-m}{NA-m} \exp[\Lambda_2 NA\,(NA-1)/2 + \Lambda_1 NA] = c_m, \quad m = 1, 2. \tag{5.42}$$

We omit the full proof of this statement: it is a standard application of the maximum-entropy procedure [e.g.: 36, 82, 115, 174–176, 187, 315, 345], combined with the equality (Eq. 5.29) of subnetwork and network expectations, e.g.

$$c_2 = \mathrm{E}(\widehat{\boldsymbol{x}\boldsymbol{x}}|\, H') = \binom{N}{2}^{-1} \sum_{NA=0}^{N} \binom{NA}{2} \mathrm{P}(\overline{\boldsymbol{X}} = A|\, H'), \tag{5.43}$$

and with relations (Eq. 5.24), (Eq. 5.26). This example is easily generalized to any number m of constraints such that $m \leq n$.

Note again that, as remarked in Section 5.2.4, the subnetwork from which the averages in the expectations (Eq. 5.41) are calculated has a probability distribution $\mathrm{P}(\widehat{\boldsymbol{x}} = a|\, H)$ determined by the marginalization (Eq. 5.27) and does *not* have a maximum-entropy form with the same number of constraints.

Second example: subnetwork-distribution constraint. Consider another state of knowledge H'' leading to the following properties:

$H''1$. the average $\widehat{\boldsymbol{x}}$ of a particular subnetwork has a probability distribution q:

$$\mathrm{P}(\widehat{\boldsymbol{x}} = a|\, H'') = q(a); \tag{5.44}$$

$H''2$. the probability distribution for the network, $\mathrm{P}(\boldsymbol{X} = \boldsymbol{R}|\, H'')$, has maximum relative entropy with respect to the uniform one, given the constraint above.

Then the probability distribution for the network given H'' is completely determined and satisfies the symmetry property (Eq. 5.22):

$$\mathrm{P}(\boldsymbol{X} = \boldsymbol{R}|\, H'') = \exp\left[\sum_{na=0}^{n} \mathit{\Lambda}_a \binom{n}{na}\binom{N-n}{N\overline{\boldsymbol{R}}-na}\right]$$

with $\mathit{\Lambda}_a$ such that

$$\sum_{NA=0}^{N} \binom{n}{na}\binom{N-n}{NA-na} \exp\left[\sum_{na=0}^{n} \mathit{\Lambda}_a \binom{n}{na}\binom{N-n}{NA-na}\right] = q(a) \quad (5.45)$$

(the normalization constraint being unnecessary since q is normalized). This result is just another application of the maximum-entropy procedure with $n+1$ (linear) constraints given by (Eq. 5.25), where the left-hand side is now given and equal to $q(a)$.

This example is equivalent to the generalization of the previous one with n moment constraints, since knowledge of $\mathrm{P}(\widehat{\boldsymbol{x}} = a|\, H'')$ is equivalent to knowledge its first n moments.

The above examples do not mention how the values of the expectation constraints or of the subnetwork-average probability distribution can have been assigned. They cannot be assigned by a measurement, of course, since probabilities and expectations are not physical quantities and cannot be physically measured – they represent guesses of an observer and depend on the observer's state of knowledge and assumptions. Rather, such values usually come from measurements made on “copies” – in a very general sense – of the states of the network; e.g., when we have a time sequence of them and measure the frequencies of their occurrence. Such situations can again be fully analysed with the probability calculus, and one can show [270] that the maximum-entropy formulae in the examples above are just limit forms of such an analysis, employing measured physical data $\mathit{\Delta}$ like, e.g., frequencies, and repeated applications of Bayes's theorem,

$$\mathrm{P}(\boldsymbol{X} = \boldsymbol{R}|\, \mathit{\Delta}, I) \propto \mathrm{P}(\mathit{\Delta}|\, \boldsymbol{X} = \boldsymbol{R}, I)\, \mathrm{P}(\boldsymbol{X} = \boldsymbol{R}|\, I), \quad (5.46)$$

which updates the initial probability assignments on such data. But the discussion of this is again outside the scope of this Note.

On the symmetry property

The symmetry property (Eq. 5.22) is called *finite exchangeability* in the Bayesian literature and, as was mentioned in Section 5.2.4, its relation to the hypergeometric distribution in expression (Eq. 5.27) for the subnetwork average is well-known [61, 67, 68, 155, 183, 195].

This property can reflect two very different states of knowledge: either (a) knowledge that the network is somehow *physically homogeneous*, or (b) *complete ignorance* about the network's homogeneity or inhomogeneity. In the second case we are saying that the indices

or labels "i" of the units are *uninformative* – because, for example, we have no idea of how the units were labelled, hence we cannot presuppose any relation among the units, nor can we presuppose any structural or topological properties of the network they constitute. In the first case we are saying that the labels are *irrelevant*, even though they might be informative. The reason could be that, even if there is a connection between labels and, say, spatial locations of the units, each unit is nevertheless physically, homogeneously, identically connected to all the others; so we assume that spatial location does not play a relevant role.

An important consequence of the symmetry property is that no amount of new evidence about an *a*symmetry in the labelling of some units can lead to asymmetric predictions about the remaining ones. For example, if we have new data Δ saying that the first n units are in state 1 and the last n in state 0,

$$\Delta \coloneqq (X_1 = X_2 = \cdots = X_n = 1, \quad X_{N-(n-1)} = \cdots = X_{N-1} = X_N = 0), \tag{5.47}$$

with n large, our updated probabily distribution will still assign (as can be shown using (Eq. 5.22) and (Eq. 5.24)) the same probability for their respective neighbouring units with labels $n+1$ and $N-n$ to be in state 1 or 0:

$$\begin{aligned} \mathrm{P}(X_{n+1} = 1 |\, \Delta, H) &= \mathrm{P}(X_{N-n} = 1 |\, \Delta, H), \\ \mathrm{P}(X_{n+1} = 0 |\, \Delta, H) &= \mathrm{P}(X_{N-n} = 0 |\, \Delta, H). \end{aligned} \tag{5.48}$$

This may seem unreasonable – we would now say that the unit $n+1$ is more likely to be in state 1, as the first n are, than in 0; and that the unit $N-n$ is more likely in state 0, as the last n are, than in 1; i.e.

$$\begin{aligned} \mathrm{P}(X_{n+1} = 1 |\, \Delta, H) &> \mathrm{P}(X_{n+1} = 0 |\, \Delta, H), \\ \mathrm{P}(X_{N-n} = 1 |\, \Delta, H) &< \mathrm{P}(X_{N-n} = 0 |\, \Delta, H), \end{aligned} \tag{5.49}$$

which is equivalent to $\mathrm{P}(X_{n+1} = 1 |\, \Delta, H) > \frac{1}{2}$, $\mathrm{P}(X_{N-n} = 1 |\, \Delta, H) < \frac{1}{2}$; but the equalities (Eq. 5.48) make this impossible.

The two very different motivations behind the symmetry property – lack of information and lack of relevance – can of course be handled differently by the probability calculus, in such a way that the appearance of asymmetry in new data leads to asymmetry in updated predictions. But this requires a more complex set of assumptions than those embodied in H; it requires, in particular, some sort of probability distribution for the degree of physical symmetry or homogeneity of the network, appropriately quantified.

The moral is that we should use the symmetric assumption H only if we can safely exclude the presence of inhomogeneity or are uninterested in detecting its presence. Otherwise, we must resort to more appropriate (and complex) assumptions. If we repeatedly

observe new values that happen to have a very low probability according to the updated distribution $\mathrm{P}(\boldsymbol{X} = \boldsymbol{R}|\, \Delta_1, \Delta_2, \dots, H)$, this could be an indication that the symmetry property is unreasonable. Any strong departure of higher powers of measured averages from their expected behaviour given in (Eq. 5.32) and illustrated in Figure 5.3 can also be an indication that the symmetry property may have to be abandoned; hence the usefulness of (Eq. 5.32).

Summary and remarks

The main point of this section was to explicitly collect and illustrate the mathematical formulae I–VI, between the probabilities assigned to the state of a network of neurons (or similar entities), and those assigned to the state of a subnetwork thereof. The formulae hold in the simple case of binary states and under an assumption of symmetry. Such relations can be found in several classical texts on probability and statistics; but we deemed it useful to restate them in a neuroscientific context, given their fundamental importance in relating a whole to its parts.

We have indeed seen that these formulae lead to straightforward but in some cases unexpected consequences: e.g., if a network is assigned a maximum-entropy probability distribution, then its subnetworks *cannot* be assigned a maximum-entropy distribution of the same form, and vice versa. The formulae also readily suggest new ways of making inferences or of formulating starting assumptions.

5.2.5 Maximum-entropy model in probability calculus and analogy with statistical mechanics

The formulae of the pairwise maximum-entropy model are similar or even identical to the formulae of the Lenz-Ising or Sherrington-Kirkpatrick spin model [169, 193, 197, 220, 256, 264, 308]. This similarity is useful: it allows us to borrow some mathematical techniques, approximations, and intuitive pictures developed for one model, and to apply them to the other. Yet we purposely emphasize the probability-calculus viewpoint and avoid any "explanation" via statistical-mechanical analogies and their related concepts and jargon. On the whole, such analogies are conceptually limitative and pedagogically detrimental because they put the logical cart before the logical horse: the logical route is not *statistical mechanics* → *maximum-entropy*, but *probability calculus + physics* → *maximum-entropy* → *statistical mechanics* [108, chaps I–IV][17, 80, 123–125, 163, 164, 166, 174, 178, 187, 231] [see also 77, 237–239]. There are in fact important differences between the two models and their quantities. In no particular order:

First: in the case of the Lenz-Ising model, the microscopic state is unknown: we try to guess it from macroscopic properties; this is a problem of inference *within* a model (the energy-constrained maximum-entropy model itself is not brought into question). The opposite holds for the pairwise maximum-entropy model: the "microscopic state" (activity) is known, and we try to find the "macroscopic properties" that lead to a good guess about it; this is a problem of inference *of* a model [cf. 224].

Second: the Lenz-Ising model has one macroscopic quantity as constraint: the total energy, which has one associated Lagrange multiplier: the statistical temperature. The pairwise maximum-entropy model has $N+(N^2-N)/2$ constraints, with as many associated Lagrange multipliers. The difference of constraints between the two models implies essential differences between their entropies; negligence of such differences leads to variants of the Gibbs paradox [118, 119, 121, 184, 218].

Third: The couplings and external fields that appear in the energy of the Lenz-Ising model are measurable physical quantities. The mathematically similar Lagrange multipliers of the pairwise maximum-entropy model are statistical parameters and cannot be measured – they encode our ignorance. In particular, the expression "$\sum_i \mu_i s_i + \sum_{i<j} \Lambda_{ij} s_i s_j$" – what Gibbs [108] calls *index of probability* – is *not* an energy. The following exercise shows why: Assume $\mu_i = -3$, $\Lambda_{ij} = 0.04$, consider a transition from a state $S = 7$ to the state $S = 0$, and calculate by how many metres we could lift a 1 kg weight if that "energy" difference could be converted into mechanical energy.

These differences do not stop us from using mathematical techniques common to the two models to our advantage.

5.3 Appendix to Chapter 4 “Population unitary events (PUE) analysis”

5.3.1 Analytical derivation of $f_A^0(\xi)$ and λ_{cp}^0

Given the mean pairwise correlation coefficient ρ, total number of spikes across all neurons D and the null order ξ_0, the CPP parameters for constructing the null distribution, $f_A^0(\xi)$ and λ_{cp}^0, can be calculated as follows:

In the CPP model mean pairwise correlation coefficient can be expressed knowing the amplitude distribution and the number of neurons [141]:

$$\rho = \frac{\left(\frac{\mathbb{E}[\xi^2]}{\mathbb{E}[\xi]} - 1\right)}{N-1} \tag{5.50}$$

where $\mathbb{E}[\xi] = \sum \xi \cdot f_A(\xi)$ and $\mathbb{E}[\xi^2] = \sum \xi^2 \cdot f_A(\xi)$. For the given correlation order ξ_0, a double peaked amplitude distribution is defined as $f_A^0(\xi) = \eta\delta_{1,\xi} + (1-\eta)\,\delta_{\xi_0,\xi}$. Therefore the expectency of ξ and ξ^2 is:

$$\mathbb{E}[\xi] = \eta + (1-\eta)\cdot\xi_0 \text{ and } \mathbb{E}\left[\xi^2\right] = \eta + (1-\eta)\cdot\xi_0^2. \tag{5.51}$$

Putting equation (Eq. 5.51) into (Eq. 5.50), η is obtained as:

$$\eta = \frac{\xi_0\cdot(\xi_0-1) - \rho\cdot(N-1)\cdot\xi_0}{\xi_0\cdot(\xi_0-1) - \rho\cdot(N-1)\cdot\xi_0 + (N-1)\cdot\rho}, \tag{5.52}$$

therefore, knowing ρ and the null order ξ_0, the amplitude distribution is written as:

$$f_A^0(\xi) = \left(\frac{\xi_0\cdot(\xi_0-1) - \rho\cdot(N-1)\cdot\xi_0}{\xi_0\cdot(\xi_0-1) - \rho\cdot(N-1)\cdot\xi_0 + (N-1)\cdot\rho}\right)\cdot\delta_{1,\xi} \tag{5.53}$$

$$+\left(\frac{(N-1)\cdot\xi_0}{\xi_0\cdot(\xi_0-1) - \rho\cdot(N-1)\cdot\xi_0 + (N-1)\cdot\rho}\right)\cdot\delta_{\xi_0,\xi}. \tag{5.54}$$

For calculating λ_{cp}^0 we make use of relation between carrier process rate and the rate of individual neurons as follows [141]:

$$\lambda_{cp} = \frac{\lambda\cdot N}{\mathbb{E}[\xi]},$$

where λ is the rate of individual neurons and can be written as $\lambda = \frac{D}{N\cdot T}$, hence:

$$\lambda_{cp}^0 = \frac{D}{T\cdot\mathbb{E}[\xi]}. \tag{5.55}$$

Basic Notation

Abbreviations:

AI	asynchronous irregular
BL	Boltzmann learning
CCG	cross-correlogram
CDF	cumulative distribution function
CPP	compound Poisson process
CV	coefficient of variation
FF	Fano factor
FP	false positive
HOC	higher-order correlation
IME	inhibited maximum entropy
ISI	inter-spike interval
ME	maximum entropy
MPST	massively parallel spike trains
UE	unitary event
PUE	population unitary event
RBN	random balanced network
TP	true positive
WOI	window of interest

Frequent mathematical symbols:

$:=$	defined as or denoted by
$\equiv$	identically equal to
$\approx$	approximately equal to
$\hat{}$	estimated values
$\mathbb{H}_0$	null hypothesis
α	significance threshold of a statistical test
ξ	true order of correlatin among spike trains
ξ_c	test order used in the test statistics of the PUE method
ξ_0	null order assumed in the null hypothesis of the PUE method
N	total number of neurons (spike trains) in a data set
f_A	amplitude distribution of the CPP model
λ_{cp}	rate of the carrier process of the CPP model
n_{ξ_c}	population coincidence count
C	population spike count
S	surprise
g_{ij}	raw covariance or coupled activity of the neuron pair ij
c_{ij}	covariance of the neuron pair ij
ρ_{ij}	Pearson correlation coefficient of the neuron pair ij
P_{p}	probability distribution of the pairwise maximum entropy model
P_{r}	probability distribution of the reduced maximum entropy model
P_{i}	probability distribution of the inhibited maximum entropy model
$\boldsymbol{\Lambda}$, $\boldsymbol{\mu}$	lagrange multipliers

Acknowledgments

"Dasein", as Heidegger says, of this work would not have been possible without numerous amazing people who supported me by their knowledge, affection and friendship of all kind (Agápe, Éros, Philia and Storge, as Greeks say). I wish to express my sincere gratefullness to all of these people.

First I wish to give thanks to my advisor, Prof. Dr. Sonja Grün. She has taught me, both consciously and unconsciously, a great deal of important ingredients that I needed as a graduate student. She supported me at the beginning of my graduate studies to catch up with the field of computational Neuroscience and, then, put trust in me to pursue different projects with different collaborators. I learned from her to actively and skillfully question, criticize and evaluate my thinking as well as new information to reach an answer.

After Sonja, I wish to thank my second advisor, Dr. Junji Ito. You taught me how to combine intelligence with humbleness, respect and patience. I was always amazed by your smartness, quick understanding and your accuracy when you were commenting on different scientific topics. I am still training my eyes and perception to be able to interpret new results as concisely and quickly as you.

I would also like to express my appreciation to Prof. Dr. Moritz Helias. You are a great role model of an excellent scientist to me. I learned from you not only the theoretical Neuroscience but also the great beauty in it. I always enjoyed all our meetings, all our brain storming and your way of explaining the complicated matters in science. It is truly a great pleasure to work with you.

Together with Moritz, I wish to thank Dr. PierGianLuca Mana. You are a valuable individual and a great friend before being an astonishing scientist. You helped me a lot, I learned a lot from you and I owe you all of my deep insights on maximum entropy models and Bayesian theory. I always remember the story you told me about the influence of E. T. Jaynes on you to discover the beauty of the probability theory, and I want to tell you that you did the same to me.

Dozens of people have taught me and helped me during my graduate studies on different projects. Among them I wish to express my gratitude to our external collaborators, Dr. Farzad Farkhooi, Dr. Alexa Riehle and Dr. Thomas Brochier which I had a chance to know and work with them.

Thinking about the effect of the external world and the environment on one's productivity, I wish to thank all my colleagues at the INM-6/IAS-6 institute. You people were like a family to me and it is indeed something exceptional how I feel close to all of you. I

am grateful for having so many friends among my colleagues. Philipp: we will continue our search for an optimum place to eat in Südstadt in Cologne (with one unique answer: Veggie burger place!) while discussing how your baby robot can behave more appropriately and can have a better life if you send him to the "temporal coding" school! Lyuba: thanks for all your kindness and the great chats we had always at the institute, in your car and in Cologne. Max: thanks for being available to help me with all the German bureaucracy; one day we have to solve the puzzle of how you can be that fast in saying "shotgun" and fall asleep even faster in the car! Carlos: thanks for all the fun moments and your warmness; all the detours to Düren with Lyuba were worth it! Emiliano: thanks for all the logical and thoughtful advice you always had at hand to help me in making decisions. Serge: thanks for all the random walks that we did while we were talking; I liked the stream of consciousness the most! Hanna: thanks for nice chit-chats and random discussions about everything. Tammo: thanks for all the great musics, photos and fun! Alper: thanks a lot for all your help and the great trip to Japan. Jenia: thanks for all the fun, discussion and nice moments that we experienced! Jakob: thanks for all the great conversations and ideas we had which will be relevant for science in 200 years (sure!); but meanwhile I am pretty sure there will be soon a programming language named "Jakob" much better than "Julia"! Jannis: thanks for all the great discussion we had in the train, in the bus, in the printing room, in the kitchen and in the corridor! I hope Mönchengladbach wins the World Cup! Nicole: thanks a lot for your help and your kindness; I got to know you very late but enjoyed a lot working with you. Marcel: thanks for your friendship and all the great times in Marseille. Janine: meine sehr nette Deutschlehrerin! Danke für schnelle bürokratische Hilfe! David, Johanna, Sacha, Tiziano, Espen, Julia, Sven, Renato, Jyotika, Martina, Michael, Tom, Claudia, Robin, Tobias, Pietro: thanks to all of you for all your help and the great moments we experienced together in the lab focus, in the journal club and in the institute.

Thinking about the importance of the internal world on one's productivity, I wish to express my intimate thanks to Nina: thanks for all the beauty, the happiness, the intimacy and the ViNi world. This world helped me so much to handle all the difficulty in the past years, mille fois merci Makhsh! Thinking about all the complexity deep inside me in defining the philosophy of life, I want to thank Abragh, my other half, my only true Neokarasist mate: Khodeto Eshghe Baaaad va Auch Nicht Zendast, heiter uns bis immer und ewig!

I am grateful to all my friends who were beside me during this time: Sfandiar, Anita, Philipp, Christina, Jenia, Chanzaa, Lucija, Danie, Claudia L, Claudia S, Bahar, Christoph, Silvia and Bizhan. You brought me amazing chats, discussions and fun to pass some difficult moments or just to increase the level of happiness and to help me find the beauty in the resonance!

Finally, I warmly thank my family for all their love and encouragement. My parents who supported me unconditionally along the way by sometimes sacrificing many things for me. My beloved sister, Shima, who was always my best friend and we experienced together happiness, sadness and love. My lovely brother, Mamat, who was always a big support. Thank you all and I am missing all the moments when we were living under the same ceiling.

References

[1] Moshe Abeles. *Corticonics: Neural Circuits of the Cerebral Cortex*. 1st. Cambridge: Cambridge University Press, 1991.

[2] Moshe Abeles. "Role of cortical neuron: integrator or coincidence detector?" In: *Int. J. Math. Sci.* 18 (1982), pp. 83–92.

[3] Milton Abramowitz, Irene A. Stegun, eds. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. tenth printing, with corrections. Vol. 55. National Bureau of Standards Applied mathematics series. First publ. 1964. Washington, D.C.: U.S. Department of Commerce, 1972.

[4] David H. Ackley, Geoffrey E. Hinton, Terrence J. Sejnowski. "A Learning Algorithm for Boltzmann Machines". In: *Cognit. Sci.* 9.1 (1985), pp. 147–169.

[5] Janos Aczél, Z. Daróczy. *On Measures of Information and Their Characterizations*. Vol. 115. Mathematics in science and engineering. New York: Academic Press, 1975.

[6] E. D. Adrian. "The impulses produced by sensory nerve-endings: Part I." In: *The Journal of physiology* 61.1 (Apr. 1926), pp. 49–72. ISSN: 0022-3751. DOI: 16993776 . URL: http://www.ncbi.nlm.nih.gov/pubmed/16993827$%5Cbackslash$nhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1514880.

[7] Stoewer Adrian, Kellner Christian, Benda Jan, Wachtler Thomas, Grewe Jan. "File format and library for neuroscience data and metadata". In: *Frontiers in Neuroinformatics* 8 (2014). ISSN: 1662-5196. DOI: 10.3389/conf.fninf.2014.18.00027. URL: http://www.frontiersin.org/Community/AbstractDetails.aspx?ABS%7B%5C_%7DDOI=10.3389/conf.fninf.2014.18.00027.

[8] Asohan Amarasingham, Ting-Li Chen, Matthew T Harrison, David L Sheinberg. "Spike Count Reliability and the Poisson Hypothesis". In: 26.4 (Jan. 2006), pp. 801–809.

[9] Shun-ichi Amari, Shiro Ikeda, Hidetoshi Shimokawa. "Information Geometry of α-Projection in Mean Field Approximation". In: [257]. http://www.ism.ac.jp/~shiro/papers/techreport/TR2000-6.pdf. 2001. Chap. ch. 16.

[10] Daniel J. Amit, Nicolas Brunel. "Dynamics of a recurrent network of spiking neurons before and following learning". In: *Network: Comput. Neural Syst.* 8 (1997), pp. 373–404.

[11] Christophe Andrieu, Nando de Freitas, Arnaud Doucet, Michael I. Jordan. "An Introduction to MCMC for Machine Learning". In: *Mach. Learn.* 50.1–2 (2003). http://cis.temple.edu/~latecki/Courses/RobotFall07/PapersFall07/andrieu03introduction.pdf, pp. 5–43.

[12] A Arieli, A Sterkin, A Grinvald, A Aertsen. "Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses". In: 273.5283 (Sept. 1996), pp. 1868–1871.

[13] Rosa Badia, Andrew Davison, Michael Denker, Andre Giesler, Satra Gosh, Carole Goble, Jan Grewe, Sonja Grün, Nicho Hatsopoulos, Yann LeFranc, Jeff Muller, Robert Pröpper, Jeff Teeters, Thomas Wachtler, Michael Weeks, Lyuba Zehl. *INCF Program on Standards for data sharing: new perspectives on workflows and data management for the analysis of electrophysiological data.* Techn. Report. 2015. URL: https://www.incf.org/about-us/history/incf-scientific-workshops.

[14] John C. Baez, Tobias Fritz. "A Bayesian characterization of relative entropy". In: *Theory and Applications of Categories* 29.16 (2014). arXiv:1402.3067, pp. 421–456.

[15] Wyeth Bair, Ehud Zohary, William T. Newsome. "Correlated firing in macaque visual area MT: Time scales and relationship to behavior". In: *J. Neurosci.* 21.5 (2001). http://invibe.net/biblio_database_dyva/woda/data/att/fc5e.file.pdf, pp. 1676–1697.

[16] Konstantin I. Bakhurin, Victor Mac, Peyman Golshani, Sotiris C. Masmanidis. "Temporal correlations among functionally specialized striatal neural ensembles in reward-conditioned mice". In: *J. Neurophysiol.* 115.3 (2016), pp. 1521–1532.

[17] Roger Balian. *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics. Vol. I.* 2nd printing. Theoretical and mathematical physics. Transl. by D. ter Haar and J. F. Gregg; first publ. in French 1982. Berlin: Springer, 2007.

[18] Arunava Banerjee, Peggy Series, Alexandre Pouget. "Dynamical Constraints on Using Precise Spike Timing to Compute in Recurrent Cortical Networks". In: *Neural Computation* 20.4 (2008), pp. 974–993.

[19] H. B. Barlow. "Single units and sensation: a neuron doctrine for perceptual psychology?" In: 1 (1972), pp. 371–394.

[20] Ole E. Barndorff-Nielsen, A. Philip Dawid, Persi Diaconis, Søren Johansen, Steffen L. Lauritzen. "Discussion of Steffen Lauritzen's paper ['Extreme Point Models in Statistics']". In: *Scand. J. Statist.* 11.2 (1984). See [217], pp. 83–91.

[21] Andrea K. Barreiro, Julijana Gjorgjieva, Fred Rieke, Eric Shea-Brown. *When Do Microcircuits Produce Beyond-Pairwise Correlations?* arXiv:1011.2797. 2010.

[22] M. S. Bartlett. “The statistical significance of odd bits of information”. In: *Biometrika* 39.3–4 (1952), pp. 228–237.

[23] Thomas Bayes. “An Essay towards Solving a Problem in the Doctrine of Chances”. In: *Phil. Trans. R. Soc. Lond.* 53 (1763). http://www.stat.ucla.edu/history/essay.pdf; with an introduction by Richard Price. The *Scholium* of p. 392 is reprinted and analysed in [330], pp. 370–418.

[24] Antal Berényi, Zoltán Somogyvári, Anett J. Nagy, Lisa Roux, John D. Long, Shigeyoshi Fujisawa, Eran Stark, Anthony Leonardo, Timothy D. Harris, György Buzsáki. “Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals”. In: *J. Neurophysiol.* 111.5 (2014). http://www.buzsakilab.com/content/PDFs/Berenyi2013.pdf, pp. 1132–1149.

[25] Antal Berènyi, Zoltàn Somogyvàri, Anett J. Nagyand Lisa Roux, John D. Long, Shigeyoshi Fujisawa, Eran Stark, Anthony Leonardo, Timothy D. Harris, György Buzsàki. “Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals”. en. In: *Journal of Neurophysiology* 111.5 (Mar. 2014), pp. 1132–1149. DOI: 10.1152/jn.00785.2013. URL: http://jn.physiology.org/content/111/5/1132.long.

[26] James O. Berger. *Statistical Decision Theory and Bayesian Analysis.* 2nd ed. Springer series in statistics. First publ. 1980. New York: Springer, 1985.

[27] José-Miguel Bernardo, M. H. DeGroot, D. V. Lindley, A. F. M. Smith, eds. *Bayesian Statistics 2: Proceedings of the Second Valencia International Meeting.* Amsterdam and Valencia: Elsevier and Valencia University Press, 1985.

[28] José-Miguel Bernardo, Adrian F. Smith. *Bayesian Theory.* Wiley series in probability and mathematical statistics. Chichester: Wiley, 1994.

[29] Elie Bienenstock. “A model of neocortex”. In: 6 (1995), pp. 179–224.

[30] Kurt Binder. “Applications of Monte Carlo methods to statistical physics”. In: *Rep. Prog. Phys.* 60.5 (1997). http://fisica.ciencias.uchile.cl/~gonzalo/cursos/SimulacionII/rpphys_binder97.pdf, pp. 487–559.

[31] Kurt Binder, ed. *Applications of the Monte Carlo Method in Statistical Physics.* 2nd ed. Vol. 36. Topics in current physics. First publ. 1984. Berlin: Springer, 1987.

[32] J. J. Binney, N. J. Dowrick, A. J. Fischer, M. E. J. Newman. *The Theory of Critical Phenomena: An Introduction to the Renormalization Group.* First publ. 1992. Oxford: Oxford University Press, 2001.

[33] S. M. Bohte, H. Spekreijse, P. R. Roelfsema. “The Effects of Pair-wise and Higher-order Correlations on the Firing Rate of a Postsynaptic Neuron”. In: *Neural Comp.* 12.1 (2000), pp. 153–179.

[34] Alexander Borst, Frédéric E. Theunissen. “Information theory and neural coding”. In: *Nat. Neurosci.* 2.11 (1999), pp. 947–957.

[35] Wayne A. Bowers. “More about approximations to $n!$”. In: *Am. J. Phys.* 51.9 (1983). See also [62, 244, 261, 349, 355], p. 778.

[36] G. Larry Bretthorst. “The maximum entropy method of moments and Bayesian probability theory”. In: *Am. Inst. Phys. Conf. Proc.* 1553 (2013). `http://bayes.wustl.edu/glb/BretthorstHistograms.pdf`, pp. 3–15.

[37] Tamara Broderick, Miroslav Dudik, Gašper Tkačik, Robert E. Schapire, William Bialek. *Faster solutions of the inverse pairwise Ising problem.* `arXiv:0712.2437`. 2007.

[38] Nicolas Brunel. “Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons”. In: 8.3 (2000), pp. 183–208.

[39] Michael A. Buice, Jack D. Cowan, Carson C. Chow. “Systematic fluctuation expansion for neural network activity equations.” In: *Neural Comp.* 22.2 (2010). `arXiv:0902.3925`, pp. 377–426.

[40] Mario Bunge, ed. *Delaware Seminar in the Foundations of Physics.* Vol. 1. Studies in the foundations: methodology and philosophy of science. Berlin: Springer, 1967.

[41] John Parker Burg. “Maximum Entropy Spectral Analysis”. `http://sepwww.stanford.edu/data/media/public/oldreports/sep06/`. PhD thesis. Stanford: Stanford University, 1975.

[42] György Buzsaki. “Large-scale recording of neuronal ensembles”. In: 7.5 (May 2004), pp. 446–451.

[43] L. L. Campbell. “The relation between information theory and the differential geometry approach to statistics”. In: *Inf. Sci.* 35.3 (1985), pp. 199–210.

[44] M. Castelo-Branco, R. Goebel, S. Neuenschwander, W. Singer. “Neural synchrony correlates with surface segregation rules.” In: 8.405 (2000), pp. 685–689.

[45] Ariel Caticha. “Maximum entropy, fluctuations and priors”. In: *Am. Inst. Phys. Conf. Proc.* 568 (2001). `arXiv:math-ph/0008017`, pp. 94–105.

[46] Ariel Caticha, Roland Preuss. “Maximum entropy and Bayesian data analysis: Entropic prior distributions”. In: *Phys. Rev. E* 70.4 (2004). `arXiv:physics/0307055`, p. 046127.

[47] Carlton M. Caves. *Learning and the de Finetti representation.* `http://info.phys.unm.edu/~caves/reports/reports.html`. 2000.

[48] Nikolaj Nikolaević Ćencov [Chentsov]. *Statistical Decision Rules and Optimal Inference*. Vol. 53. Translations of mathematical monographs. Transl. by Lev J. Leifman; first publ. in Russian 1972. Providence, USA: American Mathematical Society, 1982.

[49] Marlene R. Cohen, Adam Kohn. "Measuring and interpreting neuronal correlations". In: *Nat. Neurosci.* 14.7 (2011). http://marlenecohen.com/pubs/CohenKohn2011.pdf, pp. 811–819.

[50] Joachim Confais, Bjørg Elisabeth Kilavik, Adrián Ponce-Alvarez, Alexa Riehle. "On the anticipatory precue activity in motor cortex." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32.44 (Oct. 2012), pp. 15359–68. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.1768-12.2012. URL: http://www.ncbi.nlm.nih.gov/pubmed/23115174.

[51] Imre Csiszár. "An extended maximum entropy principle and a Bayesian justification". In: [27]. 1985, pp. 83–98.

[52] Imre Csiszár. "Sanov Property, Generalized I-Projection and a Conditional Limit Theorem". In: *Ann. Prob.* 12.3 (1984), pp. 768–793.

[53] Imre Csiszár, Paul C. Shields. "Information Theory and Statistics: A Tutorial". In: *Foundations and Trends in Communications and Information Theory* 1.4 (2004). http://www.renyi.hu/~csiszar/, pp. 417–528.

[54] David Dahmen, Hannah Bos, Moritz Helias. "Correlated Fluctuations in Strongly Coupled Binary Networks Beyond Equilibrium". In: *Physical Review X* 6.3 (Aug. 2016), p. 031024. ISSN: 2160-3308. DOI: 10.1103/PhysRevX.6.031024. URL: http://link.aps.org/doi/10.1103/PhysRevX.6.031024.

[55] D. J. Daley, D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer Series in Statistics. New York: Springer-Verlag, 1988. ISBN: 3-540-96666-8, 0-387-96666-8.

[56] Paul Damien, Petros Dellaportas, Nicholas G. Polson, David A. Stephens, eds. *Bayesian Theory and Applications*. Oxford: Oxford University Press, 2013.

[57] A. Philip Dawid. "Conditional independence in statistical theory". In: *J. Roy. Stat. Soc. B* 41.1 (1979). With discussion by D. V. Lindley, A. D. McLaren, J. B. Kadane, J. M. Dickey, S. M. Rizvi, E. F. Harding, G. A. Barnard, P. J. Bickel, M. H. DeGroot, D. A. S. Fraser, R. F. Galbraith, S. Geisser, V. P. Godambe, D. V. Hinkley, H. Kudō, K. V. Mardia, M. Mouchart, I. Nimmo-Smith, M. R. Novick, Donald B. Rubin, R. A. Wijsman, and the author; http://distancecovariance.googlecode.com/svn/trunk/References/Conditional%20Independence%20in%20Statistical%20Theory.pdf, pp. 1–31.

[58] A. Philip Dawid. “Exchangeability and its Ramifications”. In: [56]. 2013. Chap. ch. 2, pp. 19–29.

[59] A. Philip Dawid. “Intersubjective statistical models”. In: [199]. 1982, pp. 287–232.

[60] Jaime De la Rocha, Brent Doiron, Eric Shea-Brown, Josic Kresimir, Alex Reyes. “Correlation between neural spike trains increases with firing rate”. In: 448.16 (Aug. 2007), pp. 802–807.

[61] Bruno de Finetti. “Sulla proseguibilità di processi aleatori scambiabili”. In: *Rend. Ist. Mat. Univ. Trieste* 1 (1969), pp. 53–67.

[62] Joseph B. Dence. “Reply to a letter by Weissman on Stirling’s approximation”. In: *Am. J. Phys.* 51.9 (1983). See also [35, 244, 261, 349, 355], pp. 776–778.

[63] Michael Denker, Sonja Grün. “Brain Inspired Computing”. In: ed. by Katrin Amunts, Lucio Grandinetti, Thomas Lippert, Nicolai Petkov. Springer Series Lecture Notes in Computer Science, in press. Chap. Designing Workflows for the Reproducible Analysis of Electrophysiological Data.

[64] Michael Denker, Alexa Riehle, Markus Diesmann, Sonja Grün. “Estimating the contribution of assembly activity to cortical dynamics from spike and population measures”. In: 29.3 (2010), pp. 599–613. DOI: 10.1007/s10827-010-0241-8.

[65] Michael Denker, Sébastien Roux, Henrik Lindén, Markus Diesmann, Alexa Riehle, Sonja Grün. “The Local Field Potential Reflects Surplus Spike Synchrony”. In: *Cerebral Cortex* 21 (Dec. 2011), pp. 2681–2695. DOI: 10.1093/cercor/bhr040.

[66] Michael R. DeWeese, Anthony M. Zador. “Non-Gaussian Membrane Potential Dynamics Imply Sparse, Synchronous Activity in Auditory Cortex”. In: 26.47 (2006), pp. 12206–12218.

[67] Persi Diaconis. “Finite forms of de Finetti’s theorem on exchangeability”. In: *Synthese* 36.2 (1977). http://statweb.stanford.edu/~sabatti/Stat370/synthese.pdf, pp. 271–281.

[68] Persi Diaconis, David Freedman. “Finite Exchangeable Sequences”. In: *Ann. Prob.* 8.4 (1980), pp. 745–764.

[69] Markus Diesmann. “personal communication”. In: (2016).

[70] Markus Diesmann, Marc-Oliver Gewaltig, Ad Aertsen. “Stable propagation of synchronous spiking in cortical neural networks”. In: 402.6761 (1999), pp. 529–533.

[71] C. Domb, M. S. Green, eds. *Phase Transitions and Critical Phenomena. Vol. 2.* London: Academic Press, 1972.

[72] Kenji Doya, Shin Ishii, Alexandre Pouget, Rajesh P. N. Rao. *Bayesian Brain: Probabilistic Approaches to Neural Coding.* Computational neuroscience. Cambridge, USA: The MIT Press, 2007.

[73] Alexander S. Ecker, Philipp Berens, Georgios A. Keliris, Matthias Bethge, Nikos K. Logothetis. “Decorrelated Neuronal Firing in Cortical Microcircuits”. In: 327.5965 (Jan. 2010), pp. 584–587.

[74] Jos J. Eggermont. *The Correlative Brain*. Vol. 16. Studies of Brain Function. Berlin: Springer-Verlag-Verlag, 1990. ISBN: 354052326X.

[75] Werner Ehm, Benjamin Staude, Stefan Rotter. “Decomposition of neuronal assembly activity via empirical de-Poissonization”. In: *Electron. J. Statist.* 1 (2007), pp. 473–495. ISSN: 1935-7524. DOI: 10.1214/07-EJS095.

[76] Albert Einstein. “Kinetic theory of thermal equilibrium and of the second law of thermodynamics”. In: [79]. 1989, pp. 30–47.

[77] Albert Einstein. “Kinetische Theorie des Wärmegleichgewichtes und des zweiten Hauptsatzes der Thermodynamik”. In: *Ann. der Phys.* 9 (1902). Repr. in [78], transl. in [76], pp. 417–433.

[78] Albert Einstein. “Kinetische Theorie des Wärmegleichgewichtes und des zweiten Hauptsatzes der Thermodynamik”. In: *Ann. der Phys.* 14.S1 (2005), pp. 117–134.

[79] Albert Einstein. *The Collected Papers of Albert Einstein. Vol. 2: The Swiss Years: Writings, 1900–1909. (English translation)*. Transl. by Anna Beck and Peter Havas; http://einsteinpapers.press.princeton.edu/. Princeton: Princeton University Press, 1989.

[80] Gérard G. Emch, Chuang Liu. *The Logic of Thermostatistical Physics*. Berlin: Springer, 2002.

[81] Gary J. Erickson, C. Ray Smith, eds. *Maximum-Entropy and Bayesian Methods in Science and Engineering. Vol. 2: Applications*. Fundamental theories of physics. Dordrecht: Kluwer, 1988.

[82] Shu-Cherng Fang, J. R. Rajasekera, H.-S. J. Tsao. *Entropy optimization and mathematical programming*. reprint. Vol. 8. International series in operations research & management science. New York: Springer, 1997.

[83] William Feller. *An Introduction to Probability Theory and Its Applications. Vol. I*. 3rd ed. Wiley series in probability and mathematical statistics. First publ. 1950. New York: Wiley, 1968.

[84] R. Fernández, C.-E. Pfister. “Global specifications and nonquasilocality of projections of Gibbs measures”. In: *Ann. Prob.* 25.3 (1997), pp. 1284–1315.

[85] Marco A.R. Ferreira, Herbert K.H. Lee. *Multiscale Modeling: A Bayesian Perspective*. Springer series in statistics. New York: Springer, 2007.

[86] David Ferster, Nelson Spruston. “Cracking the Neuronal Code”. In: *Science* 270.5237 (1995), pp. 756–757.

[87] K. H. Fischer, J. A. Hertz. *Spin glasses*. reprint. Vol. 1. Cambridge studies in magnetism. First publ. 1991. Cambridge: Cambridge University Press, 1993.

[88] K. W. Ford, ed. *Statistical Physics*. Vol. 3. 1962 Brandeis Summer Institute lectures in theoretical physics. New York: W. A. Benjamin, 1963.

[89] Sandra Fortini, Lucia Ladelli, Eugenio Regazzini. "Exchangeability, Predictive Distributions and Parametric Models". In: *Sankhyā A* 62.1 (2000), pp. 86–109.

[90] Paul F. Fougère, ed. *Maximum Entropy and Bayesian Methods: Dartmouth, U.S.A., 1989*. Vol. 39. Fundamental theories of physics. Dordrecht: Kluwer, 1990.

[91] David Fowler. "The Binomial Coefficient Function". In: *Am. Math. Monthly* 103.1 (1996), pp. 1–17.

[92] David A. Freedman. "Invariants Under Mixing which Generalize de Finetti's Theorem". In: *Ann. Math. Stat.* 33.3 (1962), pp. 916–923.

[93] B. Roy Frieden. "Statistical Models for the Image Restoration Problem". In: *Comput. Graph. Image Process.* 12 (1980), pp. 40–59.

[94] Hiroshi Fujii, Hiroyuki Ito, Kazuyuki Aihara, Natsuhiro Ichinose, Minoru Tsukada. "Dynamical Cell Assembly Hypothesis – Theoretical Possibility of Spatio-temporal Coding in the Cortex". In: 9.8 (1996). special issue, pp. 1303–1350.

[95] Shigeyoshi Fujisawa, Asohan Amarasingham, Matthew T. Harrison, György Buzsáki. "Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex". In: 11.7 (2008), pp. 823–833.

[96] Elad Ganmor, Ronen Segev, Elad Schneidman. "Sparse low-order interaction network underlies a highly correlated and learnable neural population code". In: *Proc. Natl. Acad. Sci. (USA)* 108.23 (2011). `http://www.weizmann.ac.il/neurobiology/labs/schneidman/The_Schneidman_Lab/Publications.html`, pp. 9679–9684.

[97] Samuel Garcia, Domenico Guarino, Florent Jaillet, Todd Jennings, Robert Pröpper, Philipp Rautenberg, Chris Rodgers, Andrey Sobolev, Thomas Wachtler, Pierre Yger, Andrew Davison. "Neo: an object model for handling electrophysiology data in multiple formats". In: *Frontiers in Neuroinformatics* 8 (2014), p. 10.

[98] Crispin W. Gardiner. *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*. 3rd ed. Springer series in synergetics. First publ. 1983. Berlin: Springer, 2004.

[99] Anthony J. M. Garrett. "Maximum entropy with nonlinear constraints: physical examples". In: [90]. 1990, pp. 243–249.

[100] Andrew Gelman, John B. Carlin, Hal S. Stern, Donald B. Rubin. *Bayesian Data Analysis*. 2nd ed. Texts in statistical science. First publ. 1995. Boca Raton, USA: Chapman & Hall/CRC, 2004.

[101] Hans-Otto Georgii. “Probabilistic Aspects of Entropy”. In: [130]. http://www.mathematik.uni-muenchen.de/~georgii/papers/Dresden.pdf. 2003. Chap. ch. 3, pp. 37–54.

[102] G. L. Gerstein. “Searching for significance in spatio-temporal firing patterns”. In: *Acta Neurobiol. Exp.* 64 (2004), pp. 203–207.

[103] G. L. Gerstein, P. Bedenbaugh, A. M. H. J. Aertsen. “Neuronal Assemblies”. In: 36 (1989), pp. 4–14.

[104] George L. Gerstein, Donald H. Perkel. “Simultaneously recorded trains of action potentials: analysis and functional interpretation.” In: 881.164 (May 1969), pp. 828–830.

[105] Sebastian Gerwinn, Jakob H. Macke, Matthias Bethge. “Bayesian inference for generalized linear models for spiking neurons”. In: *Front. Comput. Neurosci.* 4 (2010), p. 12.

[106] Marc-Oliver Gewaltig, Markus Diesmann. “NEST (NEural Simulation Tool)”. In: *Scholarpedia* 2.4 (2007), p. 1430.

[107] Malay Ghosh, Glen Meeden. *Bayesian Methods for Finite Population Sampling*. reprint. Vol. 79. Monographs on statistics and applied probability. Dordrecht: Springer, 1997.

[108] Josiah Willard Gibbs. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*. http://gallica.bnf.fr/notice?N=FRBNF37257076; repr. in [109]. New York: Charles Scribner’s Sons, 1902.

[109] Josiah Willard Gibbs. *The Collected Works of J. Willard Gibbs. Vol. II: Statistical Mechanics and Dynamics*. First publ. 1906. New York: Longmans, Green and Co., 1928.

[110] Iris Ginzburg, Haim Sompolinsky. “Theory of correlations in stochastic neural networks”. In: *Phys. Rev. E* 50.4 (1994), pp. 3171–3191.

[111] Roy J. Glauber. “Time-Dependent Statistics of the Ising Model”. In: *J. Math. Phys.* 4.2 (1963), pp. 294–307.

[112] Irving John Good. “Corrections to “The Surprise Index for the Multivariate Normal Distribution””. In: *Ann. Math. Stat.* 28.4 (1957). See [117], p. 1055.

[113] Irving John Good. *Good Thinking: The Foundations of Probability and Its Applications*. Minneapolis, USA: University of Minnesota Press, 1983.

[114] Irving John Good. “How to Estimate Probabilities”. In: *J. Inst. Maths. Applics* 2.4 (1966), pp. 364–383.

[115] Irving John Good. “Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables”. In: *Ann. Math. Stat.* 34.3 (1963), pp. 911–934.

[116] Irving John Good. “The Appropriate Mathematical Tools for Describing and Measuring Uncertainty”. In: [113]. First publ. 1957. 1957. Chap. ch. 16, pp. 173–177.

[117] Irving John Good. “The Surprise Index for the Multivariate Normal Distribution”. In: *Ann. Math. Stat.* 27.4 (1956). See also corrections [112], pp. 1130–1135.

[118] Harold Grad. “Levels of description in statistical mechanics and thermodynamics”. In: [40]. 1967. Chap. ch. 5, pp. 49–76.

[119] Harold Grad. “Statistical Mechanics of Dynamical Systems with Integrals Other than Energy”. In: *J. Phys. Chem.* 56.9 (1952). See also [120], pp. 1039–1048.

[120] Harold Grad. “Statistical mechanics, thermodynamics, and fluid dynamics of systems with an arbitrary number of integrals”. In: *Comm. Pure Appl. Math.* 5.4 (1952). See also [119], pp. 455–494.

[121] Harold Grad. “The many faces of entropy”. In: *Comm. Pure Appl. Math.* 14 (1961), pp. 323–354.

[122] Franck Grammont, Alexa Riehle. “Precise spike synchronization in monkey motor cortex involved In preparation for movement”. In: 128 (1999), pp. 118–122.

[123] Walter T. Grandy Jr. *Entropy and the Time Evolution of Macroscopic Systems*. Vol. 141. International series of monographs on physics. Oxford: Oxford University Press, 2008.

[124] Walter T. Grandy Jr. *Foundations of Statistical Mechanics. Vol. I: Equilibrium Theory*. Fundamental theories of physics. Dordrecht: D. Reidel, 1987.

[125] Walter T. Grandy Jr. *Foundations of Statistical Mechanics. Vol. II: Nonequilibrium Phenomena*. Fundamental theories of physics. Dordrecht: D. Reidel, 1988.

[126] Einat Granot-Atedgi, Gašper Tkačik, Ronen Segev, Elad Schneidman. “Stimulus-dependent Maximum Entropy Models of Neural Population Codes”. In: *PLoS Computational Biology* 9.3 (2013). `arXiv:1205.6438`, e1002922.

[127] C. M. Gray. “The Temporal Correlation Hypothesis of Visual Feature Integration: Still Alive and Well”. In: *Neuron* 24 (1999), pp. 31–47.

[128] Charles M. Gray, Wolf Singer. “Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex”. In: 86 (1989), pp. 1698–1702.

[129] Phil Gregory. *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with* Mathematica *Support*. Cambridge: Cambridge University Press, 2005.

[130] Andreas Greven, Gerhard Keller, Gerald Warnecke, eds. *Entropy*. Princeton series in applied mathematics. Princeton and Oxford: Princeton University Press, 2003.

[131] Jan Grewe, Thomas Wachtler, Jan Benda. "A Bottom-up Approach to Data Annotation in Neurophysiology". In: *Frontiers in Neuroinformatics* 5 (2011). ISSN: 1662-5196. DOI: 10.3389/fninf.2011.00016. URL: http://journal.frontiersin.org/article/10.3389/fninf.2011.00016/abstract.

[132] B. Grigelionis. "On the Convergence of Sums of Random Step Processes to a Poisson Process". In: *Theory of Probability & Its Applications* 8.2 (Jan. 1963), pp. 177–182. ISSN: 0040-585X. DOI: 10.1137/1108017. URL: http://epubs.siam.org/doi/abs/10.1137/1108017.

[133] P. J. Grosbøl, R. H. Warmels, eds. *Third ESO/ST-EFC Data Analysis Workshop, held in Garching, 22–23 April, 1991*. ESO conference and workshop proceedings 38. Garching: European Southern Observatory, 1991.

[134] S. Grün. *Unitary Joint-Events in Multiple-Neuron Spiking Activity: Detection, Significance, and Interpretation*. Reihe Physik, Band 60. ISBN 3-8171-1506-7. Thun, Frankfurt/Main: Verlag Harri Deutsch, 1996.

[135] S. Grün, M. Diesmann, A. Aertsen. "'Unitary Events' in Multiple Single-Neuron Spiking Activity. I. Detection and Significance". In: 14.1 (2002), pp. 43–80.

[136] S. Grün, M. Diesmann, A. Aertsen. "'Unitary Events' in Multiple Single-Neuron Spiking Activity. II. Non-Stationary Data". In: 14.1 (2002), pp. 81–119.

[137] S. Grün, M. Diesmann, F. Grammont, A. Riehle, A. Aertsen. "Detecting unitary events without discretization of time". In: *JNeurosciMethods* 94.1 (1999), pp. 67–79.

[138] S. Grün, M. Diesmann, A. Riehle, A. Aertsen. "Unitary events without discretization of time". In: *From Molecular Neurobiology to Clinical Neuroscience*. Ed. by N. Elsner, U. Eysel. Stuttgart: Thieme Verlag, 1999, p. 96.

[139] S. Grün, A. Riehle, M. Diesmann. "Effect of cross-trial nonstationarity on joint-spike events". In: *BiolCybern* 88.5 (2003), pp. 335–351.

[140] Sonja Grün. "Data-driven significance estimation of precise spike correlation." In: *JNeurophysiol* 101 (2009). (invited review), pp. 1126–1140.

[141] Sonja Grün, Stefan Rotter, eds. *Analysis of Parallel Spike Trains*. Springer, 2010.

[142] Stephen F. Gull, G. J. Daniell. "Image reconstruction from incomplete and noisy data". In: *Nature* 272.5655 (1978), pp. 686–690.

[143] Stephen F. Gull, John Skilling. "Maximum entropy method in image processing". In: *IEE Proc. F* 131.6 (1984), pp. 646–659.

[144] Stephen F. Gull, John Skilling. *Quantified Maximum Entropy: `MemSys5`. Users' Manual*. Tech. rep. `http://www.mpe.mpg.de/~aws/integral/issw/memsys5.pdf`; first publ. 1990. Suffolk, UK: Maximum Entropy Data Consultants, 1999.

[145] Stephen F. Gull, John Skilling. "The maximum entropy method". In: [281]. 1984, pp. 267–279.

[146] Theodore Hailperin. "Best Possible Inequalities for the Probability of a Logical Function of Events". In: *Am. Math. Monthly* 72.4 (1965), pp. 343–359.

[147] Theodore Hailperin. *Logic with a Probability Semantics: Including Solutions to Some Philosophical Problems*. Plymouth, UK: Lehigh University Press, 2011.

[148] Theodore Hailperin. "Probability Logic". In: *Notre Dame J. Formal Logic* 25.3 (1984), pp. 198–212.

[149] Theodore Hailperin. "Probability logic and combining evidence". In: *Hist. Philos. Logic* 27.3 (2006), pp. 249–269.

[150] Theodore Hailperin. *Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications*. London: Associated University Presses, 1996.

[151] Hermann Haken, ed. *Complex Systems – Operational Approaches: in Neurobiology, Physics, and Computers. Proceedings of the International Symposium on Synergetics at Schloß Elmau, Bavaria, May 6–11, 1985*. Vol. 31. Springer series in synergetics. Berlin: Springer, 1985.

[152] K.D. Harris. "Neural signatures of cell assembly organization". In: *natrevnsci* 5.6 (May 2005), pp. 339–407.

[153] Douglas Rayne Hartree. "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods". In: *Proc. Cambridge Philos. Soc.* 24.1 (1928). `http://sci-prew.inf.ua/index1.htm`; see also [154], pp. 89–110.

[154] Douglas Rayne Hartree. "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion". In: *Proc. Cambridge Philos. Soc.* 24.1 (1928). `http://sci-prew.inf.ua/index1.htm`; see [153], pp. 111–132.

[155] David Heath, William Sudderth. "De Finetti's Theorem on Exchangeable Variables". In: *American Statistician* 30.4 (1976), pp. 188–189.

[156] D. O. Hebb. *The organization of behavior: A neuropsychological theory*. New York: John Wiley & Sons, 1949.

[157] Donald Olding Hebb. *The Organization of Behaviour: A Neuropsychological Theory*. John Wiley & Sons, 1967.

[158] Moritz Helias, Tom Tetzlaff, Markus Diesmann. "Echoes in correlated neural systems". In: *New J. Phys.* 15 (2013), p. 023002.

[159] Moritz Helias, Tom Tetzlaff, Markus Diesmann. “The correlation structure of local cortical networks intrinsically results from recurrent dynamics”. In: *PLoS Computational Biology* 10.1 (2014), e1003428.

[160] John A. Hertz, Yasser Roudi, Joanna Tyrcha. “Ising Models for Inferring Network Structure from Spike Data”. In: [275]. `arXiv:1106.1752`. 2013. Chap. ch. 27, pp. 527–546.

[161] G. E. Hinton, T. J. Sejnowski. “Learning and Relearning in Boltzmann Machines”. In: [294]. First publ. 1986; `https://papers.cnl.salk.edu/PDFs/Learning%20and%20Relearning%20in%20Boltzmann%20Machines%201986-3239.pdf`. 1999. Chap. ch. 7, pp. 282–317, 507–516.

[162] Arthur Hobson. “A new theorem of information theory”. In: *J. Stat. Phys.* 1.3 (1969), pp. 383–391.

[163] Arthur Hobson. *Concepts in Statistical Mechanics*. New York: Gordon and Breach, 1971.

[164] Arthur Hobson. “Irreversibility in simple systems”. In: *Am. J. Phys.* 34 (1966), pp. 411–416.

[165] Arthur Hobson, Bin-Kang Cheng. “A comparison of the Shannon and Kullback information measures”. In: *J. Stat. Phys.* 7.4 (1973), pp. 301–310.

[166] Arthur Hobson, David N. Loomis. “Exact classical nonequilibrium statistical-mechanical analysis of the finite ideal gas”. In: *Phys. Rev.* 173.1 (1968). `http://ergodic.ugr.es/statphys_grado/bibliografia/hoobson.pdf`, pp. 285–295.

[167] Sungho Hong, Stéphanie Ratté, Steven A. Prescott, Erik De Schutter. “Single neuron firing properties impact correlation-based population coding”. In: 32.4 (Jan. 2012), pp. 1413–1428.

[168] D. H. Hubel, T. N. Wiesel. “Receptive fields of single neurones in the cat’s striate cortex”. In: *J Physiol* 148 (1959), pp. 574–591.

[169] Ernst Ising. “Beitrag zur Theorie des Ferromagnetismus”. In: *Z. Phys.* 31.1 (1925), pp. 253–258.

[170] J Ito, P Maldonado, W Singer, S Grün. “Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes”. In: *Cereb Cortex* 21.11 (Nov. 2011), pp. 2482–2497.

[171] F. James. “Monte Carlo theory and practice”. In: *Rep. Prog. Phys.* 43.9 (1980), pp. 1145–1189.

[172] Edwin Thompson Jaynes. “Bayesian methods: general background”. In: [190]. `http://bayes.wustl.edu/etj/node1.html`. 1986, pp. 1–25.

[173] Edwin Thompson Jaynes. *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*. reprint. Ed. by R. D. Rosenkrantz. First publ. 1983. Dordrecht: Kluwer, 1989.

[174] Edwin Thompson Jaynes. "Foundations of probability theory and statistical mechanics". In: [40]. `http://bayes.wustl.edu/etj/node1.html`. 1967, pp. 77–101.

[175] Edwin Thompson Jaynes. "Information theory and statistical mechanics". In: *Phys. Rev.* 106.4 (1957). `http://bayes.wustl.edu/etj/node1.html`, see also [177], pp. 620–630.

[176] Edwin Thompson Jaynes. "Information theory and statistical mechanics". In: [88]. Repr. in [173, ch. 4, pp. 39–76]; `http://bayes.wustl.edu/etj/node1.html`. 1963, pp. 181–218.

[177] Edwin Thompson Jaynes. "Information theory and statistical mechanics. II". In: *Phys. Rev.* 108.2 (1957). `http://bayes.wustl.edu/etj/node1.html`, see also [175], pp. 171–190.

[178] Edwin Thompson Jaynes. "Macroscopic prediction". In: [151]. Updated version 1996 at `http://bayes.wustl.edu/etj/node1.html`. 1985, pp. 254–269.

[179] Edwin Thompson Jaynes. "Monkeys, kangaroos, and N". In: [190]. See also the rev. and corrected version [180]. 1986, pp. 26–58.

[180] Edwin Thompson Jaynes. *Monkeys, kangaroos, and N*. `http://bayes.wustl.edu/etj/node1.html`; first publ. 1986. (Errata: in equations (29)–(31), (33), (40), (44), (49) the commas should be replaced by gamma functions, and on p. 19 the value 0.915 should be replaced by 0.0915). 1996.

[181] Edwin Thompson Jaynes. "On the Rationale of Maximum-Entropy Methods". In: *Proc. IEEE* 70.9 (1982). `http://bayes.wustl.edu/etj/node1.html`, p. 939.

[182] Edwin Thompson Jaynes. *Probability Theory: The Logic of Science*. Ed. by G. Larry Bretthorst; `http://omega.albany.edu:8008/JaynesBook.html`, `http://omega.albany.edu:8008/JaynesBookPdf.html`, `http://www-biba.inrialpes.fr/Jaynes/prob.html`; first publ. 1994. Cambridge: Cambridge University Press, 2003.

[183] Edwin Thompson Jaynes. "Some applications and extensions of the de Finetti representation theorem". In: *Bayesian Inference and Decision Techniques with Applications: Essays in Honor of Bruno de Finetti*. Ed. by Prem K. Goel, Arnold Zellner. `http://bayes.wustl.edu/etj/node1.html`. Amsterdam: North-Holland, 1986, p. 31.

[184] Edwin Thompson Jaynes. "The Gibbs paradox". In: *Maximum-Entropy and Bayesian Methods*. Ed. by C. R. Smith, G. J. Erickson, P. O. Neudorfer. `http://bayes.wustl.edu/etj/node1.html`. Dordrecht: Kluwer, 1992, pp. 1–22.

[185] Edwin Thompson Jaynes. "The minimum entropy production principle". In: *Annu. Rev. Phys. Chem.* 31 (1980). http://bayes.wustl.edu/etj/node1.html, p. 579.

[186] Edwin Thompson Jaynes. "Where do we go from here?" In: [320]. http://bayes.wustl.edu/etj/node1.html. 1985, pp. 21–58.

[187] Edwin Thompson Jaynes. "Where do we Stand on Maximum Entropy?" In: [221]. http://bayes.wustl.edu/etj/node1.html; repr. with an introduction in [173], pp. 210–314. 1979, pp. 15–118.

[188] Harold Jeffreys. *Scientific Inference*. 3rd ed. First publ. 1931. Cambridge: Cambridge University Press, 1973.

[189] Harold Jeffreys. *Theory of Probability*. 3rd ed. First publ. 1939. London: Oxford University Press, 2003.

[190] James H. Justice, ed. *Maximum Entropy and Bayesian Methods in Applied Statistics: Proceedings of the Fourth Maximum Entropy Workshop, University of Calgary, 1984*. First publ. 1986. Cambridge: Cambridge University Press, 2008.

[191] Olav Kallenberg. *Probabilistic Symmetries and Invariance Principles*. Probability and its applications. New York: Springer, 2005.

[192] Eric R. Kandel, James H. Schwartz, Thomas M. Jessell, Steven A. Siegelbaum, A. J. Hudspeth, eds. *Principles of Neural Science*. 5th ed. New York: McGraw-Hill, 2013.

[193] Kyozi Kawasaki. "Kinetics of Ising Models". In: [71]. 1972. Chap. ch. 11, pp. 443–501.

[194] F. P. Kelly. *Reversibility and Stochastic Networks*. http://www.statslab.cam.ac.uk/~frank/BOOKS/kelly_book.html. Chichester: Wiley, 1979.

[195] D. G. Kendall. "On Finite and Infinite Sequences of Exchangeable Events". In: *Studia Sci. Math. Hung.* 2 (1967), pp. 319–327.

[196] B. E. Kilavik, S. Roux, A. Ponce-Alvarez, J. Confais, S. Gruen, A. Riehle. "Long-term Modifications in Motor Cortical Dynamics induced by intensive practice". In: *JNeurosci* 29 (2009), pp. 12653–12663.

[197] Scott Kirkpatrick, David Sherrington. "Infinite-ranged models of spin-glasses". In: *Phys. Rev. B* 17.11 (1978), pp. 4384–4403.

[198] Kevin H. Knuth. *Optimal Data-Based Binning for Histograms*. arXiv:physics/0605197. First publ. 2006. 2013.

[199] Giorgio Koch, Fabio Spizzichino, eds. *Exchangeability in Probability and Statistics: Proceedings of the International Conference on Exchangeability in Probability and Statistics*. Amsterdam: North-Holland, 1982.

[200] Adam Kohn, Matthew A. Smith. “Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque”. In: *J. Neurosci.* 25.14 (2005). `http://www.smithlab.net/publications.html`, pp. 3661–3673.

[201] Peter König, Andreas K. Engel, Pieter R. Roelfsema, Wolf Singer. “How Precise is Neuronal Synchronization”. In: *NeuralComput* 7 (1995), pp. 469–485.

[202] Peter König, Andreas K. Engel, Wolf Singer. “Integrator or coincidence detector? The role of the cortical neuron revisited”. In: *tins* 19.4 (1996), pp. 130–137.

[203] Werner Krauth. *Statistical Mechanics: Algorithms and Computations*. Oxford master series in statistical, computational, and theoretical physics. Oxford: Oxford University Press, 2006.

[204] Birgit Kriener, Tom Tetzlaff, Ad Aertsen, Markus Diesmann, Stefan Rotter. “Correlations and population dynamics in cortical networks”. In: *NeuralComput* 20 (2008), pp. 2185–2226.

[205] Paruchuri R. Krishnaiah, P. K. Sen, eds. *Handbook of Statistics 4: Nonparametric Methods*. Amsterdam: North-Holland, 1984.

[206] A. Kuhn, A. Aertsen, S. Rotter. “Higher-order statistics of input ensembles and the response of simple model neurons”. In: *NeuralComput* 15.1 (2003), pp. 67–101.

[207] Solomon Kullback. *Information Theory and Statistics*. Republ. with a new preface and corrections and additions by the author; first publ. 1959. New York: Dover, 1978.

[208] Solomon Kullback. “The Kullback-Leibler distance”. In: *American Statistician* 41.4 (1987), pp. 340–341.

[209] Solomon Kullback, John C. Keegel. “Categorical Data Problems Using Information Theoretic Approach”. In: [205]. 1984. Chap. ch. 35, pp. 831–873.

[210] Solomon Kullback, John C. Keegel, Joseph H. Kullback. *Topics in Statistical Information Theory*. Vol. 42. Lecture notes in statistics. Berlin: Springer, 1987.

[211] Solomon Kullback, R. A. Leibler. “On Information and Sufficiency”. In: *Ann. Math. Stat.* 22.1 (1951), pp. 79–86.

[212] David P. Landau, Kurt Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. 4th ed. `http://el.us.edu.pl/ekonofizyka/images/6/6b/A_guide_to_monte_carlo_simulations_in_statistical_physics.pdf`, `http://iop.vast.ac.vn/~nvthanh/cours/simulation/MC_book.pdf`. First publ. 2000. Cambridge: Cambridge University Press, 2015.

[213] Pierre Simon Laplace (Marquis de). *Essai philosophique sur les probabilités*. 4th ed. Repr. as the Introduction of [215], pp. V–CLIII; `http://gallica.bnf.fr/document?O=N077595`. Paris: Courcier, 1819.

[214] Pierre Simon Laplace (Marquis de). *Œuvres complètes de Laplace. Tome septième: Théorie analytique des probabilités*. 'Publiées sous les auspices de l'Académie des sciences, par MM. les secrétaires perpétuels'; http://gallica.bnf.fr/notice?N=FRBNF30739022. Paris: Gauthier-Villars, 1886.

[215] Pierre Simon Laplace (Marquis de). *Théorie analytique des probabilités*. 3rd ed. First publ. 1812; repr. in [214]; http://gallica.bnf.fr/document?O=N077595. Paris: Courcier, 1820.

[216] Peter E. Latham, Yasser Roudi. "Role of Correlations in Population Coding". In: [275]. arXiv:1109.6524. 2013. Chap. ch. 7, pp. 121–138.

[217] Steffen L. Lauritzen. "Extreme Point Models in Statistics". In: *Scand. J. Statist.* 11.2 (1984). See also discussion [20], pp. 65–83.

[218] Joel L. Lebowitz, Christian Maes. "Entropy: a Dialogue". In: [130]. 2003. Chap. ch. 13, pp. 269–276.

[219] Julian Lee, Steve Pressé. "A derivation of the master equation from path entropy maximization". In: *J. Chem. Phys.* 137.7 (2012), p. 074103.

[220] Wilhelm Lenz. "Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern". In: *Physik. Zeitschr.* XXI (1920). http://www.physik.uni-rostock.de/fileadmin/Physik/Mahnke/Geschichte/Wilhelm_Lenz/Lenz_1920.pdf, pp. 613–615.

[221] Raphael D. Levine, Myron Tribus, eds. *The Maximum Entropy Formalism: A Conference Held at the Massachusetts Institute of Technology on May 2–4, 1978*. Cambridge, USA: The MIT Press, 1979.

[222] Michael London, Arnd Roth, Lisa Beeren, Michael Häusser, Peter E. Latham. "Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex". In: 466.1 (July 2010), pp. 123–128. DOI: 10.1038/nature09086.

[223] S. Louis, G. L. Gerstein, S. Grün, M. Diesmann. "Surrogate spike train generation through dithering in operational time". In: *Front. Comput. Neurosci.* 4.127 (2010).

[224] David J. C. MacKay. "Bayesian Interpolation". In: *Neural Comp.* 4.3 (1992). http://www.inference.phy.cam.ac.uk/mackay/PhD.html, pp. 415–447.

[225] David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. http://www.inference.phy.cam.ac.uk/mackay/itila/; first publ. 1995. Cambridge: Cambridge University Press, 2003.

[226] Jakob H. Macke, Philipp Berens, Alexander S. Ecker, Andreas S. Tolias, Matthias Bethge. "Generating Spike Trains with Specified Correlation Coefficients". In: *Neural Comp.* 21.2 (2009), pp. 397–423.

[227] Jakob H. Macke, Lars Buesing, John P. Cunningham, Byron M. Yu, Krishna V. Shenoy, Maneesh Sahani. "Empirical models of spiking in neural populations". In: *Advances in Neural Information Processing Systems (NIPS proceedings)* 24 (2011), pp. 1350–1358.

[228] Jakob H. Macke, Iain Murray, Peter E. Latham. "Estimation Bias in Maximum Entropy Models". In: *Entropy* 15.8 (2013). `http://www.gatsby.ucl.ac.uk/~pel/papers/maxentbias.pdf`, pp. 3109–3129.

[229] Jakob H. Macke, Manfred Opper, Matthias Bethge. "Common Input Explains Higher-Order Correlations and Entropy in a Simple Model of Neural Population Activity". In: *Phys. Rev. Lett.* 106.20 (2011). `arXiv:1009.2855`, p. 208102.

[230] Jakob H. Macke, Manfred Opper, Matthias Bethge. *The effect of pairwise neural correlations on global population statistics*. Tech. rep. 183. `http://www.kyb.tuebingen.mpg.de/publications/attachments/MPIK-TR-183_%5B0%5D.pdf`. Tübingen: Max-Planck-Institut für biologische Kybernetik, 2009.

[231] Christian Maes, Karel Netočný. "Time-Reversal and Entropy". In: *J. Stat. Phys.* 110.1–2 (2003). `arXiv:cond-mat/0202501`, pp. 269–310.

[232] Christian Maes, Frank Redig, Annelies Van Moffaert. "The Restriction of the Ising Model to a Layer". In: *J. Stat. Phys.* 96.1 (1999). `arXiv:math/9810094`, pp. 69–107.

[233] Pedro Maldonado, Cecilia Babul, Wolf Singer, Eugenio Rodriguez, Denise Berger, Sonja Grün. "Synchronization of Neuronal Responses in Primary Visual Cortex of Monkeys Viewing Natural Images". In: 100.3 (2008), pp. 1523–1532.

[234] Olivier Marre, S. El Boustani, Y. Frégnac, A. Destexhe. "Prediction of Spatiotemporal Patterns of Neural Activity from Pairwise Correlations". In: *Phys. Rev. Lett.* 102.13 (2009). `arXiv:0903.0127`, p. 138101.

[235] L. Martignon, H. Von Hassein, Sonja Grün, Ad Aertsen, Günther Palm. "Detecting higher-order interactions among the spiking events in a group of neurons". In: *Biol. Cybern.* 73.1 (1995), pp. 69–81.

[236] L. Martignon, H. von Hasseln, S. Grün, A. Aertsen, G. Palm. "Detecting higher-order interactions among the spiking events in a group of neurons". In: *BiolCybern* 73 (1995), pp. 69–81.

[237] James Clerk Maxwell. *Introductory Lecture on Experimental Physics*. Repr. in [240], pp. 241–255. 1871.

[238] James Clerk Maxwell. "On the Dynamical Evidence of the Molecular Constitution of Bodies". In: *Nature* XI.279, 280 (1875). Repr. in [240], pp. 418–438, pp. 357–359, 374–377.

[239] James Clerk Maxwell. "Tait's "Thermodynamics"". In: *Nature* 17.431, 432 (1878). Repr. in [240], pp. 660–671, pp. 257–259, 278–280.

[240] James Clerk Maxwell. *The Scientific Papers of James Clerk Maxwell. Vol. Two.* Ed. by W. D. Niven. Two volumes bound as one. http://num-scd-ulp.u-strasbg.fr:8080/347/. First publ. 1890. New York: Dover, 1965.

[241] Mark E. Mazurek, Michael N. Shadlen. "Limits to the temporal fidelity of cortical spike rate signals". In: *Nat. Neurosci.* 5.5 (2002). https://www.shadlenlab.columbia.edu/publications/publications/mike/mazurek_shadlen2002.pdf, pp. 463–471.

[242] C. Alden Mead. "The special role of maximum entropy in the application of "mixing character" to irreversible processes in macroscopic systems". In: [221]. 1979, pp. 273–287.

[243] Lawrence R. Mead, N. Papanicolaou. "Maximum entropy in the problem of moments". In: *J. Math. Phys.* 25.8 (1984). http://bayes.wustl.edu/Manual/MeadPapanicolaou.pdf, pp. 2404–2417.

[244] N. David Mermin. "Improving an improved analytical approximation to $n!$". In: *Am. J. Phys.* 51.9 (1983). See also [35, 62, 261, 349, 355], p. 776.

[245] Tomislav Milekovic, Wilson Truccolo, Sonja Grün, Alexa Riehle, Thomas Brochier. "Local field potentials in primate motor cortex encode grasp kinetic parameters". In: *NeuroImage* 114 (2015), pp. 338–355. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2015.04.008.

[246] P M Milner. "A model for visual shape recognition." In: *Psychological review* 81.6 (Nov. 1974), pp. 521–35. ISSN: 0033-295X. URL: http://www.ncbi.nlm.nih.gov/pubmed/4445414.

[247] Thierry Mora, Stéphane Deny, Olivier Marre. "Dynamical Criticality in the Collective Activity of a Population of Retinal Neurons". In: *Phys. Rev. Lett.* 114.7 (2015). arXiv:1410.6769, p. 078105.

[248] Rubén Moreno-Bote, Jeffrey Beck, Ingmar Kanitscheider, Xaq Pitkow, Peter Latham, Alexandre Pouget. "Information-limiting correlations". In: *Nature Neuroscience* 17.10 (Sept. 2014), pp. 1410–1417. ISSN: 1097-6256. DOI: 10.1038/nn.3807. URL: http://www.nature.com/doifinder/10.1038/nn.3807.

[249] M P Nawrot, C Mehring, V Rodriguez-Molina, D Heck, A Aertsen, S Rotter. "How renewal-like are cortical neurons? Investigation of serial interval statistics in vitro and in vivo". In: *Proceedings of the 6th International IBRO World Congress 2003.* Prague, 2003, p. 4382.

[250] Martin P Nawrot. "Ongoing activity in cortical networks: noise, variability and context". PhD thesis. Albert-Ludwigs-University Freiburg, June 2003.

[251] Martin P Nawrot, C Boucsein, V Rodriguez Molina, A Riehle, A Aertsen, S Rotter. "Measurement of variability dynamics in cortical spike trains". In: *JNeurosciMethods* 169 (2008), pp. 374–390.

[252] John W. Negele, Henri Orland. *Quantum Many-Particle Systems*. Advanced book classics. First publ. 1988. Reading, USA: Perseus Books, 1998.

[253] Jerzy Neyman, ed. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Held at the Statistical Laboratory, University of California, June 20–July 30, 1960. Vol. I: Contributions to the Theory of Statistics*. Berkeley and Los Angeles: University of California Press, 1961.

[254] M. A. L. Nicolelis, ed. *Methods for Neural Ensemble Recordings*. Boca Raton, Florida: CRC Press, 1998.

[255] Keit B. Oldham, Jan C. Myland, Jerome Spanier. *An Atlas of Functions: with* Equator*, the Atlas Function Calculator*. 2nd ed. First publ. 1987. New York: Springer, 2009.

[256] Lars Onsager. "Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition". In: *Phys. Rev.* 65.3–4 (1944), pp. 117–149.

[257] Manfred Opper, David Saad, eds. *Advanced Mean Field Methods: Theory and Practice*. Neural information processing series. Cambridge, USA: The MIT Press, 2001.

[258] Manfred Opper, Ole Winther. "From Naive Mean Field Theory to the TAP Equations". In: [257]. `http://cogsys.imm.dtu.dk/staff/winther/opper.fromnaive.pdf`. 2001. Chap. ch. 2, pp. 7–20.

[259] Stefano Panzeri, Jakob H. Macke, Joachim Gross, Christoph Kayser. "Neural population coding: combining insights from microscopic and mass signals". In: *Trends Cognit. Sci.* 19.3 (2015), pp. 162–172.

[260] Stefano Panzeri, Simon R. Schultz, Alessandro Treves, Edmund T. Rolls. "Correlations and the encoding of information in the nervous system". In: *Proc. R. Soc. Lond. B* 266.1423 (1999). `http://people.sissa.it/~ale/Pan+99a.pdf`, pp. 1001–1012.

[261] R. K. Pathria. "Mathematical basis for Weissman's approximation to $n!$". In: *Am. J. Phys.* 53.1 (1985). See also [35, 62, 244, 349, 355], pp. 81–82.

[262] Judea Pearl. *Causality: Models, Reasoning, and Inference*. 2nd ed. First publ. 2000. Cambridge: Cambridge University Press, 2009.

[263] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. rev. second printing. The Morgan Kaufmann series in representation and reasoning. San Francisco: Morgan Kaufmann, 1988.

[264] Rudolf E. Peierls. “Quelques propriétés typiques des corps solides”. In: *Ann. Inst. Henri Poincaré* 5.3 (1935), pp. 177–222.

[265] Donald H. Perkel, George L. Gerstein, George P. Moore. “Neuronal Spike Trains and Stochastic Point Processes. II. Simultaneous Spike Trains”. In: *Biophys. J.* 7.4 (1967), pp. 419–440.

[266] Volker Pernice, Benjamin Staude, Stefano Cardanobile, Stefan Rotter. “How Structure Determines Correlations in Neuronal Networks”. In: *PLoS Comput. Biol.* 7.5 (May 2011), e1002059.

[267] G. Pipa, S. Grün. “Non-Parametric Significance Estimation of Joint-Spike Events by Shuffling and Resampling”. In: 52–54 (2003), pp. 31–37.

[268] Gordon Pipa, Markus Diesmann, Sonja Grün. “Significance of Joint-Spike Events Based on Trial-Shuffling by Efficient Combinatorial Methods”. In: *Complexity* 8.4 (2003), pp. 79–86.

[269] Gordon Pipa, Sonja Grün, Carl Van Vreeswijk. “Impact of Spike Train Autostructure on Probability Distribution of Joint Spike Events”. In: *Neural Comp.* 1163 (2013), pp. 1123–1163.

[270] Piero Giovanni Luca Porta Mana. *On the relation between plausibility logic and the maximum-entropy principle: a numerical study.* `arXiv:0911.2197`. Also presented as invited talk at the 31st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering ‘MaxEnt 2011’, Waterloo, Canada. 2009.

[271] Piero Giovanni Luca Porta Mana, Emiliano Torre, Vahid Rostami. *Inferences from a network to a subnetwork and vice versa under an assumption of symmetry.* `bioRxiv:2015/12/11/034199`. 2015.

[272] R. B. Potts. “Note on the Factorial Moments of Standard Distributions”. In: *Aust. J. Phys.* 6.4 (1953), pp. 498–499.

[273] William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery. *Numerical Recipes: The Art of Scientific Computing.* 3rd ed. First publ. 1988. Cambridge: Cambridge University Press, 2007.

[274] Steve Pressé, Kingshuk Ghosh, Julian Lee, Ken A. Dill. “Principles of maximum entropy and maximum caliber in statistical physics”. In: *Rev. Mod. Phys.* 85.3 (2013). `http://statphysbio.physics.iupui.edu/Presse_rmp.pdf`, pp. 1115–1141.

[275] Rodrigo Quian Quiroga, Stefano Panzeri, eds. *Principles of Neural Coding.* Boca Raton, USA: CRC Press, 2013.

[276] Alfonso Renart, Jaime De La Rocha, Peter Bartho, Liad Hollender, N’estor Parga, Alex Reyes, Kenneth D. Harris. “The asynchronous State in Cortical Circuits”. In: 327 (Jan. 2010), pp. 587–590.

[277] Alfred Rényi. “On measures of entropy and information”. In: [253]. 1961, pp. 547–561.

[278] Alexa Riehle, Sonja Grün, Markus Diesmann, Ad Aertsen. “Spike Synchronization and Rate Modulation Differentially Involved in Motor Cortical Function”. In: *science* 278 (1997), pp. 1950–1953.

[279] Alexa Riehle, Sarah Wirtssohn, Sonja Grün, Thomas Brochier. “Mapping the spatio-temporal structure of motor cortical LFP and spiking activities during reach-to-grasp movements”. In: *Frontiers in Neural Circuits* 7 (2013). doi: 10.3389/fncir.2013.00048, p. 48.

[280] Herbert Robbins. “A Remark on Stirling’s Formula”. In: *Am. Math. Monthly* 62.1 (1955), pp. 26–29.

[281] J. A. Roberts, ed. *Indirect Imaging: Measurement and processing for indirect imaging (Proceedings of an International Symposium, Sydney, Australia, 30 August to 2 September 1983)*. Cambridge: Cambridge University Press, 1984.

[282] R. Tyrrell Rockafellar. “Lagrange Multipliers and Optimality”. In: *SIAM Rev.* 35.2 (1993). `http://www.math.washington.edu/~rtr/papers.html`, pp. 183–238.

[283] Carlos C. Rodríguez. *A Geometric Theory of Ignorance*. `http://omega.albany.edu:8008/`. 2003.

[284] Carlos C. Rodríguez. *Entropic Priors*. `http://omega.albany.edu:8008/`. 1991.

[285] Carlos C. Rodríguez. *Entropic Priors for Discrete Probabilistic Networks and for Mixtures of Gaussians Models*. `arXiv:physics/0201016`; `http://omega.albany.edu:8008/`. 2002.

[286] Sheldon Ross. *A First Course in Probability*. 8th ed. First publ. 1976. Upper Saddle River, USA: Pearson, 2010.

[287] Vahid Rostami. “Pull Request of the Unitary Events method in Elephant”. In: *Elephant Github* (2016). URL: `https://github.com/NeuralEnsemble/elephant/pull/64`.

[288] Vahid Rostami, Junji Ito, Sonja Grün. “Assessment of Higher-Order Synchrony in Massively Parallel Spike Trains by the Population Unitary Events Analysis”. In: *Bernstein Conference 2015* (2015).

[289] Vahid Rostami, Piero Giovanni Luca Porta Mana, Moritz Helias. *Pairwise maximum-entropy models and their Glauber dynamics: bimodality, bistability, non-ergodicity problems, and their elimination via inhibition*. `arXiv:1605.04740`. 2016.

[290] Yasser Roudi, Erik Aurell, John A. Hertz. “Statistical physics of pairwise probability models”. In: *Frontiers in Computational Neuroscience* 3.22 (2009), doi: 10.3389/neuro.10.022.2009.

[291] Yasser Roudi, John Hertz. "Mean Field Theory for Nonequilibrium Network Reconstruction". In: *Phys. Rev. Lett.* 106.4 (2011). arXiv:1009.5946, p. 048702.

[292] Yasser Roudi, Sheila Nirenberg, Peter E. Latham. "Pairwise Maximum Entropy Models for Studying Large Biological Systems: When They Can Work and When They Can't". In: *PLoS Computational Biology* 5.5 (2009). arXiv:0811.0903, e1000380.

[293] Yasser Roudi, Joanna Tyrcha, John Hertz. "Ising model for neural data: Model quality and approximate methods for extracting functional connectivity". In: *Phys. Rev. E* 79.5 (2009). arXiv:0902.2885, p. 051915.

[294] David E. Rumelhart, James L. McClelland, the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*. 12th printing. Computational models of cognition and perception. First publ. 1986. Cambridge, USA: The MIT Press, 1999.

[295] Y. Sakurai. "Population coding by cell assemblies - What it really is in the brain". In: *Neuroscience Research* 26 (1996), pp. 1–16.

[296] Emilio Salinas, Terrence J. Sejnowski. "Correlated neuronal activity and the flow of neural information". In: *NatRevNeurosci* 2.8 (2001), pp. 539–550.

[297] Elad Schneidman, Michael J. Berry II, Ronen Segev, William Bialek. "Weak pairwise correlations imply strongly correlated network states in a neural population". In: *Nature* 440.7087 (2006). arXiv:q-bio/0512013, http://www.weizmann.ac.il/neurobiology/labs/schneidman/The_Schneidman_Lab/Publications.html, pp. 1007–1012.

[298] Roberto H. Schonmann. "Projections of Gibbs measures may be non-Gibbsian". In: *Commun. Math. Phys.* 124.1 (1989), pp. 1–7.

[299] Matthias Schultze-Kraft, Markus Diesmann, Sonja Gruen, Moritz Helias. "Noise Suppression and Surplus Synchrony by Coincidence Detection". In: *PLoS Comput Biol* 9.4 (2013), e1002904. DOI: 10.1371/journal.pcbi.1002904.

[300] Vitor Sessak. "Inverse problems in spin models". arXiv:1010.1960. PhD thesis. Paris: École Normale Supérieure, 2010.

[301] Vitor Sessak, Rémi Monasson. "Small-correlation expansions for the inverse Ising problem". In: *J. Phys. A* 42.5 (2009). arXiv:0811.3574, p. 055001.

[302] Michael N. Shadlen, William T. Newsome. "Noise, neural codes and cortical organization". In: 4.4 (1994), pp. 569–579.

[303] Michael N. Shadlen, William T. Newsome. "The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding". In: 18.10 (1998), pp. 3870–3896.

[304] Claude Elwood Shannon. “A mathematical theory of communication”. In: *Bell Syst. Tech. J.* 27.3, 4 (1948). http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html, http://www.cparity.com/it/demo/external/shannon.pdf, pp. 379–423, 623–656.

[305] Claude Elwood Shannon. “Communication In The Presence Of Noise”. In: *Proc. IRE* 37.1 (1949). Repr. in [306], pp. 10–21.

[306] Claude Elwood Shannon. “Communication In The Presence Of Noise”. In: *Proc. IEEE* 86.2 (1998). Repr. of [305], with an introduction [361], pp. 447–457.

[307] Eric Shea-Brown, Kresimir Josic, Jaime de la Rocha, Brent Doiron. “Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding”. In: *Phys. Rev. Lett.* 100 (Mar. 2008), p. 108102.

[308] David Sherrington, Scott Kirkpatrick. “Solvable Model of a Spin-Glass”. In: *Phys. Rev. Lett.* 35.26 (1975), pp. 1792–1796.

[309] Hideaki Shimazaki, Shun-ichi Amari, Emery N. Brown, Sonja Grün. “State-Space Analysis of Time-Varying Higher-Order Spike Correlation for Multiple Neural Spike Train Data”. In: *PLoS Computational Biology* 8.3 (2012). http://hdl.handle.net/1721.1/71752, e1002385.

[310] Hideaki Shimazaki, Kolia Sadeghi, Tomoe Ishikawa, Yuji Ikegaya, Taro Toyoizumi. “Simultaneous silence organizes structured higher-order interactions in neural populations”. In: *Sci. Rep.* 5 (2015), p. 9821.

[311] Jonathon Shlens, Greg D. Field, Jeffrey L. Gauthier, Matthew I. Grivich, Dumitru Petrusca, Alexander Sher, Alan M. Litke, E. J. Chichilnisky. “Correction, The Structure of Multi-Neuron Firing Patterns in Primate Retina”. In: *J. Neurosci.* 28.5 (2008). See [312], p. 1246.

[312] Jonathon Shlens, Greg D. Field, Jeffrey L. Gauthier, Matthew I. Grivich, Dumitru Petrusca, Alexander Sher, Alan M. Litke, E. J. Chichilnisky. “The Structure of Multi-Neuron Firing Patterns in Primate Retina”. In: *J. Neurosci.* 26.32 (2006). See also correction [311], pp. 8254–8266.

[313] W Singer. “Neuronal synchrony: a versatile code for the definition of relations?” In: *Neuron* 24.1 (Sept. 1999), pp. 49–65.

[314] W. Singer, A. K. Engel, A. K. Kreiter, M. H. J. Munk, S. Neuenschwander, P. R. Roelfsema. “Neuronal assemblies: necessity, signature and detectability”. In: *Trends in Cognitive Sciences* 1.7 (1997), pp. 252–261.

[315] Devinderjit Singh Sivia. “Bayesian Inductive Inference, Maximum Entropy, & Neutron Scattering”. In: *Los Alamos Science* 4.19 (1990). http://www.fas.org/sgp/othergov/doe/lanl/pubs/number19.htm, pp. 180–207.

[316] John Skilling. “Classic maximum entropy”. In: [**skilling1989**]. 1989, pp. 45–52.

[317] John Skilling. “Quantified maximum entropy”. In: [90]. 1990, pp. 341–350.

[318] John Skilling. “Theory of maximum entropy image reconstruction”. In: [190]. 1986, pp. 156–178.

[319] John Skilling, R. K. Bryan. “Maximum entropy image reconstruction: general algorithm”. In: *Mon. Not. Roy. Astron. Soc.* 211.1 (1984), pp. 111–124.

[320] C. Ray Smith, Walter T. Grandy Jr., eds. *Maximum-Entropy and Bayesian Methods in Inverse Problems*. Fundamental theories of physics. Dordrecht: D. Reidel, 1985.

[321] Matthew A. Smith, Adam Kohn. “Spatial and Temporal Scales of Neuronal Correlation in Primary Visual Cortex”. In: *J. Neurosci.* 28.48 (2008). `http://www.smithlab.net/publications.html`, pp. 12591–12603.

[322] Donald L Snyder, Michael I Miller. *Random Point Processes in Time and Space*. Springer, 1991.

[323] William R. Softky, Christof Koch. “The Highly Irregular Firing of Cortical Cells Is Inconsistent with Temporal Integration of Random EPSPs”. In: 13.1 (1993), pp. 334–350.

[324] H. Sompolinsky, H. Yoon, K. Kang, M. Shamir. “Population coding in neuronal systems with correlated noise”. In: *Phys. Rev. E* 64.5 (Nov. 2001), p. 51904.

[325] Sompolinsky, Crisanti, Sommers. “Chaos in random neural networks.” eng. In: *Phys. Rev. Lett.* 61.3 (July 1988), pp. 259–262.

[326] Wolfgang Spohn. “Stochastic independence, causal independence, and shieldability”. In: *J. Philos. Logic* 9.1 (1980). `http://www.uni-konstanz.de/philosophie/files/06_spohn_stochastic__1c403d.pdf`, pp. 73–99.

[327] Benjamin Staude, Sonja Grün, Stefan Rotter. “Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference”. In: 4 (July 2010). Article nr 16, doi: 10.3389/fncom.2010.00016.

[328] Benjamin Staude, Stefan Rotter, Sonja Grün. “CuBIC: cumulant based inference of higher-order correlations in massively parallel spike trains”. In: 29 (2010). DOI 10.1007/s10827-009-0195-x, pp. 327–350.

[329] Ian H Stevenson, Konrad P Kording. “How advances in neural recording affect data analysis”. In: *Nature Neuroscience* 14.2 (Feb. 2011), pp. 139–142. ISSN: 1097-6256. DOI: `10.1038/nn.2731`. URL: `http://www.nature.com/doifinder/10.1038/nn.2731`.

[330] Stephen M. Stigler. “Thomas Bayes’s Bayesian Inference”. In: *J. Roy. Stat. Soc. A* 145.2 (1982), pp. 250–258.

[331] Toshiyuki Tanaka. “Information Geometry of Mean-Field Approximation”. In: [257]. 2001. Chap. ch. 17, pp. 259–273.

[332] Tom Tetzlaff, Moritz Helias, Gaute Einevoll, Markus Diesmann. “Decorrelation of neural-network activity by inhibitory feedback”. In: *PLoS Computational Biology* 8.8 (2012), e1002596.

[333] Tom Tetzlaff, Stefan Rotter, Eran Stark, Moshe Abeles, Ad Aertsen, Markus Diesmann. “Dependence of neuronal correlations on filter characteristics and marginal spike-train statistics”. In: *NeuralComput* 20.9 (2008), 2133–2184.

[334] D. J. Thouless, P. W. Anderson, R. G. Palmer. “Solution of ‘Solvable model of a spin glass’”. In: *Phil. Mag.* 35.3 (1977), pp. 593–601.

[335] Gašper Tkačik, Olivier Marre, Dario Amodei, Elad Schneidman, William Bialek, Michael J. Berry II. “Searching for Collective Behavior in a Large Network of Sensory Neurons”. In: *PLoS Computational Biology* 10.1 (2014). `arXiv:1306.3061`, e1003408.

[336] Gašper Tkačik, Olivier Marre, Thierry Mora, Dario Amodei, Michael J. Berry II, William Bialek. “The simplest maximum entropy model for collective behavior in a neural network”. In: *J. Stat. Mech.* 2013 (2013). `arXiv:1207.6319`, P03011.

[337] Gašper Tkačik, Thierry Mora, Olivier Marre, Dario Amodei, Stephanie E. Palmer, Michael J. Berry II, William Bialek. “Thermodynamics and signatures of criticality in a network of neurons”. In: *Proc. Natl. Acad. Sci. (USA)* 112.37 (2014). `arXiv:1407.5946`, pp. 11508–11513.

[338] Gašper Tkačik, Elad Schneidman, Michael J. Berry II, William Bialek. *Ising models for networks of real neurons*. `arXiv:q-bio/0611072`. 2006.

[339] Gašper Tkačik, Elad Schneidman, Michael J. Berry II, William Bialek. *Spin glass models for a network of real neurons*. `arXiv:0912.5409`. 2009.

[340] E. Torre, P. Quaglio, M. Denker, T. Brochier, A. Riehle, S. Grun. “Synchronous Spike Patterns in Macaque Motor Cortex during an Instructed-Delay Reach-to-Grasp Task”. In: *Journal of Neuroscience* 36.32 (Aug. 2016), pp. 8329–8340. ISSN: 0270-6474. DOI: `10.1523/JNEUROSCI.4375-15.2016`. URL: `http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4375-15.2016`.

[341] A Treisman. “Solutions to the binding problem: Progress through controversy and convergence”. In: *Neuron* 24 (1999), pp. 105–110.

[342] Myron Tribus. *Rational Descriptions, Decisions and Designs*. New York: Pergamon, 1969.

[343] Bryan Tripp, Chris Eliasmith. “Neural populations can induce reliable postsynaptic currents without observable spike rate changes or precise spike timing”. In: *CerebCortex* 17.8 (2007), pp. 1830–1840.

[344] J Trousdale, Y Hu, E Shea-Brown, K Josic. "Impact of network structure and cellular response on spike time correlations." In: *PLoS Comput. Biol.* 8.3 (2012), e1002408.

[345] Jan M. van Campenhout, Thomas M. Cover. "Maximum entropy and conditional probability". In: *IEEE Trans. Inform. Theor.* IT-27.4 (1981), pp. 483–489.

[346] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry.* 3rd ed. First publ. 1981. Amsterdam: North-Holland, 2007.

[347] Carl van Vreeswijk, Haim Sompolinsky. "Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity". In: 274 (Dec. 1996), pp. 1724–1726.

[348] Carl van Vreeswijk, Haim Sompolinsky. "Chaotic Balanced State in a Model of Cortical Circuits". In: 10 (1998), pp. 1321–1371.

[349] John D. Vedder. "Analytical approximations to $n!$". In: *Am. J. Phys.* 51.9 (1983). See also [35, 62, 244, 261, 355], pp. 778–779.

[350] Christoph von der Malsburg. "Am I thinking assemblies?" In: *Brain Theory.* Ed. by G Palm, A Aertsen. Berlin: Springer-Verlag, 1986, pp. 161–176.

[351] Christoph von der Malsburg. "Binding in models of perception and brain function". In: 5 (1995), pp. 520–526.

[352] Christoph von der Malsburg. *The Correlation Theory of Brain Function.* Internal report 81-2. Göttingen, Germany: Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, 1981.

[353] C. von der Malsburg. "The what and why of binding: the modeler's perspective." In: *Neuron* 24.1 (Sept. 1999), pp. 95–104. ISSN: 0896-6273. URL: `http://view.ncbi.nlm.nih.gov/pubmed/10677030`.

[354] Nicholas Weir. "Applications of Maximum Entropy Techniques to HST Data". In: [133]. 1991, pp. 115–129.

[355] Yitzhak Weissman. "An improved analytical approximation to $n!$". In: *Am. J. Phys.* 51.1 (1983). See also [35, 62, 244, 261, 349], p. 9.

[356] Mike West, Jeff Harrison. *Bayesian Forecasting and Dynamic Models.* 2nd ed. Springer series in statistics. First publ. 1989. New York: Springer, 1997.

[357] David H. Wolpert, David R. Wolf. "Erratum: Estimating functions of probability distributions from a finite set of samples". In: *Phys. Rev. E* 54.6 (1996). See [359], p. 6973.

[358] David H. Wolpert, David R. Wolf. *Estimating Functions of Distributions from A Finite Set of Samples, Part 2: Bayes Estimators for Mutual Information, Chi-Squared, Covariance and other Statistics.* `arXiv:comp-gas/9403002`, see also [359, 360]. 1994.

[359] David H. Wolpert, David R. Wolf. “Estimating functions of probability distributions from a finite set of samples”. In: *Phys. Rev. E* 52.6 (1995). See also erratum [357] and [358, 360], pp. 6841–6854.

[360] David H. Wolpert, David R. Wolf. *Estimating Functions of Probability Distributions from a Finite Set of Samples, Part 1: Bayes Estimators and the Shannon Entropy.* arXiv:comp-gas/9403001, see also [358, 359]. 1994.

[361] Aaron D. Wyner, Shlomo Shamai. “Introduction To “Communication In The Presence Of Noise””. In: *Proc. IEEE* 86.2 (1998). See [306], pp. 442–446.

[362] Lyuba Zehl, Florent Jaillet, Adrian Stoewer, Jan Grewe, Andrey Sobolev, Thomas Wachtler, Thomas G. Brochier, Alexa Riehle, Michael Denker, Sonja Grün. “Handling Metadata in a Neurophysiology Laboratory”. In: *Frontiers in Neuroinformatics* 10 (July 2016), p. 26. ISSN: 1662-5196. DOI: 10.3389/fninf.2016.00026. URL: http://journal.frontiersin.org/Article/10.3389/fninf.2016.00026/abstract.

[363] Arnold Zellner. *An Introduction to Bayesian Inference in Econometrics.* Wiley classics library. First publ. 1971. New York: Wiley, 1996.

[364] Ehud Zohary, Michael N. Shadlen, William T. Newsome. “Correlated neuronal discharge rate and its implications for psychophysical performance”. In: 370 (1994), pp. 140–143.