Analysis of Higher-Order Coincident Activity in Multiple Parallel Processes

August 2002

Gaby Schneider

Diplomarbeit

Eingereicht am Fachbereich Mathematik

der Johann Wolfgang Goethe-Universität Frankfurt am Main

slightly modified version of Diploma Thesis from June 2002

Acknowledgments

I would like to express my largest gratitude to my advisor Sonja Grün who introduced me in the field of computational neuroscience. During the cooperation with her, I had the pleasure to profit from her extensive knowledge, her clear and resourceful mind and her ever constructive comments. With her patience and enthusiasm, she manages to strike the perfect balance between providing direction and encouraging independence.

My sincere gratefulness goes to my mathematical advisor Brooks Ferebee whose continuous guidance and inventive ideas have been essential throughout the whole development of my thesis. During our long discussions, I had the opportunity to benefit from his astuteness, his kindness and his dedication.

I wish to thank Prof. Wakolbinger for his time and his effort, and above all for his interest in this interdisciplinary thesis. With his analytic expertise, his encouragement and his commitment, he provided invaluable guidance during the second period of my thesis.

I am very grateful to Prof. Singer for his inspiration and his constructive ideas, which have influenced me in many ways. I greatly appreciate his openness, his promotion and his constant support.

Thanks also to my peer Gordon Pipa for all helpful suggestions, critical remarks and unconventional ideas during our intensive Wednesday discussions.

Zusammenfassung

Motivation

Trotz umfangreicher Forschung auf dem Gebiet der Neurowissenschaften besteht bisher noch Unklarheit über die grundlegenden Prinzipien der Informationsverarbeitung im Neokortex. Im wesentlichen stehen sich zwei Hypothesen gegenüber, die einander nicht ausschließen müssen. Die "single-neuron doctrine" (Barlow, 1972) sieht die differenzierte Feuerrate einzelner Neurone mit spezifischen Antworteigenschaften und deren hierarchische Verknüpfung und Konvergenz auf Zellen mit hoch spezialisierten Stimuluseigenschaften als wesentliches Element der Informationsverarbeitung. Aufgrund konzeptueller Schwächen dieser Theorie (siehe z.B. Engel et al., 1992) wurde vorgeschlagen, daß Gruppen (Assemblies) von Neuronen und deren zeitlich fein koordinierte Feueraktivität zur Informationsverarbeitung beitragen (v.d. Malsburg, 1981; Abeles, 1982b).

Fragestellung

In dieser Arbeit wird eine Methode entwickelt und diskutiert, mit deren Hilfe man Daten, die bei der experimentellen Untersuchung der zeitlich abgestimmten Aktivität von Nervenzellen anfallen, hinsichtlich gleichzeitiger und fast-gleichzeitiger Feueraktivität analysieren kann. Aus parallelen diskreten binären Prozessen stammende Daten sollen auf korrelierte Feueraktivität hin untersucht werden, um überzufällig häufig auftretende gleichzeitige Aktivität zu detektieren. Ziel ist dabei außerdem, festzustellen, welchen Untergruppen der beobachteten Prozesse eine erkannte Korrelation zuzuordnen ist. So fallen durch zufällige Ko-aktivierung von einzelnen Neuronen und/oder echt korrelierten Gruppen von Neuronen zufällige Koinzidenzen höherer Ordnung an ("Scheinkorrelationen"), die nicht als "echte" Korrelationen erkannt werden sollen.

Modellbeschreibung

Im ersten Teil der Arbeit wird ein Modell (Modell of Independent Interaction Processes, "MIIP") vorgestellt, mit dessen Hilfe die Analyse von "echten" Korrelationen, die sich durch überzufällig häufige gleichzeitige Feueraktivität ausdrücken und nicht - wie die Scheinkorrelationen - auf

Korrelationen niedrigerer Ordnung zurückführbar sind, ermöglicht wird. Es enthält Prozesse O_i, $i = 1, ..., n$, die die beobachtete Feueraktivität der n Neurone beschreiben sollen und Superpositionen von unabhängigen Basisprozessen darstellen. Alle Basisprozesse B_M, $\emptyset \neq M \subseteq \{1, ..., n\}$ sind stationäre Bernoulli Prozesse mit eigenen Feuerwahrscheinlichkeiten $\lambda_M \in [0, 1)$, wobei jeder nicht-leeren Teilmenge M der Menge N der n Neurone genau ein Basisprozeß zugeordnet wird, der pro Zeitpunkt t mit Wahrscheinlichkeit λ_M feuert und dabei zur gleichen Zeit t Aktionspotentiale in allen beobachtbaren Prozessen O_i mit $i \in M$ erzeugt. Formal:

$$B_M(t) = 1_M \cdot Z_M(t) \qquad \text{mit } Z_M(t) \text{ Bernoulli}(\lambda_M),$$

alle B_M unabhängig, und

$$O_i(t) := \sup\{B_M(t) | \emptyset \neq M \subseteq \{1, ..., n\}, i \in M\} \qquad i = 1, ..., n.$$

Die Basisprozesse mit $|M| = 1$ werden als unabhängige Hintergrundprozesse interpretiert, wohingegen alle anderen Basisprozesse zu Korrelationen zwischen beobachtbaren Prozessen führen.

ML-Schätzung und Asymptotische Varianz

Mit Hilfe des MIIP kann man sehr direkt und anschaulich zwischen beobachteten Korrelationen, die sich in den Prozessen O_i ausdrücken, und "echten" Korrelationen, die in den Basisprozessen mit $\lambda_M > 0$ erzeugt werden, unterscheiden. Mit Hilfe der Maximum-Likelihood (ML) Methode werden die Feuerwahrscheinlichkeiten der Basisprozesse aus den relativen Häufigkeiten der 2^n beobachtbaren Feuerkonfigurationen pro Zeitpunkt geschätzt, da die ML-Schätzer einer Multinomialverteilung genau die relativen Häufigkeiten der entsprechenden Ereignisse sind. Eine Formel für die Maximum-Likelihood Schätzer im MIIP für n Neurone wird entwickelt und bewiesen, mit deren Hilfe man unter Verwendung der multidimensionalen δ-Methode die asymptotische Normalität der ML-Schätzer für alle λ_M zeigen und die asymptotische Varianz ausrechnen kann, was für zwei und drei Neurone bei der Korrelation der jeweils höchsten Ordnung exemplarisch durchgeführt wird.

Empirische und Asymptotische Eigenschaften des Tests

Die Teststatistik des Quotienten aus dem ML-Schätzer und seiner asymptotischen Varianz wird in Bezug auf den empirischen und asymptotischen Fehler erster und zweiter Ordnung für typische Parameterbereiche diskutiert. In Simulationsstudien konnte gezeigt werden, daß der Test konservativ ist, d.h. daß für die untersuchten Parameterbereiche der asymptotische Signifikanzlevel unterschritten wird. Des weiteren werden in analytischen Überlegungen prinzipielle Zusammenhänge zwischen den beteiligten Variablen demonstriert, wie z.B. das Ansteigen der Testmacht mit der Länge des untersuchten Datenstücks, mit fallenden Hintergrundraten oder mit steigender Feuerwahrscheinlichkeit des Basisprozesses der Korrelation der höchsten Ordnung, jeweils bei festgehaltenen verbleibenden Parametern. Zudem hängt die Testmacht für konstante verbleibende Parameter im wesentlichen von dem Verhältnis aus "echten" Koinzidenzen und Zufallskoinzidenzen ab.

Erweitertes Modell für "Unscharfe" Koinzidenzereignisse

Modellbeschreibung

Experimentelle Daten deuten darauf hin, daß koinzidentes Feuern mit einer gewissen Unschärfe einhergeht (Abeles et al., 1993; Riehle et al., 1997). Im zweiten Teil der Arbeit wird daher das MIIP erweitert, um die Analyse von fast-gleichzeitiger Aktivität bei der Suche nach "echten" Korrelationen in parallelen Prozessen zu ermöglichen. Ein weiterer Parameter $J \in \mathbb{N}_0$ bestimmt den maximalen Abstand (in Zeitschritten), den zwei Aktionspotentiale verschiedener Zellen haben dürfen, um noch als koinzident interpretiert zu werden.

Für eine m-elementige Teilmenge M von Neuronen ($m > 1$) aus N wird der Begriff der Konfiguration eingeführt: Eine Konfiguration von M unter der Verzögerung ("Jitter") $j \in \{1, ..., J\}$ beschreibt genau eine Möglichkeit, eine "gejitterte" Koinzidenz der Neurone aus M mit Jitter j zu erzeugen, in der alle Neurone aus M in einem Fenster der Länge j feuern, wobei das erste und das letzte Aktionspotential genau j Zeitschritte voneinander entfernt sind. Für jede Teilmenge M und jede Verzögerung $j \in \{1, ..., J\}$ wird pro Konfiguration von M unter j je ein zusätzlicher stationärer Bernoulli Basisprozeß eingeführt, der Aktionspotentiale in den O_i mit $i \in M$ in der durch die Konfiguration bestimmten Reihenfolge erzeugt. Für jeden

zusätzlichen Jitter-Basisprozeß existiert ein weiterer Parameter, der die Wahrscheinlichkeit der entsprechenden Konfiguration festlegt.

Parameterschätzung

Mit Hilfe der Momentenmethode werden Formeln zur Berechnung der Parameter für symmetrisches und asymmetrisches Jitter bei zwei Neuronen und für $n = 3$, $J = 1$ und symmetrisches Jitter gezeigt. Da die Unabhängigkeit zwischen den Zeitschritten im erweiterten Modell nicht mehr gegeben ist, werden anhand exemplarischer Rechnungen verschiedene Schätzungen für die beteiligten Wahrscheinlichkeiten evaluiert. In den diskutierten Fällen besitzt die Schätzung, die viele, abhängige Intervalle verwendet, gegenüber derjenigen mit unabhängigen Intervallen geringerer Anzahl eine reduzierte Varianz. Des weiteren hängt die Schätzung der Hintergrundfeuerwahrscheinlichkeit von J ab. Da dies im Experiment nicht bekannt ist, werden in einer Simulationsstudie für verschiedene Parameter und unterschiedliche J verschiedene Schätzungen evaluiert. Die Ergebnisse legen nahe, das größte plausible J zu verwenden.

Eigenschaften des Tests

Da die durch das Jitter J injizierten Abhängigkeiten der zur Schätzung verwendeten Intervalle von endlicher Reichweite sind, besitzen auch die Schätzer für die Parameter im erweiterten Modell asymptotische Normalverteilung. Anhand eines eingeschränkten Modells für zwei Neurone, in dem die Wahrscheinlichkeit jeder Konfiguration gleich hoch ist, werden exemplarisch Signifikanzlevel und Testmacht der vorgeschlagenen Teststatistik untersucht. Diese wird gebildet aus dem Quotienten der Summe der geschätzten Koinzidenzwahrscheinlichkeiten über alle Jitter j, $j = 1, ..., a$, wobei a das angenommene J ist, und deren empirisch ermittelter Standardabweichung. Der Signifikanzlevel liegt für alle untersuchten a sehr nahe an dem vorgegebenen Wert. Die Testmacht ist für $a = J$ maximal für konstantes J. Eine Wahl von $a > J$ führt zu einer geringfügigen Reduktion der Testmacht, wohingegen $a < J$ dieselbe stark reduziert.

Ausblick

Die durchgeführten Analysen und Simulationen weisen darauf hin, daß das MIIP gut geeignet ist, um die Struktur der Korrelationen innerhalb einer Gruppe von Neuronen zu untersuchen. Die Maximum-Likelihood Schätzer der Modellparameter sind trotz der Komplexität des Modells anschaulich und besitzen kurze Formeln. Dagegen hat das Modell einige Einschränkungen, so u.a. die mangelnde Analysierbarkeit systematischer Desynchronisation der Prozesse und die exponentielle Zunahme der Anzahl der zu schätzenden Parameter mit der Anzahl der Neurone. Des weiteren ist das MIIP zur Analyse von stationären Prozessen entwickelt worden. Um auf nichtstationäre Daten angewandt werden zu können, müssen weitere Untersuchungen durchgeführt werden, die eine Erweiterung des Ansatzes ermöglichen. Zur Anwendung des Modells in der experimentellen Praxis sollten somit weitere Studien durchgeführt werden, um abweichendes Verhalten der Teststatistik unter weniger idealen Bedingungen einschätzen und aufgrund dieser Erkenntnisse die Ergebnisse der Datenanalyse kritisch beurteilen zu können.

Contents

Frequently Used Symbols

Name	Description
n	number of observed neurons
N	set $N = \{1, 2, ..., n\}$ of observed neurons
T	length of observed data piece (in bins)
t	$\in \{1, ..., T\}$ time index
$\lambda_1,\ \lambda_2,\ \lambda_3$	background firing probabilities
λ_b	background firing probability under $\lambda_1 = \lambda_2 = \lambda_3$
$\lambda_{12}, \lambda_{13}, \lambda_{23}$	pairwise coincidence firing probabilities
λ_c	pairwise coincidence firing probabilities under $\lambda_{12} = \lambda_{13} = \lambda_{23}$
λ_{123}	firing probability of the triplet process
O_1, O_2, O_3	observable processes
z^*	threshold value for the test, leading to α
α	significance level
$1 - \beta$	test power
p^c_{11}	probability of a chance doublet in the MIIP with $n = 2$
p^c_{111}	probability of a chance triplet in the MIIP with $n = 3$
NO	probability of a non-overlapping chance triplet under $H_0 :\ \lambda_{123} = 0$
O	probability of an overlapping chance triplet under $H_0 :\ \lambda_{123} = 0$
MIIP	Model of Independent Interaction Processes, Exact Coincidences
E-MIIP	Extended Model of Independent Interaction Processes

1 Introduction

During the last century, the amount of research in the field of neuroscience has grown enormously in the struggle to reveal the mechanisms underlying the immense variety of the brain's capabilities. Despite this fact, we are far from understanding the principles of information processing in the cortex.

The human neocortex contains about 10^{14} neurons, which are thought to be the basic elements of information processing. Communication between neurons is carried out by short, transient electrical signals ("spikes") of about one millisecond duration, which are transmitted across synapses - the physical connections between nerve cells - and can be measured with a microelectrode.

In a simplified view, the flow of information is unidirectional for a single nerve cell: the dendrites propagate their synapses' electrical input to the soma of the cell, where all inputs are integrated. If this sum reaches threshold, a spike is evoked, which then travels along the axon to the synapses connecting the cell with other neurons (for a detailed description see e.g. Kandel, Schwartz and Jessell, 1996, chapter 2).

The human neocortex is highly interconnected. One neuron receives about 20000 inputs from other cells. These connections are far from being organized in a unidirectional way. There are feed-forward as well as feedback and intrinsic connections (DeYoe and von Essen, 1988).

Regarding anatomical properties, there is relatively large agreement among neuroscientists. Yet, concerning the way information is coded by the electrical signals, two different hypotheses have been proposed, which could as well be complementary.

1.1 Rate Coding and the Single Neuron Doctrine

The first hypothesis made use of the known fact that neurons 'respond' to special stimuli by a change in their firing rate. Based on this observation, Barlow (1972, 1992a, b, c, d) formulated the "single neuron doctrine" that regards single neurons as the main building blocks of information processing. Every neuron responds optimally to a single feature or a combination of features, and an enhancement of its firing rate signals the presence of the corresponding stimulus. Simple neurons situated at the sensory periphery converge onto more complex neurons

with highly specific response properties. “Cardinal cells” at the highest level were claimed to respond to the appearance of entire objects.

A cell’s firing response is usually measured with a Peri-Stimulus-Time Histogram (PSTH). Across several trials, time is divided into short intervals to measure the neuron’s mean firing frequency and to detect the time of its enhancement.

Consistent with Barlow’s approach, specialization can be found on different levels in the brain: On a global level, localized lesions of the brain cause special behavioral defects, a fact which led to the conceptual division of the brain into different regions which are thought to be essential for vision, audition, motor control or speech. A more detailed view shows that single brain areas responsible for vision or motor control are organized into cortical maps, which means e.g. that objects next to each other in the visual field are represented by adjacent areas in the visual cortex (for a detailed description see e.g. Kandel et al., 1996, chapter 1). And finally, on the micro-level, each segment that represents a small area in the visual field contains various specific cells that respond to special stimulus properties such as orientations of moving light bars or colors (e.g. Hubel and Wiesel, 1962).

During the last two decades, more and more experimental evidence and theoretical considerations evoked doubt that the described mechanism is the only way information is processed in the cortex.

First, if the single neuron doctrine were true, the robustness of the signal against the loss of single neurons should be expected to be much smaller than is now known. Also the speed of information processing would be much slower than measured in psychophysical experiments (Thorpe, Fize and Marlot, 1996), if several hierarchical layers of neurons needed to integrate their firing input over time. As Singer, Engel, Kreiter, Munk, Neuenschwander and Roelfsema (1997) conclude, “decisions must be reached on the basis of the first few spikes that are sent by the preceding processing stage (because) maximally it takes a few tens of milliseconds per processing stage to perform the computations necessary for the analysis and recognition of patterns”.

Apart from that there is a combinatorial problem. "The possible combinations that confront the visual system are virtually unlimited" (Singer and Gray, 1995). Thus, the requirement regarding the immense number of cardinal cells representing complex percepts can never be fulfilled. If we needed one cardinal cell for any special grouping of objects in every possi-

ble combination and from any possible perspective, not only for visual but also for auditory and sensory objects and their combinations, the existent variety and adaptiveness of perception could not be accomplished.

Moreover, for any new object or new combination of objects, spare cells needed to be reserved. On the other hand, cells that are extremely specialized could be idle for decades, which would reduce the efficiency of information processing (see e.g. Engel, König, Kreiter, Schillen and Singer, 1992).

And finally, bottleneck problems would arise (see e.g. Singer, 1999): Just as any visual scene at first evokes the activation of thousands of neurons, the same number is necessary for the planning and execution of a movement. It is hardly conceivable that a single cardinal cell that finally represents the percept of the visual scene is able to coordinate that complex response.

1.2 Assembly Coding and Synchrony

To accomplish for these and other restrictions, groups of neurons ("assemblies") have been proposed to represent objects with their coordinated spiking activity. Hebb (1949) was the first to put forward the idea that information is to be found in the coherent activity within functional groups of neurons.

From an anatomical point of view, the immense interconnectivity mentioned at the beginning conforms with a high degree of divergence and convergence. Braitenberg and Schüz (1991) concluded from anatomical and mathematical considerations that "no neuron is farther than two synapses away from any other neuron ... (which implies that) ... any sufficiently large portion of the cortex is informed about the activity in the rest of the cortex".

Assembly coding would allow individual cells to participate at different times in the representation of different patterns and thus reduce the number of cells required as well as allow for a greater flexibility in the generation of new representations (Singer et al., 1995). It would as well be more robust against the loss of single neurons, because the joint assembly activity remains nearly unchanged by the lack of one neuron.

Taking into consideration that assembly coding is used by the cortex, another question arises: How can a given subset of cells that represents a percept be identified by the cortex? In the literature, this is commonly referred to as the "binding problem" (Singer et al., 1995).

The required enhancement of the signal's saliency could be accomplished by an enhancement of the firing rate. This would give rise to ambiguities, because the discharge rates of feature-selective cells which vary as a function of the match between stimulus and receptive field properties would not be distinguishable from the modulations signaling the relatedness of responses (Singer et al., 1997).

Thus, an additional property of the signal has been proposed to solve the binding problem, namely synchronization of individual discharges of neurons of the same assembly (v. d. Malsburg, 1981; Abeles, 1982b). Parallel recordings of different neurons were introduced to allow for the study of synchronicity in the millisecond range.

The next paragraphs will be a summary of evidence as well as theoretical considerations indicating that exact spike timing is important for cortical information processing. This presentation claims by no means to be complete but should instead give the reader an overview over this work's experimental background.

Already Hebb (1949) suggested that coincident firing was important for the dynamical formation of assemblies. Support for Hebb's rule - in its short version to be read as "cells that fire together wire together" - could be found recently by Makram, Lübke, Frotscher and Sakmann (1997), who showed that synaptic efficacy can be regulated by - roughly speaking - coincident activity of both pre- and postsynaptic cell. To continue with evidence on the cellular level, Mainen and Sejnowski (1995) showed that under certain circumstances, neurons are capable of precise and stable timing of the moment of their firing, which is a necessary condition for the coordination of joint spiking activity. In a theoretical work, Abeles (1982a) showed that a neuron is more sensitive to a few synchronous excitatory inputs than to the same number of inputs arriving in an asynchronous random manner, which led to the notion of a neuron as a "coincidence detector" rather than an integrator over its input rate (Abeles, 1982a; König, Engel, Singer, 1996). As König et al. (1996) pointed out, coding by coincidence has multiple advantages. First, the processing time is fast, as only one event of synchronous input evokes a response. Secondly, the brain would be less susceptible to noise, because only noise that coincides is considered. Finally, by the additional use of another temporal property of the signal, information processing becomes resistant to amplitude fluctuations.

From a theoretical point of view, the binding problem can be solved. The temporal binding model "predicts that neurons that respond to the same sensory object might fire in temporal

synchrony with a precision in the millisecond range. However, no such synchronization should occur between cells that are activated by different objects in the sensory periphery" (Engel, Fries, Singer, 2001). This selective mechanism would allow the system to enable figure-ground segregation by establishing a distinct representational pattern for each object. It is important to mention that one single cell can take part in many different assemblies. This reduces redundancy and allows for adaptive combinations of neuronal groups to new percepts. It has been found that assemblies can form dynamically, depending on the context and nature of the computational task (Vaadia, Haalman, Abeles, Bergman, Prut, Slovin, Aertsen, 1995; Riehle, Grün, Diesmann, Aertsen, 1997, Grün, Diesmann, Aertsen, 2002b).

Fries, Roelfsema, Engel, König and Singer (1997) showed that "upon dichoptic stimulation, neurons responding to the stimulus that continued to be perceived increased the synchronicity (...) of their (...) patterning, while the reverse was true for neurons responding to the stimulus that was no longer perceived", using a paradigm of inter-ocular rivalry (Fries, Schröder, Singer, Engel, 2001). This provided great support for the hypothesis that response synchronization could serve as a mechanism for perceptual grouping as pointed out in the last paragraph.

Finally, synchrony has been shown to be correlated to expectation of a stimulus and motivation (Riehle et al., 1997). During a delayed-pointing task the recorded neurons in the primary motor cortex of macaque monkeys showed spike synchronization without modulation of discharge rates in relation to internal events such as expectation. Coincidence was found to provide information beyond that expected by simple rate changes and independence of neuronal firing.

1.3 Mathematical Analysis Methods

Now that was pointed out why synchrony is thought to be important for information processing in the cortex, a short description will follow regarding some analytical methods used to interprete the recorded data. Of course, there is a wide variety of methods, however those will be discussed that are of direct relevance for the work here.

As mentioned above, a microelectrode can measure the voltage changes of the electrical potential of a nerve cell. The times of spiking events are detected by defining a threshold: Whenever the potential crosses threshold from below, a spike is detected. An upper bound avoids the detection of artefacts whose fluctuations considerably exceed those of spikes. The

times of spiking events provide the raw data that underly many of the techniques used. They are often represented as "spike trains", i.e. time series of point processes, displayed as vertical bars at the time of spike occurrences.

With the goal to detect fine temporal relationships between the firing of different cells, it became necessary to use parallel recordings. Thus, the data that need to be analyzed are represented by parallel point processes. Moreover, the used data are often averaged over many trials in order to improve the statistics.

In a first attempt to describe the data with a view to temporal coordination of spike-timing, cross-correlation histograms (CCH, Perkel, Gerstein, Moore, 1967) are computed. They can deal with the data of two parallel processes and are based on the assumption that the firing rates of the recorded cells do not change during the observed period. A CCH provides information about precise spike-timing of the two cells in question. For two spike trains $z_1(t)$ and $z_2(t)$ of length T with

$$z_i(t) = \sum_{j=1}^{N_i} \delta_{t\, t_{ij}} \qquad i = 1, 2,$$

where $t \in \{1, ..., T\}$ and t_{ij}, $j = 1, ..., N_i$ indicate the times of spike occurrences in spike train i, the cross-correlation function can be written as (see e.g. Eggermont, 1990, p. 143):

$$R_{z_1 z_2}(\tau) = \frac{1}{T} \sum_{j=1}^{N_1} \sum_{k=1}^{N_2} \delta_{\tau\,(t_{1j} - t_{2k})}$$

This means that per time shift τ, the empirical frequency of cell 1 to fire τ time units before cell 2 is plotted. Theoretical considerations about the relation between the shape of the cross-correlogram and the connectivity between the cells can be found in Aertsen and Gerstein (1985) and in Gerstein and Aertsen (1985). In Epping and Eggermont (1987) and Eggermont (1990), theoretical and experimental findings are integrated. Detailed discussions of experimental findings about cross-correlation analysis can be found in Toyama, Kimura and Tanaka (1981a, b), Nelson, Salin, Munk, Arzi and Bullier (1992), Munk, Nowak, Nelson and Bullier (1995), and Nowak, Munk, Nelson, James and Bullier (1995). Usually, centered peaks are observed, which means that the participating cells have a tendency to fire in synchrony, i.e. without a delay. Still, the peaks have a certain width which implies that synchronous firing does not occur without a

small imprecision of a few milliseconds (Nowak et al., 1995; Grün, Diesmann, Grammont, Riehle, Aertsen, 1999).

As the cross-correlogram can only be applied onto pairs of neurons, the Unitary-Event (UE) method (Grün, 1996; Grün, Diesmann, Aertsen, 2002a, b) was developed that could deal with virtually any number of neurons. The method uses multiple-trial data from parallel recordings and is based on the assumption that the times of spiking events of one single neuron can be described as a stationary Poisson process and on the null-hypothesis of full independence between all recorded neurons. Rate fluctuations that are the same over all trials can be treated by applying the method in sliding-window fashion. Time is first discretized and thus the spike trains are assumed to be describable as Bernoulli processes. The occurrence of at least one spike in a time-unit ("bin") is coded as a one, whereas the counter-event is coded as a zero, such that every bin is coded as a binary vector of length n, where n denotes the number of neurons. Whenever at least two neurons fire a spike in a designated bin, this event is called a coincidence. In the UE method, all $2^n - n - 1$ binary vectors describing coincidences are analyzed in parallel to find those subgroups of neurons which emit significantly more coincidences than expected under full independence.

In a given period of stationary firing rates, one can compare the expected number of coincidences with its observed number for all subgroups. The expected number can easily be derived in the following way:

On the basis of the number of spikes $S_{i,v}$, $i = 1, ..., n$, $v = 1, ..., V$ per neuron i and trial v in a time segment of length T (in bins), the firing probabilities p_i of the neurons are estimated over all trials as $\widehat{p_i} := \frac{\sum_v S_{i,v}}{VT}$. Thus the expected number n_{exp} of coincidences with spikes in a subset of neurons $M \subseteq N$, $N = \{1, 2, ..., n\}$ and non-spikes in $N \setminus M$ is $n_{exp} = TV \prod_{i \in M} \widehat{p_i} \prod_{j \in N-M} (1 - \widehat{p_j})$, because the null-hypothesis assumes independence between all processes. n_{exp} parameterizes the Poisson distribution used to identify those coincidences whose number exceeds the expected number more than up to a pre-defined significance level. They are called "Unitary Events".

This method can be used to treat relatively short pieces of data and many neurons. It has been used to identify dynamical changes in assembly formation (Riehle et al., 1997). The method was extended to allow for the analysis of near-coincidences (Grün et al., 1999).

1.4 Statement of the Problem

In the UE-method, detection of “Unitary Events”, i.e. the occurrence of significant excess coincidences, implies existence of correlations between the processes. Due to the null-hypothesis of full independence of the processes, deviation from expectation due to a correlation within a subgroup of neurons is not identified. Thus, if coincidences of a group of more than two neurons are found to occur more often than expected by full independence, one cannot clarify the following question:

“Is a correlation between all neurons of the group necessary to explain the coincidences, or is the existence of lower-order correlations sufficient to produce the observed amount of events?”

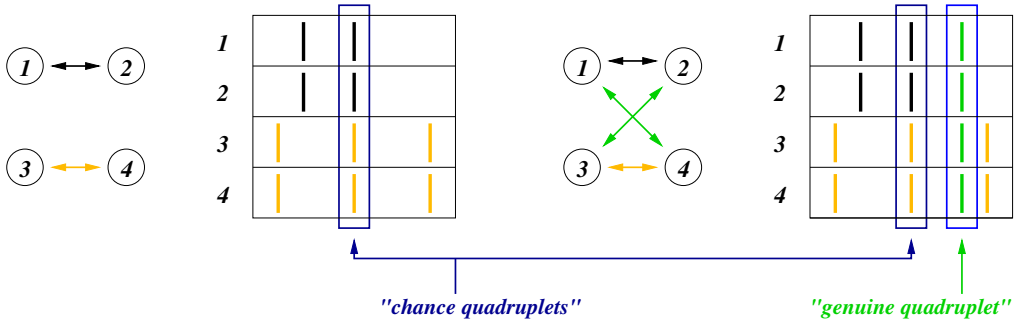

Figure 1: Visualization of the difference between genuine and chance coincidences.

The work presented here focuses on the identification of those (sub-)groups of neurons participating in “genuine” higher-order correlations, defined here as those coincidences that cannot be explained by any kind of chance co-activated firing of subgroups or single neurons. Figure 1 demonstrates the difference between genuine and chance coincidences. The correlations (arrows) between the neurons (numbers in circles) show up in coincident firing of spikes (vertical bars). On the left, a chance co-activation of two genuine pairwise correlations produces a chance quadruplet. This can as well happen on the right side. However, a genuine quadruplet correlation is added, which produces genuine coincidences of order four. This work provides methods for deciding which one of all 2^{2^n-n-1} possible combinations of genuine correlations is underlying the joint spiking activity of the observed neurons. Here, every subgroup of neurons showing a genuine correlation is interpreted as an assembly.

The model used in this work to describe spiking activity of groups of neurons should meet the following constraints based on assumptions about information processing in the cortex:

1. Following the hypothesis that coincident activity expresses activation of cell assemblies, the model should provide methods to distinguish between independence and correlation of the observed neurons on the basis of fine temporal structuring. It will thus contain processes for independent background activity as well as processes describing correlations between the neurons.

2. All $2^n - n - 1$ subgroup-correlations are allowed to exist in the observed time period. Thus, assemblies can as well overlap or include one another. This is consistent with the hypothesis that single neurons can take part in different assemblies to allow for the recombination to different but overlapping representations.

3. Considering the hypothesis that different assemblies activated by distinct parts of the same percept can be distinguished by synchronization among cells of the same group and desynchronization between the groups, the processes representing correlations between different assemblies must be independent. This implies that the co-activation of two different assemblies can only occur on a chance level.

Finally, the model should be able to deal with data from single trials. Therefore, everything is done on a single-trial basis. If stationarity can be assured over the trials, the use of the data from all trials leads to an increase of the length of the data piece and thus to an improvement of the statistics.

In the first part, a model for the analysis of exact coincidences is developed that can treat any subgroup correlation separately from the others. With the help of this model, which is formulated for any number of neurons, it is possible to estimate the probability of synchronous firing for each subset of neurons and to evaluate the significance of the number of coincidences based on the existence of subgroup correlations. The performance of the proposed analysis method is demonstrated for two and three neurons.

In the second part, the model is extended to allow for coincidences with a time-lag (“jitter”). Exemplary considerations will show the usefulness of the extended model and its applicability onto two and more neurons.

Part I

The Model for Exact Coincidences

2 A General Framework for the MIIP

The model on the basis of which the analysis methods are developed will be presented now. It conforms with all constraints made in the introduction and thus provides adequate means to analyze neuronal spike trains if the assumptions are met. The model will be named 'MIIP' (Model of Independent Interaction Processes). In part II, an extension of the model will be developed to allow for the analysis of less exact coincidences. This extended version will be called 'E-MIIP'.

Like in the original Unitary Event method (Grün, 1996), the observed individual spike trains will be assumed to be describable as stationary Poisson processes. Due to the discretization of time, these will always reduce to Bernoulli processes of length T (in bins of length b (in ms)), where every spike is coded as a 1 and every non-spike as a 0. To distinguish between independent and correlated firing, the observed activity of one single neuron is assumed to consist of an independent part, which will be referred to as "background", and a part that is correlated with groups of other neurons. Moreover, as it is necessary to decide which correlations exist, the correlated part needs to be split up into different processes representing the correlations with all the different subgroups of neurons.

In order to allow for all these different processes, one must distinguish between "observable" and "basic" processes. Each observable process O_i, $i = 1, ..., n$ which represents the firing activity of one observed neuron is assumed to be a superposition of several basic processes. As a whole:

- Let $N := \{1, ..., n\}$ be the set of n observed neurons,
- $1, ..., T$ the indices of the time steps in the observed period of time, $t \in \{1, ..., T\}$.
- For each of the $2^n - 1$ non-empty subsets of neurons $M \subseteq N$, a basic process $P_M(t)$ is introduced.

- All P_M are assumed to be stationary Bernoulli processes with firing probabilities $\lambda_M \in [0, 1)$, and independent.
- Those P_M with $|M| = 1$ represent the independent background processes. Per time step, they produce a spike in that same O_i with $\{i\} = M$ with probability λ_M. From now on they will be named B_M ("background").
- All other P_M represent correlation processes. A success at time t produces one spike in each observable process $O_i(t)$ with $i \in M$ at time t. These basic correlation processes will be called C_M ("correlation").

Thus, to observe the event $\{O_i(t) = 1\}$ - "a spike in O_i is observed at time t", either $\{B_i(t) = 1\}$ must occur, or some of the $\{C_M(t) = 1\}$ with $i \in M$, or both. Every C_M is the origin of the genuine correlation between the subgroup of neurons that is given by M. A genuine correlation between all members of M is said to exist if and only if $\lambda_M > 0$. It can lead to a chance correlation that is observed between the members of another subgroup M' with $M \subset M'$. The estimation of all parameters λ_M will help to extract the genuine correlations. In subsection 4.2, a general formula for the maximum-likelihood estimates of all λ_M in the MIIP will be proven.

3 Two Neurons

3.1 The Model and Its Estimates

For $N = \{1, 2\}$, there are only three non-empty subsets of N: $\{1\}$, $\{2\}$ and $\{1, 2\}$. Two of them lead to background processes, and the third produces coincidences between the two observed neurons. Figure 2 shows the effect this superposition has on both observed processes. For simplicity, in the examples, the processes P_M and their firing probabilities λ_M, where $M = \{x_1, x_2, ..., x_K\} \subseteq N$ are denoted shortly as $P_{x_1,x_2,...,x_K}$ and $\lambda_{x_1,x_2,...,x_K}$, respectively.

The two observed processes are correlated if and only if the firing probability λ_{12} of the process C_{12} is positive. The question of higher- and lower-order correlations is not approached in the two-neurons-case because only one correlation can exist at all, and its absence is equivalent to independence of the processes.

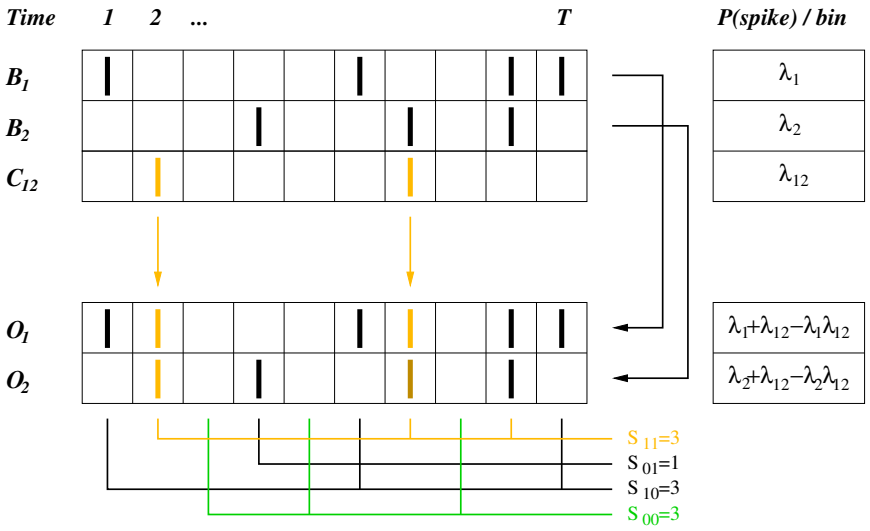

Figure 2: MIIP for two neurons: events from the background processes are copied onto the corresponding observed processes, whereas events in the correlation process are copied onto both observed processes to produce genuine coincidences.

One approach to estimate the firing probabilities is the method of moments: For all fixed $t \in \{1, ..., T\}$

$$p_{00} := P(\{O_1(t) = 0, O_2(t) = 0\}) = (1 - \lambda_1)(1 - \lambda_2)(1 - \lambda_{12}) \quad (1)$$

$$p_{0+} := P(\{O_1(t) = 0\}) = (1 - \lambda_1)(1 - \lambda_{12}) \quad (2)$$

$$p_{+0} := P(\{O_2(t) = 0\}) = (1 - \lambda_2)(1 - \lambda_{12}), \quad (3)$$

which implies

$$(1 - \lambda_1) = \frac{p_{00}}{p_{+0}} \quad (4)$$

$$(1 - \lambda_2) = \frac{p_{00}}{p_{0+}} \quad (5)$$

$$(1 - \lambda_{12}) = \frac{p_{0+}p_{+0}}{p_{00}}. \quad (6)$$

Lemma 1

The maximum-likelihood estimates of the probabilities p_{00}, p_{0+}, p_{+0} are the relative frequencies of the corresponding events.

Proof

As this is also needed in later parts of this work, it will directly be shown for all numbers of neurons n.

The joint observed process

$$(O(t))_{t=1,...,T} := \begin{pmatrix} O_1(t) \\ \vdots \\ O_n(t) \end{pmatrix}_{t=1,...,T}$$

has values in $\{0,1\}^n$ for every $t \in \{1,...,T\}$. As every single $(O_i(t))_{t=1,...,T}$ is a superposition of stationary and independent Bernoulli processes, every bin is independent from every other bin, and the probabilities

$$p_M(t) := P(\{O_i(t) = 0 \,\forall i \in M,\, O_j(t) = 1 \,\forall j \in N - M\})$$

are constant for $t \in \{1,...,T\}$ and will thus be written as p_M. Thus, the random vector

$$X := ((X_M)_{M \subseteq N})$$

is multinomial$(T, p_{M_{M \subseteq N}})$ for $X_M := \sum_{t=1}^{T} (1_{\{O_i(t)=0 \,\forall i \in M\}} \cdot 1_{\{O_j(t)=1 \,\forall j \in N-M\}})$.

It remains to be shown that for any multinomial$(T, p_1, ..., p_k)$ distribution with outcomes $X_1, ..., X_T \in \{o_1, ..., o_k\}$, the maximum-likelihood estimates for the p_i, $i = 1, ..., k$ are the relative frequencies $f_i := \frac{n_i}{T}$ with $n_i := \sum_{j=1}^{T} 1_{\{X_j = o_i\}}$.

The probability of getting the sequence $(x_1, ..., x_T)$ of outcomes is given by:

$$P_{(p_1,...,p_k)}(X_1 = x_1, ..., X_T = x_T) = p_1^{n_1} p_2^{n_2} \cdot ... \cdot p_k^{n_k}$$

leading to the log-likelihood

$$L_{(p_1,...,p_k)}(X_1 = x_1, ..., X_n = x_n) = n_1 \log p_1 + ... + n_k \log p_k$$

Maximization with the constraint $\sum_{i=1}^{k} p_i = 1$ and a Lagrange multiplier λ gives the equations

$$\frac{\partial L}{\partial p_i}(\widehat{p}_i) - \lambda \frac{\partial}{\partial p_i} \left(\sum_{i=1}^{k} p_i \right) = 1, \quad i = 1, ..., k$$

$$\iff \frac{n_i}{\widehat{p}_i} - \lambda = 0 \Leftrightarrow \widehat{p}_i = \frac{n_i}{\lambda}, \quad i = 1, ..., k$$

With

$$1 = \sum_{i=1}^{k} \widehat{p}_i = \frac{1}{\lambda} \sum_{i=1}^{k} n_i = \frac{T}{\lambda}$$

follows $\lambda = T$ and thus $\widehat{p}_i = \frac{n_i}{T} = f_i$. □

Corollary 1

With

$$S_{ij} := \sum_{t=1}^{T} 1_{\{O_1(t)=i, O_2(t)=j\}}, \quad i, j \in \{0, 1\} \tag{7}$$

$$S_{0+} := \sum_{t=1}^{T} 1_{\{O_1(t)=0\}} \tag{8}$$

$$S_{+0} := \sum_{t=1}^{T} 1_{\{O_2(t)=0\}}, \tag{9}$$

the maximum-likelihood estimates of the probabilities p_{00}, p_{0+}, p_{+0} are

$$\widehat{p}_{00} = \frac{S_{00}}{T}; \quad \widehat{p}_{0+} = \frac{S_{0+}}{T}; \quad \widehat{p}_{+0} = \frac{S_{+0}}{T}. \tag{10}$$

Indeed, note that

$$\begin{aligned} T &= S_{00} + S_{10} + S_{01} + S_{11} \\ &= S_{+0} + S_{01} + S_{11} \\ &= S_{0+} + S_{10} + S_{11}, \end{aligned}$$

and the probability p_{0+} (and in analogy p_{+0}) corresponds to the union of the disjoint events $\{O_1 = 0, O_2 = 0\} \cup \{O_1 = 0, O_2 = 1\}$. Thus, the probability of this latter event is the sum of the single probabilities.
A direct consequence of (10) together with (4)-(6) is

$$(1 - \widehat{\lambda}_1) = \frac{S_{00}}{S_{+0}} \tag{11}$$

$$(1 - \widehat{\lambda}_2) = \frac{S_{00}}{S_{0+}} \tag{12}$$

$$(1 - \widehat{\lambda}_{12}) = \frac{S_{0+}S_{+0}}{S_{00}T} \tag{13}$$

$$\widehat{\lambda}_1 = \frac{S_{10}}{S_{+0}} \tag{14}$$

$$\widehat{\lambda}_2 = \frac{S_{01}}{S_{0+}} \tag{15}$$

$$\widehat{\lambda}_{12} = \frac{S_{00}(S_{00} + S_{01} + S_{10} + S_{11}) - (S_{00} + S_{01})(S_{00} + S_{10})}{S_{00}T} \tag{16}$$

$$= \frac{S_{00}S_{11} - S_{01}S_{10}}{S_{00}T} \tag{17}$$

3.2 Developing a Test for the Hypothesis H_0: $\lambda_{12} = 0$

After having computed the maximum-likelihood estimates, the original question concerning the existence of correlations can be approached. The hypothesis $H_0 : \lambda_{12} = 0$ needs to be tested against its alternative $H_1 : \lambda_{12} > 0$. The question that needs to be answered now is the following: what is the probability that $\widehat{\lambda}_{12}$ has at least the same amount as found in the data, given the model holds true and $\lambda_{12} = 0$? To cope with that problem, we construct a test: Define an interval I such that (error of first order) the probability that $\widehat{\lambda}_{12}$ falls outside of I,

given $\lambda_{12} = 0$ does not exceed a given level α, i.e.

$$P(\widehat{\lambda}_{12} \notin I \mid H_0) \leq \alpha. \tag{18}$$

With the pre-defined significance level, one can examine the test power, i.e. the probability that $\widehat{\lambda}_{12}$ falls outside of I, given $\lambda_{12} > 0$, i.e. $P(\widehat{\lambda}_{12} \notin I \mid H_1) =: 1 - \beta$.

In the described model, the estimates have asymptotic normal distribution (see subsection 3.2.1). With the asymptotic variance, a test with the asymptotic property (18) can be developed.

3.2.1 Asymptotic Normality and Variance of $\widehat{\lambda}_{12}$

To show the asymptotic normality of the maximum-likelihood estimates and to compute their asymptotic variance, the multivariate δ-method will be used (see e.g. Bishop, Fienberg, Holland, 1991, pp. 486-502). It is a generalized version of the one-dimensional delta-method, and based on a Taylor-expansion up to the second term around the components' expected values.

Proposition 1 (Multidimensional δ-method)

Let $\widehat{\theta}_T$ be a d-dimensional random vector: $\widehat{\theta}_T = (\widehat{\theta}_{T,1}, ..., \widehat{\theta}_{T,d})$,
let θ be a d-dimensional vector parameter: $\theta = (\theta_1, ..., \theta_d)$.
Let further $\widehat{\theta}_T$ have asymptotic normal distribution in the sense that

$$\mathcal{L}\left[\sqrt{T}(\widehat{\theta}_T - \theta)\right] \xrightarrow[T\to\infty]{} \mathcal{N}(0, \Sigma(\theta)),$$

where $\Sigma(\theta)$ is the $d \times d$ asymptotic covariance matrix of $\widehat{\theta}_T$. This implies that for large T, $\widehat{\theta}_T$ has approximately $\mathcal{N}(\theta,\, T^{-1}\Sigma(\theta))$ distribution.
Let further $f : O \to \mathbb{R}$ be defined on an open subset $O \subseteq \mathbb{R}$.
Let f have a differential at θ, i.e. f has the following expansion as $x \to \theta$:

$$f(x) = f(\theta) + \sum_{j=1}^{d} (x_j - \theta_j) \left.\frac{\partial f}{\partial x_j}\right|_{x=\theta} + o(||x - \theta||),$$

or in matrix notation:

$$f(x) = f(\theta) + (x - \theta) \left(\frac{\partial f}{\partial \theta} \right)^t + o(||x - \theta||).$$

Then the asymptotic distribution of $f(\widehat{\theta}_T)$ is given by

$$\mathcal{L}\left[\sqrt{T}(f(\widehat{\theta}_T) - f(\theta)) \right] \xrightarrow[T \to \infty]{} \mathcal{N}\left(0, \left(\tfrac{\partial f}{\partial \theta} \right) \Sigma(\theta) \left(\tfrac{\partial f}{\partial \theta} \right)^t \right).$$

The asymptotic variance of $f(\widehat{\theta}_T)$ is

$$\frac{1}{T} \sum_{k,k'=1}^{d} \sigma_{kk'}(\theta) \left(\frac{\partial f}{\partial \theta_k} \right) \left(\frac{\partial f}{\partial \theta_{k'}} \right)$$

Corollary 2

The maximum-likelihood estimate $\widehat{\lambda}_{12}$ of the parameter λ_{12} has asymptotic normal distribution with asymptotic variance

$$\sigma^2_{\widehat{\lambda}_{12}} = \frac{(1 - \lambda_{12})(\lambda_{12}(1 - \lambda_1)(1 - \lambda_2) + \lambda_1 \lambda_2)}{T(1 - \lambda_1)(1 - \lambda_2)} \tag{19}$$

Proof of Corollary 2

This will be shown by application of the multidimensional δ-method: Take $d = 3$ and

$$\begin{aligned} \widehat{\theta}_T &:= (\widehat{\theta}_{T,1}, \widehat{\theta}_{T,2}, \widehat{\theta}_{T,3}) := \left(\left(\frac{S_{00}}{T} \right)_T, \left(\frac{S_{+0}}{T} \right)_T, \left(\frac{S_{0+}}{T} \right)_T \right) \\ \theta &:= (\theta_1, \theta_2, \theta_3) := (p_{00}, p_{+0}, p_{0+}). \end{aligned}$$

All S_{ij}, $i, j \in \{0, 1\}$ are binomially distributed random variables with parameters p_{ij} and T. Thus, as T tends to infinity, all $\widehat{p}_{ij} := \frac{S_{ij}}{T}$ have asymptotic normal distribution with mean p_{ij} and variance $\sigma^2_{ij} = \frac{p_{ij} \cdot (1 - p_{ij})}{T}$ (local limit theorem for the binomial distribution, e.g. Krengel (1991, p. 80)). One can thus approximate every $\frac{S_{ij}}{T}$ by a normally distributed variable with mean p_{ij} and variance σ^2_{ij}, i.e. $\frac{S_{ij}}{T} \sim p_{ij} + \sigma_{ij} Z_{ij}$ for some standard normal random variable Z_{ij}. This yields

$$\mathcal{L}\left[\widehat{\theta}_T\right] \xrightarrow[T\to\infty]{} \mathcal{N}\left(\theta, T^{-1}\Sigma(\theta)\right),$$

With

$$\Sigma(\theta) = \begin{pmatrix} p_{00}(1-p_{00}) & p_{00}(1-p_{+0}) & p_{00}(1-p_{0+}) \\ p_{00}(1-p_{+0}) & p_{+0}(1-p_{+0}) & p_{00}-p_{+0}p_{0+} \\ p_{00}(1-p_{0+}) & p_{00}-p_{+0}p_{0+} & p_{0+}(1-p_{0+}) \end{pmatrix}$$

The formulas for the covariances in $\Sigma(\theta)$ can be derived as follows:

$$\begin{aligned} & \text{Cov}\left(\frac{S_{0+}}{T}, \frac{S_{00}}{T}\right) && (20) \\ = \; & \frac{1}{T^2}\left[E\left(\sum_{t_1=1}^{T} 1_{\{O_1(t_1)=0\}} \sum_{t_2=1}^{T} 1_{\{O_1(t_2)=O_2(t_2)=0\}}\right) - p_{0+}p_{00}T^2\right] && (21) \\ = \; & \frac{1}{T^2}\left[\sum_{t_1=1}^{T} E(1_{\{O_1(t_1)=O_2(t_1)=0\}}) + \sum_{t_1,t_2=1;\; t_1\neq t_2}^{T} E(1_{\{O_1(t_1)=O_2(t_2)=0\}}) - p_{0+}p_{00}T^2\right] && (22) \\ = \; & \frac{1}{T^2}\left[Tp_{00} + T(T-1)p_{00}p_{0+} - T^2 p_{0+}p_{00}\right] = \frac{p_{00}(1-p_{0+})}{T} && (23) \end{aligned}$$

In an analogous way one gets

$$\text{Cov}\left(\frac{S_{+0}}{T}, \frac{S_{00}}{T}\right) = \frac{p_{00}(1-p_{+0})}{T} \quad \text{and} \quad \text{Cov}\left(\frac{S_{0+}}{T}, \frac{S_{+0}}{T}\right) = \frac{p_{00}-p_{0+}p_{+0}}{T}.$$

Take now $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$ with $f(x) := f(x_1, x_2, x_3) := \frac{x_2 x_3}{x_1}$. f has a differential at θ for $p_{00} > 0$ and $T > 0$. Note that $f(\theta) = (1 - \lambda_{12})$. This f was used in order to simplify the computation, because λ_{12} has normal distribution iff $1 - \lambda_{12}$ has normal distribution, and $\text{Var}(1 - \lambda_{12}) = \text{Var}(\lambda_{12})$.

$\Longrightarrow f(\widehat{\theta}_T)$ has asymptotic normal distribution with asymptotic variance

$$\text{Var}(f(\widehat{\theta}_T)) \xrightarrow[T\to\infty]{} \frac{1}{T}\sum_{k,k'=1}^{3} \sigma_{kk'}(\theta)\left(\frac{\partial f}{\partial \theta_k}\right)\left(\frac{\partial f}{\partial \theta_{k'}}\right)$$

$$
\begin{aligned}
&= \frac{p_{00}(1-p_{00})p_{0+}^2p_{+0}^2}{Tp_{00}^4} - 2\frac{p_{00}(1-p_{+0})p_{0+}^2p_{+0}}{Tp_{00}^3} - 2\frac{p_{00}(1-p_{0+})p_{0+}p_{+0}^2}{Tp_{00}^3} \\
&\quad + \frac{p_{+0}(1-p_{+0})p_{0+}^2}{Tp_{00}^2} + 2\frac{(p_{00}-p_{0+}p_{+0})p_{0+}p_{+0}}{Tp_{00}^2} + \frac{p_{0+}(1-p_{0+})p_{+0}^2}{Tp_{00}^2} \\
&= \frac{p_{0+}p_{+0}}{Tp_{00}^2}\left[-p_{0+}p_{+0} - p_{+0} - p_{0+} + 2p_{00} + \frac{p_{+0}p_{0+}}{p_{00}}\right] \\
&= \frac{(1-\lambda_{12})}{T(1-\lambda_1)(1-\lambda_2)}\left[-1+\lambda_1+\lambda_2+2(1-\lambda_1)(1-\lambda_2)-(1-\lambda_1)(1-\lambda_2)(1-\lambda_{12})\right] \\
&= \frac{(1-\lambda_{12})(\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2)}{T(1-\lambda_1)(1-\lambda_2)}
\end{aligned}
$$

□

Thus, the asymptotic variance of $\widehat{\lambda}_{12}$ is

$$\sigma^2_{\widehat{\lambda}_{12}} \doteq \frac{(1-\lambda_{12})(\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2)}{(1-\lambda_1)(1-\lambda_2)} \cdot \frac{1}{T} \tag{24}$$

or if $\lambda_{12} = 0$:

$$\sigma^2_{\widehat{\lambda}_{12}|\lambda_{12}=0} \doteq \frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)} \cdot \frac{1}{T} \tag{25}$$

Unfortunately, the exact value of $\sigma^2_{\widehat{\lambda}_{12}}$ cannot be derived from the data. Just like the firing probabilities, it needs to be estimated. Still, as T tends to infinity, $\widehat{\sigma}_{\widehat{\lambda}_{12}} \to \sigma_{\widehat{\lambda}_{12}}$, and consequently $\frac{\widehat{\lambda}_{12}}{\widehat{\sigma}_{\widehat{\lambda}_{12}}} \sim \frac{\lambda_{12}+\sigma_{\widehat{\lambda}_{12}}Z}{\sigma_{\widehat{\lambda}_{12}}}$ for a standard normal distributed random variable Z. Hence:

$$P\left(\frac{\widehat{\lambda}_{12}}{\widehat{\sigma}_{\widehat{\lambda}_{12}}} > z\right) \doteq P\left(\frac{\lambda_{12}+\sigma_{\widehat{\lambda}_{12}}Z}{\sigma_{\widehat{\lambda}_{12}}} > z\right) = P(Z > z) = 1 - \Phi(z) \tag{26}$$

where

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx. \tag{27}$$

for a standard normal distributed random variable Z and $\lambda_{12} = 0$.

3.2.2 Test Power for $\sigma_{\widehat{\lambda}_{12}|\lambda_{12}=0}$ and $\sigma_{\widehat{\lambda}_{12}}$

From what was just stated, it follows that for the test, the probability to find $\frac{\widehat{\lambda}_{12}}{\widehat{\sigma}_{\widehat{\lambda}_{12}}}$ outside the interval I :=(-∞, z^*), given $\lambda_{12} = 0$, is for large T asymptotically $1 - \Phi(z^*)$ (for example for $z^* = 1.96$, we have a significance level of 2.5%)

This is true when using either of both equations (24) or (25), because they are equal under H_0. Thus, the asymptotic significance level is the same when using equation (24) or (25). However, for both methods, one needs to know the test power, i.e. the probability to find a test statistics $\frac{\widehat{\lambda}_{12}}{\widehat{\sigma}_{\widehat{\lambda}_{12}}}$ outside the interval (-∞, z^*), given $\lambda_{12} > 0$.

Notation and asymptotics derived for the significance estimation in equation (26) can be used:

$$P\left(\frac{\widehat{\lambda}_{12}}{\widehat{\sigma}_{\widehat{\lambda}_{12}}} > z^*\right) \doteq P\left(\frac{\lambda_{12} + \sigma_{\widehat{\lambda}_{12}} Z}{\sigma_{\widehat{\lambda}_{12}}} > z^*\right) = P\left(\frac{\lambda_{12}}{\sigma_{\widehat{\lambda}_{12}}} + Z > z^*\right)$$

which using $\sigma_{\widehat{\lambda}_{12}|\lambda_{12}=0}$ is the following

$$= P\left(Z > z^* - \frac{\lambda_{12}\sqrt{T}\sqrt{(1-\lambda_1)}\sqrt{(1-\lambda_2)}}{\sqrt{\lambda_1\lambda_2}}\right) = \text{power}_{\sigma_{\widehat{\lambda}_{12}|\lambda_{12}=0}} \tag{28}$$

and using $\sigma_{\widehat{\lambda}_{12}}$ the following

$$= P\left(Z > z - \frac{\lambda_{12}\sqrt{T}\sqrt{(1-\lambda_1)}\sqrt{(1-\lambda_2)}}{\sqrt{(1-\lambda_{12})}\sqrt{\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}\right) = \text{power}_{\sigma_{\widehat{\lambda}_{12}}} \tag{29}$$

This shows that using $\sigma_{\widehat{\lambda}_{12}|\lambda_{12}=0}$, one is (at least asymptotically) able to detect $\lambda_{12} > 0$ with a higher probability than with $\sigma_{\widehat{\lambda}_{12}}$.

Indeed for $\lambda_{12} > 0$:

$$\frac{\lambda_{12}\sqrt{T}\sqrt{(1-\lambda_1)}\sqrt{(1-\lambda_2)}}{\sqrt{\lambda_1\lambda_2}} \overset{!}{>} \frac{\lambda_{12}\sqrt{T}\sqrt{(1-\lambda_1)}\sqrt{(1-\lambda_2)}}{\sqrt{(1-\lambda_{12})}\sqrt{\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}$$

$$\begin{aligned}
&\iff (1-\lambda_{12})(\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2) > \lambda_1\lambda_2 \\
&\iff \lambda_{12}(1-\lambda_1)(1-\lambda_2) - \lambda_{12}^2(1-\lambda_1)(1-\lambda_2) - \lambda_{12}\lambda_1\lambda_2 > 0 \\
&\iff (1-\lambda_1-\lambda_2) - \lambda_{12}(1-\lambda_1)(1-\lambda_2) > 0
\end{aligned}$$

This is true as the following parameter ranges are assumed: $\lambda_{12} \leq 0.01$ and $0 < \lambda_1, \lambda_2 \leq 0.2$ with $b = 1$ ms, leading to

$$(1-\lambda_1-\lambda_2) - \lambda_{12}(1-\lambda_1)(1-\lambda_2) > 0.6 - 0.01 > 0.$$

The firing probabilities can directly be transformed into mean firing rates $r_M(\text{in Hz}) = \frac{\lambda_M}{b\,(\text{in sec})}$, leading to a range of $0 < r_1, r_2 \leq 200$ Hz and $0 \leq r_{12} \leq 10$ Hz, which is thought to be the physiologically relevant range of frequencies (e.g. White, Chow, Ritt, Soto-Treviño and Kopell, 1998; Grün et al., 1999).

The above result is also described in Bishop et al. (1991, p. 499), namely that using the variance of $\widehat{\lambda}_{12}$ assuming $\lambda_{12} = 0$ increases the power of the test. It would thus seem reasonable to only use $\sigma_{\widehat{\lambda}_{12}|\lambda_{12}=0}$. Still, as we will see, this leads to a bad performance of the test for short pieces of data. In the following subsections, it is thus proposed to use the full formula (24) to make sure that the pre-defined significance level is obtained for small T, too.

3.3 Performance of the Proposed Test

In the following subsection the properties of the proposed test will be discussed. Mainly two aspects must be considered:

1. Which constraints need to be imposed onto the data to ensure that the asymptotic method can be applied? This has been examined in simulation studies and will be discussed in subsection 3.3.1.

2. Given that the asymptotics are applicable, how good is the performance of the method? This question can be approached on a theoretical basis. Problems like the required number of time steps to reach a certain test power or the dependency of the required value of λ_{12} on the background firing probabilities are discussed in subsection 3.3.2.

3.3.1 Applicability to Finite Strings of Data

As was mentioned in the introduction, experimental spiking data are subject to non-stationarity over time. The Unitary Event analysis copes with rate-changes over time that are constant over all trials by application of a sliding-window analysis. The method introduced in this work however concentrates on the analysis of single-trial data that can become available in studies where the investigator cannot by any external measures identify repetitions of certain events (e.g. Schmidt, Grün, Singer and Galuske, 2001). When dealing with single-trial data, questions concerning the required length of the data piece arise. They will be dealt with in this subsection. On the other hand, the method that is presented here can as well deal with multiple-trial data when stationarity over time and trials can be granted in the inspected time period. Then, depending on the length of the trial, the answers provided in this subsection can give an idea concerning the number of trials that need to be accomplished.

We will now discuss the question of the minimal size of T necessary to reach the predefined significance level and test power. As one spike has a length of about one millisecond, the time resolution b is often chosen to be 1 ms. With background rates smaller than 200 Hz (e.g. White et al., 1998), coincidence rates up to 10 Hz and a time resolution of $b = 1$ ms, parameter values $\lambda_1, \lambda_2 < 0.2$ and $\lambda_{12} < 0.01$ need to be examined. This subsection will also serve to compare the two test statistics

$$Z := \frac{\widehat{\lambda}_{12}}{\widehat{\sigma}_{\widehat{\lambda}_{12}}} \tag{30}$$

and

$$Z_0 := \frac{\widehat{\lambda}_{12}}{\widehat{\sigma}_{\widehat{\lambda}_{12}|\lambda=0}} \tag{31}$$

that use different estimations of the variance of $\widehat{\lambda}_{12}$, the latter leading to a higher test power. It will be shown that among the two test statistics Z and Z_0, only the use of Z can be considered reasonable for small pieces of data.

3.3.1.1 Significance Level

In the figures 3 and 4, the pre-given asymptotic significance level ($\alpha = 0.025$, solid line) is compared with the empirical significance level derived from simulations. This is done for

different lengths T (corresponding to 0.1, 0.5, 1, 5, 10 and 50 seconds, $b = 1$ ms). Figure 3 uses Z as test statistics, whereas in figure 4, Z_0 is applied.

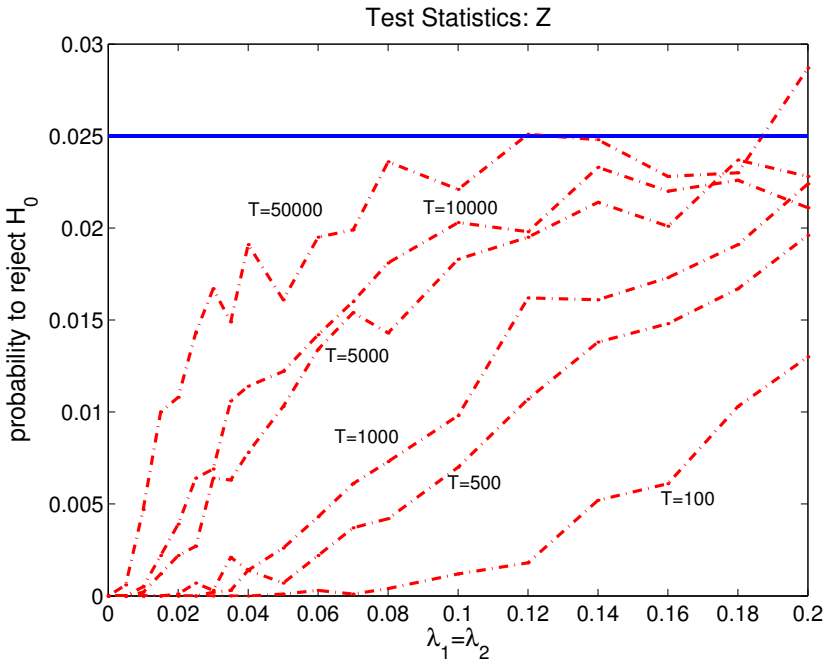

Figure 3: Comparison of empirical and asymptotic significance level for different T (100, 500, 1000, 5000, 10000, 50000). On the abscissa, the background firing probabilities λ_1, λ_2, which are chosen to be equal, range from 0 to 0.2. $\lambda_{12} = 0$. The solid line represents the pre-given asymptotic significance level α. For the data points of the dash-dotted curves, the two processes $B_1 = O_1,\ B_2 = O_2$ with firing probabilities $\lambda_1 = \lambda_2$ were simulated 10000 times. The relative number of experiments with $Z > 1.96$ is plotted (see equation (30)).

The data points of the dash-dotted curves represent the empirical percentage of experiments with $Z > 1.96$ and $Z_0 > 1.96$, respectively. For every T and every parameter set $\lambda_1 = \lambda_2 \in \{0, 0.005, 0.01, ..., 0.04, 0.05, 0.06, ..., 0.08, 0.1, 0.12, ..., 0.2\}$, 10000 datasets consisting of the processes O_1 and O_2 were generated. For the derivation of the significance level by definition the correlation process C_{12} is absent ($\lambda_{12} = 0$), so that in this case $O_1 = B_1$ and $O_2 = B_2$. The parameters λ_1 and λ_2 were chosen to be equal in order to reduce the parameter space. In subsection 3.3.2.2, inequality of both background firing probabilities will be discussed and shown to reduce to - roughly speaking - the product of the background rates, such that for these

considerations the limitation $\lambda_1 = \lambda_2$ does not represent a strong restriction of the parameters.

Figure 3 demonstrates that for the given parameters and the test statistics Z, nearly all empirical significance values stay below the required significance level of 0.025. This shows that under $H_0 : \lambda_{12} = 0$, the probability to reject the null-hypothesis is even smaller than required, leading to a conservative test.

In contrast, figure 4 shows that - except for $\lambda_1 = \lambda_2 = 0$ where $\widehat{\lambda}_{12} \equiv 0$ and thus the empirical significance level is 0, too - all empirical significance values are higher than the predefined $\alpha = 0.025$. Due to the asymptotic properties, the empirical curves approximate the value of α with growing T, still as long as background rates and T together do not produce "enough" coincidences to obtain an approximately normal distribution of counts, the required significance level is exceeded up to more than 400% of its value.

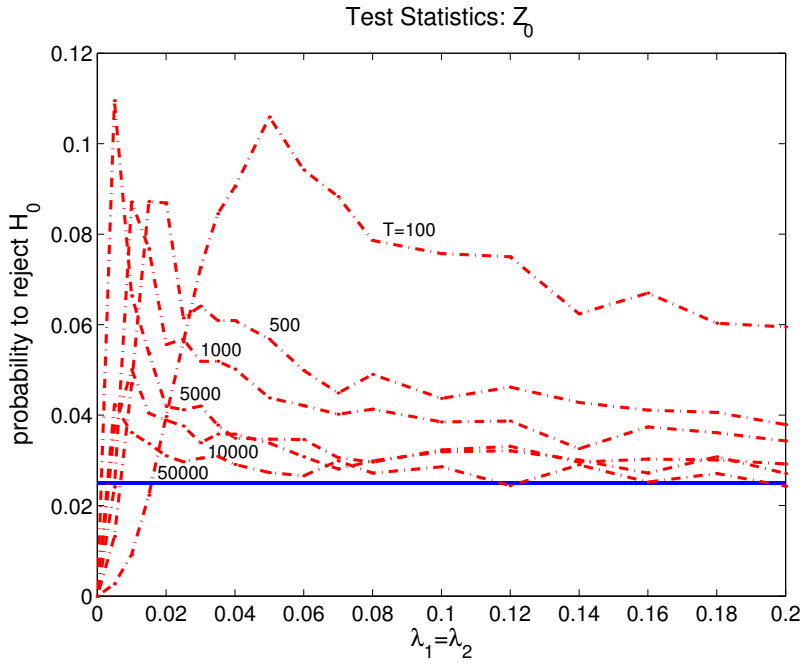

Figure 4: Comparison of empirical (dash-dotted) and asymptotic (solid) significance level for different lengths T of data. Used test statistics: Z_0 (equation (31)). All parameters and simulations as in figure 3.

To summarize the above results:

- The asymptotics are not applicable for small T and small background rates. This is true for the use of both test statistics. Although one might think that small background rates improve the clarity because less chance coincidences "disturb" the detection of genuine coincidences, another fact proves to be more important: All counts - especially the coincidence counts S_{11}, of which there are in general very few - need to be approximately normally distributed. As Sachs (1971) pointed out, for $T \cdot \lambda \cdot (1-\lambda) \geq 9$, one can use the normal approximation of the binomial distribution with parameters λ and T. In table 1, the minimal background firing probability (given $\lambda_1 = \lambda_2$) that fulfills this condition is shown for every T that was used in the simulations.
  The graph of $\lambda(1-\lambda)$ between 0 and 1 is a $\bigcap$-shaped curve with maximum $\frac{1}{4}$ at $\lambda = \frac{1}{2}$ and two minima with value zero at the borders $\lambda \in \{0, 1\}$. Thus, using the p-q-formula, one gets $\lambda = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{9}{T}}$, considering only the cases where $\lambda < 0.5$. The probability of a chance coincidence under H_0 is the product of the background rates $\frac{p^c_{11}}{(1-\lambda_{12})} = \lambda_1\lambda_2 = \lambda_1^2$ due to the independence of the basic background processes. One can observe the correspondence between table 1 and the figures 3 and 4. For $T = 50000$, already a very low background firing probability leads to an effective significance level close to $\alpha = 2.5\%$. In contrast, for $T = 100$, even for the very high background firing probability of 0.2, the effective significance level stays above even $2 \cdot \alpha$ for Z_0 (figure 4) and around $\frac{\alpha}{2}$ for Z (figure 3).

T	minimal $\lambda_1 = \lambda_2$ (rounded)	λ_1^2 (rounded)
100	0.3162	0.1
500	0.1354	0.0183
1000	0.0953	0.0091
5000	0.0425	0.0018
10000	0.0300	0.0009
50000	0.0134	0.0002

Table 1: Minimal background firing probability at $\lambda_{12} = 0$ to reach $T \cdot \lambda_{12} \cdot (1-\lambda_{12}) \geq 9$.

- In spite of the non-applicable asymptotics, the test using the test statistics Z stays conservative. It can thus be applied onto short pieces of data without risking to falsely rejecting

H_0 more often than intended. In contrast, the test using Z_0 does not meet the requirements for small pieces of data and small background rates and should thus be applied only for large T and higher background firing probabilities.

In the following subsection, the test power of both tests will be discussed. The consideration of the test with test statistics Z_0 is only added for reasons of completeness, as its shortcomings concerning the significance level are considered to be grave, especially because in a usual experimental situation the analysis of short pieces of data is necessary due to non-stationary firing rates, as mentioned above.

3.3.1.2 Test Power

In the figures 5 (test statistics: Z, equation (30) on page 26) and 6 (test statistics: Z_0, equation (31) on page 26), empirical (dash-dotted) and asymptotic (solid) test power are compared. Again for the empirical curves, the relative number of significant experiments out of 10000 is plotted. A relatively high coincidence firing probability of $\lambda_{12} = 0.004$ was chosen for the figures. Per simulation experiment, the independent Bernoulli processes B_1, B_2 and C_{12} were simulated for T time steps with firing probabilities $\lambda_1, \lambda_2,$ and λ_{12} and merged onto the observable processes O_1 and O_2, from which the test statistics was computed and compared with $z^* = 1.96$.

1. The first remarkable fact is that - as described in subsection 3.2.2 - the empirical as well as the asymptotic test power is considerably larger for Z_0 than for Z, especially for small pieces of data and small background rates. This corresponds to the fact that in those parameter ranges, the error of first order is very high for Z_0. Moreover, the asymptotic curves in figure 6 starts at a test power of nearly 100% for background rates close to zero (the test power is not defined for the extreme case $\lambda_1 = \lambda_2 = 0$). This does not seem plausible, especially for small pieces of data and small coincidence rates. Accordingly, the empirical curve belonging to $T = 100$ deviates strongly from the corresponding asymptotic solid curve, especially for small background rates. This is not the fact for the test statistics Z, although for small T the asymptotics do not fit well either.

2. Still, not surprisingly, with growing T the asymptotic and the empirical curve approach more and more for both test statistics.

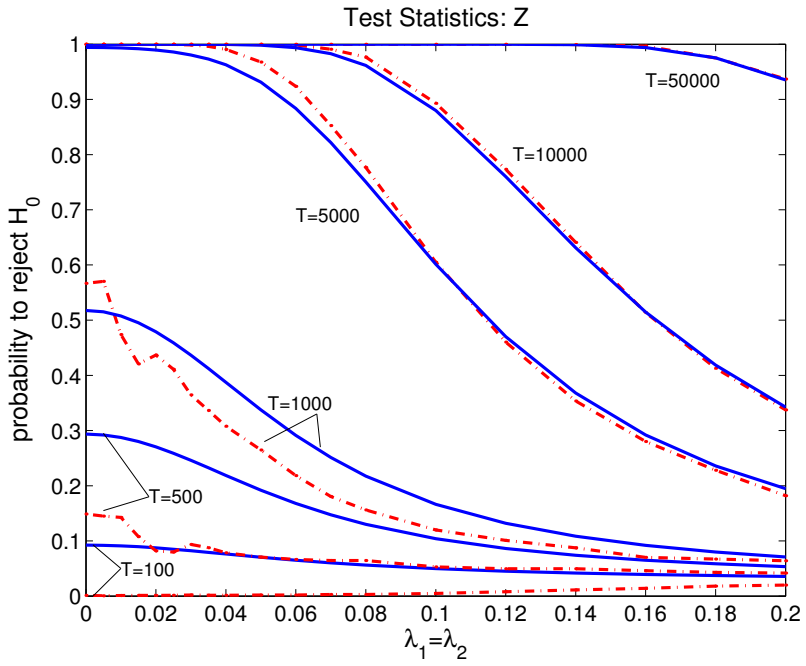

Figure 5: Comparison of empirical (dash-dotted) and asymptotic (solid) test power for T =100, 500, 1000, 5000, 10000, 50000 and $\lambda_{12} = 0.004$. λ_1 and λ_2 like in figures 3 and 4. Solid: The empirical test power is derived as the relative number of experiments with $Z > 1.96$. Per parameter set, the processes O_1 and O_2 of length T were generated 10000 times.

3. Two main factors influence the test power:

- The relation between the number of chance coincidences and the number of genuine coincidences plays a role. The more excess coincidences relative to the randomly expected coincidences, the higher the probability to reject H_0. For example for $T = 5000$, $\lambda_1 = \lambda_2 = 0.06$ and $\lambda_{12} = 0.004$, there are $0.0036 \cdot 5000 = 18$ chance coincidences to be expected, plus $0.004 \cdot 5000 = 20$ genuine coincidences. As one would expect, an excess of "20 more than 18" can easily be detected (test power of about 0.9 for Z and nearly 1 for Z_0).

- Consider the case $T = 500$ with the same firing probabilities as in the previous paragraph. Then 1.8 chance coincidences need to be compared to 2 genuine coincidences. As it is very likely to get an "excess" of more than one coincidence just by chance, two genuine

coincidences are likely to be left undetected. Thus, a high quotient of $\frac{\lambda_{12}}{\lambda_c}$ is needed, plus a relatively large number of random coincidences to reduce random fluctuations. Similar results can be found in Grün et al. (2002a) and in Roy, Steinmetz and Niebur (2000), who discuss the relation between the firing rate and the number of excess coincidences necessary for the detection of Unitary Events in the UE method.

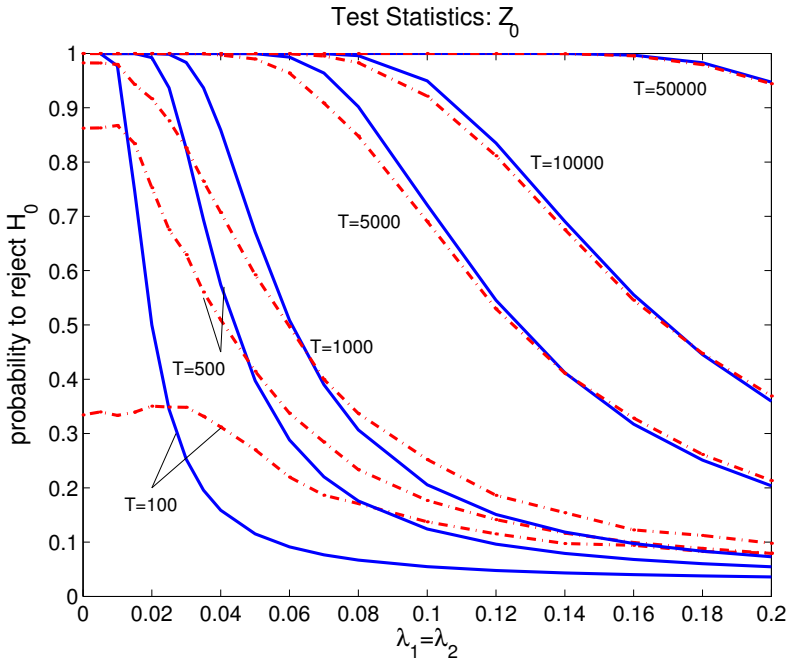

Figure 6: Comparison of empirical (dash-dotted) and asymptotic (solid) test power. Simulations, and parameters like in figure 5. Used test statistics for the empirical curves: Z_0.

3.3.2 Discussion of the Asymptotic Test Power

In this subsection, the asymptotic properties of the test will be discussed. One needs to keep in mind that the real properties may deviate from the results, especially for small T. Still, one can learn about general interdependencies between the parameters.

3.3.2.1 Relations between Background, Coincidence Rate and Test Power

Due to the high number of parameters T, λ_1, λ_1, λ_{12}, z^* and the test power, the number of dimen-

sions needs to be reduced when visualizing interdependencies. Hence in this first subsection, the background rates will be set equal, and the test power and z^* will be reduced to some typical values. When necessary, also λ_{12} and T will take on only some characteristic values. Z (equation (30) on page 26) will be the used test statistics, meaning that the full formula (24) on page 23 is used for $\sigma_{\hat{\lambda}_{12}}$.

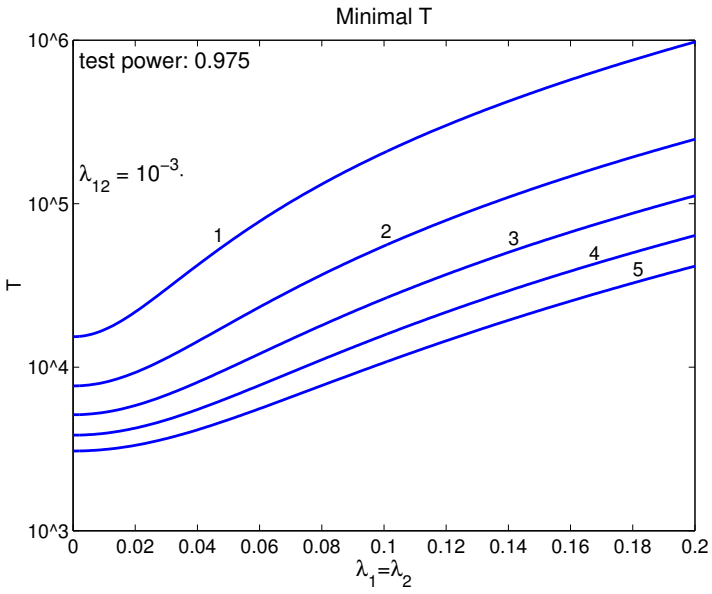

Figure 7: Required length T of data, depending on equal background firing probabilities $\lambda_1 = \lambda_2 \in (0, 0.2]$ and different $\lambda_{12} \in \{0.001, 0.002, ..., 0.005\}$ to reach an asymptotic test power of 0.975 at $z^* = 1.96$.

In figure 7, the required number of time steps to reach an asymptotic test power of 97.5% at $z^* = 1.96$ is plotted depending on the background firing probabilities for different λ_{12}.
Transforming equation (29) on page 24:

$$P(Z > z^* - \frac{\lambda_{12}\sqrt{T}\sqrt{(1-\lambda_1)(1-\lambda_2)}}{\sqrt{(1-\lambda_{12})(\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2)}}) = \text{power}_{\sigma_{\hat{\lambda}_{12}}} := 0.975 \qquad (32)$$

leads to

$$1.96 - \frac{\lambda_{12}\sqrt{T_{\min}}\sqrt{(1-\lambda_1)(1-\lambda_2)}}{\sqrt{(1-\lambda_{12})(\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2)}} = -1.96$$

and hence

$$\begin{aligned} T_{\min} &= \frac{4 \cdot 1.96^2 \cdot (1-\lambda_{12})(\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2)}{\lambda_{12}^2 \cdot (1-\lambda_1) \cdot (1-\lambda_2)} \\ &= \frac{4 \cdot 1.96^2 \cdot (1-\lambda_{12})(\lambda_{12}(1-\lambda_1)^2+\lambda_1^2)}{\lambda_{12}^2 \cdot (1-\lambda_1)^2}. \end{aligned}$$

Not surprisingly, the required number of time steps increases with growing background firing probabilities and falling λ_{12}. The smaller the relation between genuine coincidences and chance coincidences, the more time steps are needed to detect the genuine coincidences.

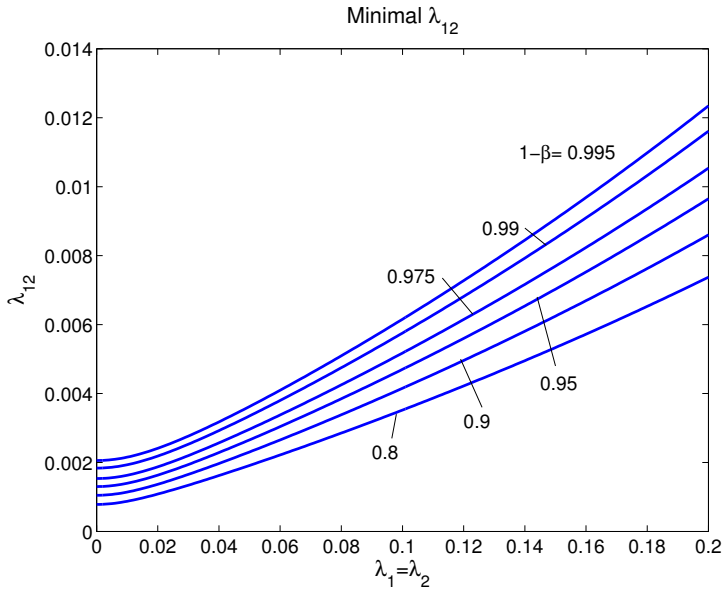

Figure 8: Required minimal λ_{12} at $T = 10000$, $\alpha = 0.025$ ($z^* = 1.96$) and $1-\beta \in \{0.8, 0.9, 0.95, 0.975, 0.99, 0.995\}$, depending on equal background rates $\lambda_1 = \lambda_2$ (x-axis).

In figure 8, the number of time steps is left constant at $T = 10000$. For the fixed significance level $\alpha = 0.025$ ($z^* = 1.96$), the minimal required genuine coincidence firing probability is plotted at six different values for the test power $1-\beta = 1-\Phi(z')$ for

$z' \in \{0.84, 1.82, 1.65, 1.96, 2.33, 2.58\}$. The corresponding formula is derived from

$$\begin{aligned} P(Z > z^* - \frac{\lambda_{12}\sqrt{T}\sqrt{(1-\lambda_1)(1-\lambda_2)}}{\sqrt{(1-\lambda_{12})(\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2)}}) &= 1-\Phi(z') \\ \iff \quad -1.96 + \frac{\lambda_{12}\sqrt{T}\sqrt{(1-\lambda_1)(1-\lambda_2)}}{\sqrt{(1-\lambda_{12})(\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2)}} &= z' \end{aligned}$$

by solving for λ_{12} and setting $\lambda_1 = \lambda_2$:

$$\begin{aligned} \lambda_{12} &= \frac{(1.96+z')^2(1-2\lambda_1)}{2(T+(1.96+z')^2)(1-\lambda_1)^2} \\ &\quad + \sqrt{\frac{(1.96+z')^4(1-2\lambda_1)^2 + 4\lambda_1^2(1.96+z')^2(T+(1.96+z')^2)(1-\lambda_1)^2}{4(T+(1.96+z')^2)^2(1-\lambda_1)^4}}, \end{aligned}$$

as $\lambda_{12} \geq 0$ and for $\lambda_1 = \lambda_2 > 0$

$$1 - 2\lambda_1 < \sqrt{(1-2\lambda_1)^2 + 4\lambda_1^2(1-\lambda_1)^2}.$$

One can see that the required λ_{12} needs to be larger for high background rates as well as for a higher test power, which shows the foregoing results from another view.

3.3.2.2 Different Background Firing Probabilities

In all previous considerations, the background firing probabilities were set equal for reasons of dimension reduction. However in the experimental situation, the background firing rates do not have to be equal. Therefore the situation for different background rates needs to be discussed. It seems to be plausible that the test power is highly connected to the relation between the probability of a chance coincidence on the one hand and the probability of a genuine coincidence on the other. As the probability of a chance coincidence, conditioned on the non-firing of the doublet process, is $\lambda_1\lambda_2$, the test power will be discussed for fixed background rate product $\lambda_1 \cdot \lambda_2$. For that reason, figure 9 shows the asymptotic test power depending on one background firing probability, with fixed $T = 10000$, $\lambda_{12} = 0.004$, $\alpha = 0.025$ and $\lambda_1 \cdot \lambda_2 = 0.05^2$. The

scaling is logarithmic to account for the symmetry of λ_1 and λ_2.

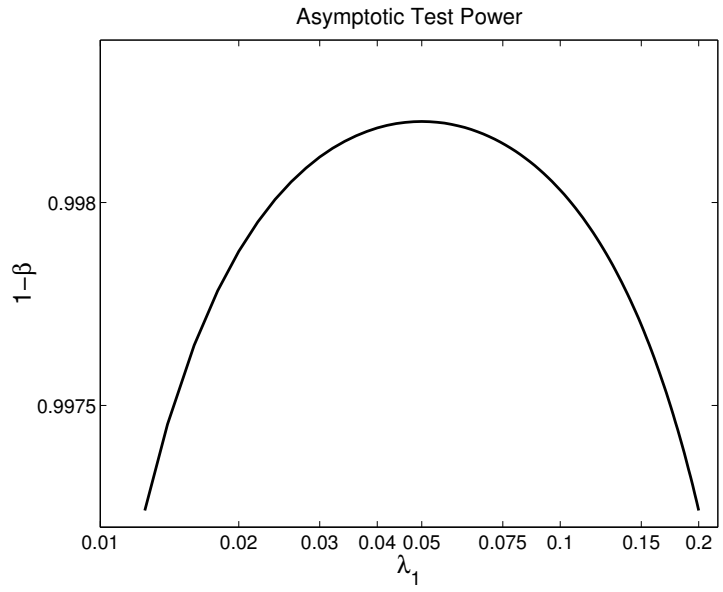

Figure 9: Asymptotic test power in relation to λ_1 for fixed background rate product $\lambda_1 \cdot \lambda_2 = 0.05^2$ at $T = 10000$, $\alpha = 0.025$ and $\lambda_{12} = 0.004$.

One can see that the test power is maximal for $\lambda_1 = \lambda_2$, namely at 0.05. This might seem a bit alerting since it says that in the previous subsections only the best case among all fixed $\lambda_1\lambda_2$ has been inspected. But the difference is relatively small. This is stressed in figure 10. Part (A) shows a three-dimensional graph of the minimal required λ_{12}, depending on both background firing probabilities. The solid curves are the level lines of this hillock at the same λ_{12} for a given test power of 0.975. The dashed curves represent those combinations of the background rates which have the same rate product as the level lines at the main diagonal. There, they require the same λ_{12} to reach the pre-defined test power. Further to the left and right, the curves for constant $\lambda_1\lambda_2$ deviate upwards and run towards levels of larger required λ_{12} to reach a test power of 0.975. Solid and dashed curves are shown as projections in figure 10 (B). Mainly the same result is to be seen in figure 9, showing that for $\lambda_1 = \lambda_2$ the test power is maximal for constant λ_{12} and $\lambda_1\lambda_2$ or that the minimal required λ_{12} is minimal for constant test power and constant $\lambda_1\lambda_2$. Still, it shows another important fact: The deviation of the curves with equal background rate product from their corresponding level lines with constant minimal λ_{12} is much

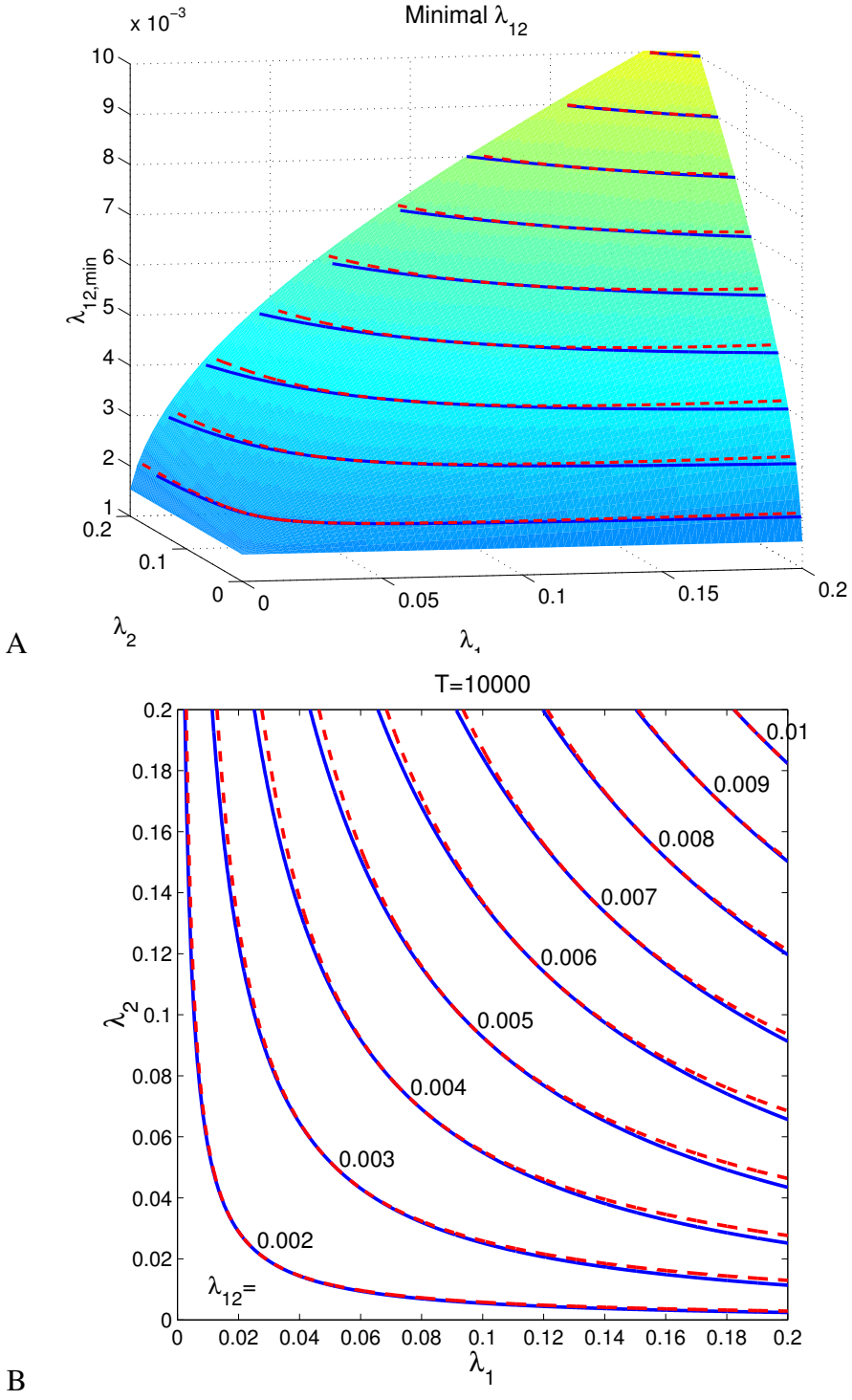

Figure 10: Comparison of the curves of equal background firing probability product (dashed) and curves of equal minimal required λ_{12} to reach a test power of 0.975 at $T = 10000$ (solid). The level lines (solid) are chosen such that $\lambda_{12} \cdot 1000 \in \mathbb{N}$. Every level line and its corresponding dashed curve intersect at the main diagonal. (A) three-dimensional plot (B) projections of the level lines and of the curves with constant $\lambda_1 \lambda_2$.

smaller than the distance to the other plotted curves even at the edges of the parameter range. The values of λ_{12} were chosen such that for a time resolution of $b = 1$ ms, they cover the whole range of integer frequencies in Hz corresponding to the plot. One can thus roughly say that for fixed lengths of data and fixed λ_{12}, the test power is essentially a function of the product of the background rates.

Taking another look at figure 7 on page 33, one sees another aspect: e.g. for $\lambda_{12} = 0.004$, the minimal required number of time steps is plotted depending on the background rates, assumed as equal. This represents the "movement" of the point of intersection of the level line for $\lambda_{12} = 0.004$ of figure 10 (B) with the main diagonal along the diagonal when T changes. To explain this in more detail: for large T, one can obviously detect $\lambda_{12} > 0$ more easily. On the other side, the larger T, the larger may the background firing probabilities be without disturbing the detection of $\lambda_{12} > 0$. So, for growing T, the level line for $\lambda_{12} = 0.004$ in figure 10 (B) (as all the others) moves upwards and to the right along the main diagonal. The situation resulting for $T = 50000$ is plotted in figure 11.

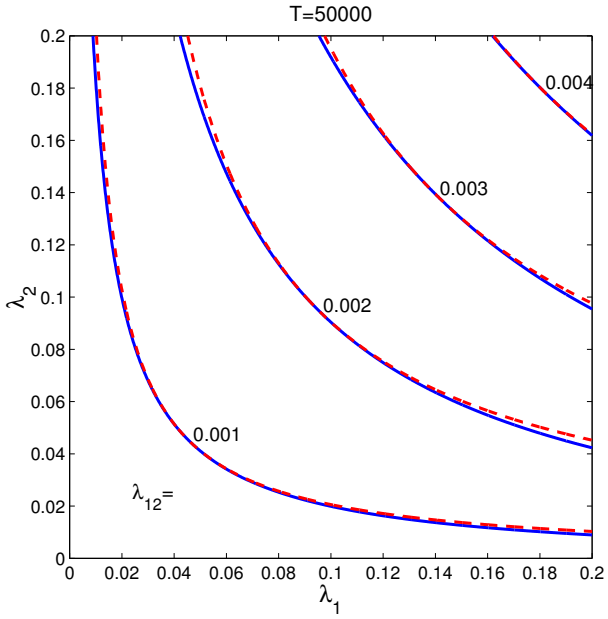

Figure 11: Comparison of the curves of equal background firing probability product (dashed) and curves of equal minimal required λ_{12} to reach a test power of 0.975 at $T = 50000$ (solid). Compare to figure 10 (B).

4 Three and More Neurons

This section will deal with the extension of the MIIP onto three neurons. In principle, the same considerations are made, and the same simulations and evaluations are carried out. The main difference between section three and four is to be found in the conceptual background. As described in the introduction, the main goal was to develop a tool that allows to distinguish between genuine and chance coincidences in the context of higher-order correlations. The problem for two neurons can as well be approached by the described Unitary-Event method, as for $n = 2$, absence of correlation is equivalent to independence of the processes. Yet for three neurons, this is no longer the case. The absence of a genuine triplet correlation does not prevent pairwise correlations, such that the null-hypothesis of the UE method needs to be modified to allow for sub-correlations. This section serves as a demonstration of the usefulness of the MIIP in the context of the original problem when applied onto more than two neurons. Therefore, all discussions are carried out in detail to show the direct extensibility of the model onto more than two parallel processes.

This section also contains a part that deals with general properties of the MIIP for n neurons. In subsection 4.2, a general formula for the maximum-likelihood estimates of all basic processes' firing probabilities will be shown for all numbers of neurons $n \in \mathbb{N}$. This will allow to draw conclusions about the estimates' asymptotic normality and variance.

4.1 The Model for Three Neurons

The model will now be demonstrated for three neurons. Mainly the correlation of highest order will be discussed, as the pairwise correlations can in principle be examined as described in section 3. Small modifications in the analysis of pairwise correlations will be accounted for in subsection 4.3.

The step from two to three neurons is regarded characteristic for the increase in complexity from n to $n + 1$ neurons. This complexity can already be seen in figure 12. The three observable processes now represent a superposition of seven basic processes. Three different triplet coincidences can be seen in the figure, one due to the triplet correlation and two due to pairwise correlations and/or background. To visualize this complexity more systematically, figure 14 shows all possibilities to get chance triplets. As we have seen that roughly the problem reduces

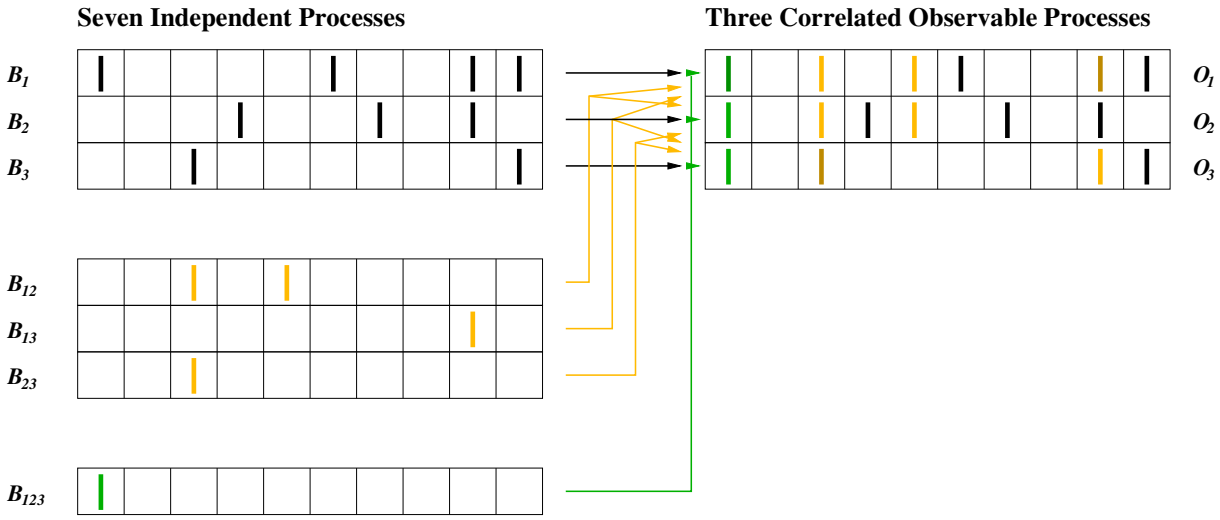

Figure 12: The MIIP for three neurons. There are $2^3 - 1$ basic processes of which three produce background spikes, three others are the origin of pairwise coincidences, and one produces genuine triplets. The basic processes are merged to get the observed processes (on the right).

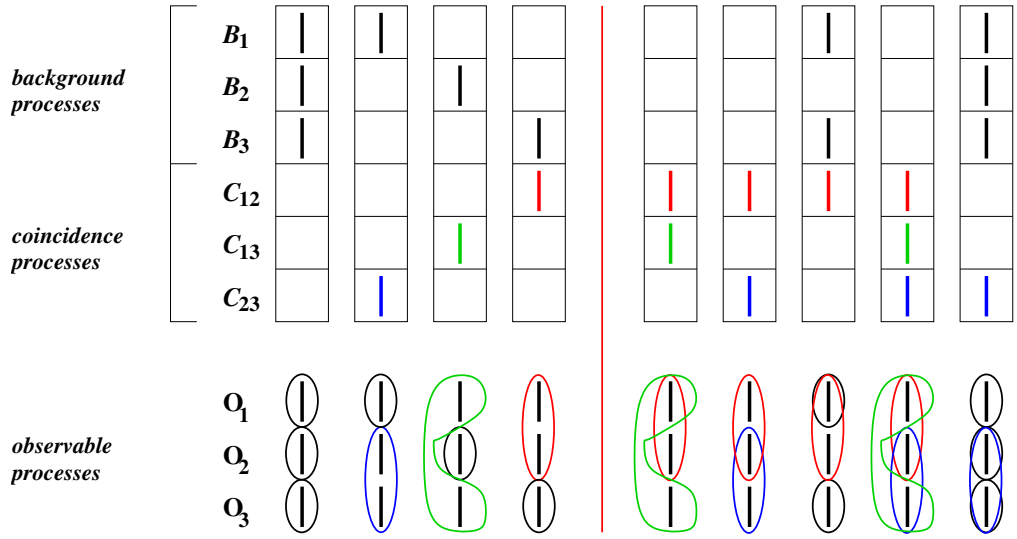

Figure 13: Difference between overlapping and non-overlapping events: Detailed representation of a part of figure 14. On the left, all four non-overlapping events produce each observed spike exactly once. In the overlapping events of which five are depicted on the right, at least one observed spike originates in at least two basic processes.

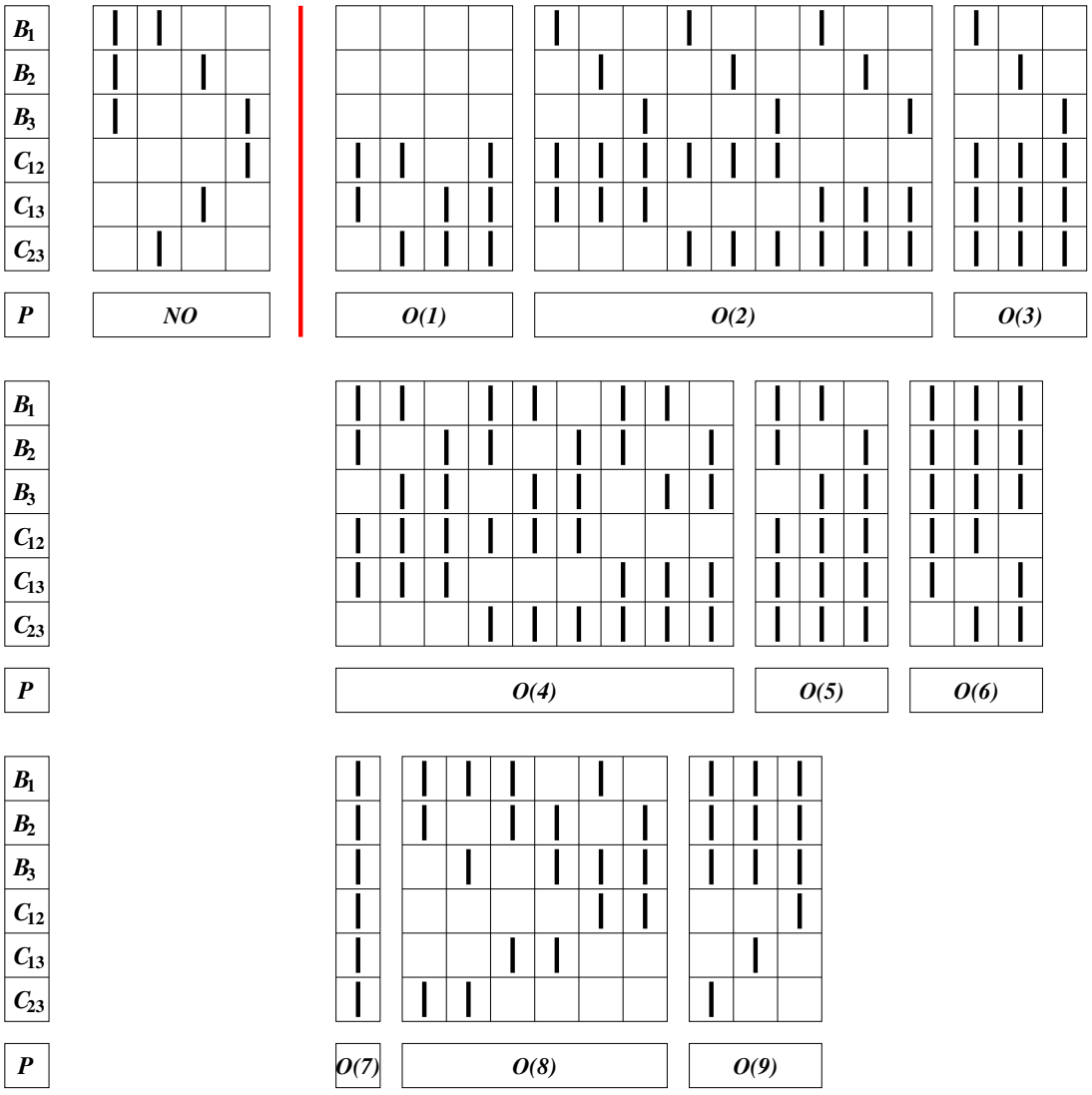

Figure 14: Visualization of all possible events that produce chance triplets. The variables below represent the probability of the events marked above.

to a comparison of chance and genuine coincidences, the increase in complexity from two to three neurons becomes quite obvious in the following: For two neurons, there was only one single possibility to get a chance coincidence, namely by a chance co-activation of both background processes. For three neurons, there are 45 combinations of successes and non-successes in the basic processes which produce a chance triplet. For later use, those will be conceptually subdivided into "overlapping" and "non-overlapping" events as follows:

In the left part of figure 13, the spikes produced in the basic processes are chosen such that in each observed process, only one spike is evoked. These will be called "non-overlapping" chance coincidences. In the right part, at least one spike per chance triplet is produced by at least two of the basic processes. Those chance coincidences will be referred to as "overlapping" events (for $n = 3$: 41 out of 45).

4.2 Maximum-Likelihood Estimates for n Neurons

The formulas for the maximum-likelihood estimates developed in section 3 can be extended onto larger models with more parallel processes. In spite of the increased complexity, the formulas are relatively small.
The notation is chosen as such:
Let $N = \{1, 2, ..., n\}$ be the set of observed neurons, $M \subseteq N$ a subset of them.
$\pi_M := P(\{O_i = 0 \; \forall \, i \in M\})$ denotes the probability to find no spike in the observed processes of the neurons in M at any fixed time $t \in \{1, ..., T\}$. For two neurons, this has been expressed with the symbol '+': $\pi_{\{1\}} = p_{0+}$ or $\pi_{\{1,2\}} = p_{00}$. Note that $\pi_\emptyset = p_{++} = 1$.
Due to Lemma 1, the following Lemma remains to be shown.

Proposition 2

For all $M_0 \subseteq N$ with $M_0 \neq \emptyset$

$$1 - \lambda_{M_0} = \frac{\prod_{M \subseteq M_0, \, |M| \neq |M_0| \text{mod} 2} \pi_{(N-M)}}{\prod_{M \subseteq M_0, \, |M| = |M_0| \text{mod} 2} \pi_{(N-M)}} \tag{33}$$

In particular one gets for $n = 3$:

$$1 - \lambda_{123} = \frac{p_{000}p_{++0}p_{+0+}p_{0++}}{p_{+00}p_{0+0}p_{00+}} \tag{34}$$

Formula (34) will be used in large parts of this section.
For the sake of completeness, we will now include a proof of Proposition 2.

Proof

Proposition 2 will be shown by induction over n.

For $n = 2$ it was already shown in section 3.1 (the indices of the λ are again written without brackets: e.g. $\lambda_{\{1\}} =: \lambda_1$):

$$1 - \lambda_1 = \frac{\pi_{N-\emptyset}}{\pi_{N-\{1\}}} = \frac{p_{00}}{p_{+0}},$$

analogous for $1 - \lambda_2$, and

$$1 - \lambda_{12} = \frac{\pi_{N-\{1\}}\pi_{N-\{2\}}}{\pi_{N-\emptyset}\pi_{N-N}} = \frac{p_{+0}p_{0+}}{p_{00}},$$

as $N = \{1, 2\}$.

To show what will be the main argument for the step from n to $n + 1$, this will be done exemplarily for $n = 2$.

1. For $n = 2$, all parameters λ_1, λ_2 and λ_{12} are determined by the following system of equations (recall the formulas (4)-(6) on page 16):

$$\begin{aligned} p_{00} &= (1 - \lambda_1)(1 - \lambda_2)(1 - \lambda_{12}) \\ p_{+0} &= (1 - \lambda_2)(1 - \lambda_{12}) \\ p_{0+} &= (1 - \lambda_1)(1 - \lambda_{12}) \\ p_{++} &= 1 \end{aligned}$$

2. For $n = 3$ and $N^+ := \{1, 2, 3\}$, there are more equations:

$$\begin{aligned}
p_{000} &= (1-\lambda_1)(1-\lambda_2)(1-\lambda_3)(1-\lambda_{12})(1-\lambda_{13})(1-\lambda_{23})(1-\lambda_{123}) \quad (35)\\
&= p_{00} \prod_{M\subseteq N^+,\, 3\in M} (1-\lambda_M)\\
p_{+00} &= (1-\lambda_2)(1-\lambda_3)(1-\lambda_{12})(1-\lambda_{13})(1-\lambda_{23})(1-\lambda_{123})\\
&= p_{+0} \prod_{M\subseteq N^+,\, 3\in M} (1-\lambda_M)\\
p_{0+0} &= (1-\lambda_1)(1-\lambda_3)(1-\lambda_{12})(1-\lambda_{13})(1-\lambda_{23})(1-\lambda_{123})\\
&= p_{0+} \prod_{M\subseteq N^+,\, 3\in M} (1-\lambda_M)\\
p_{00+} &= (1-\lambda_1)(1-\lambda_2)(1-\lambda_{12})(1-\lambda_{13})(1-\lambda_{23})(1-\lambda_{123})\\
p_{++0} &= (1-\lambda_3)(1-\lambda_{13})(1-\lambda_{23})(1-\lambda_{123})\\
&= p_{++} \prod_{M\subseteq N^+,\, 3\in M} (1-\lambda_M)\\
p_{+0+} &= (1-\lambda_2)(1-\lambda_{12})(1-\lambda_{23})(1-\lambda_{123})\\
p_{0++} &= (1-\lambda_1)(1-\lambda_{12})(1-\lambda_{13})(1-\lambda_{123})\\
p_{+++} &= 1
\end{aligned}$$

3. As far as the parameters λ_1, λ_2 and λ_{12} are concerned, the following is obvious:

$$(1-\lambda_1) = \frac{p_{00}}{p_{+0}} = \frac{p_{000}}{p_{+00}} \quad \text{and} \quad (1-\lambda_{12}) = \frac{p_{0+}p_{+0}}{p_{00}p_{++}} = \frac{p_{0+0}p_{+00}}{p_{000}p_{++0}},$$

because the same factors are added to the corresponding formulas. For symmetry reasons, the analogous is true for λ_3, λ_{13} and λ_{23}.

4. It remains to show that formula (33) is true for λ_{123}. For this, equation (35) will be solved:

$$\begin{aligned}
(1-\lambda_{123}) &= \frac{p_{000}}{(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)(1-\lambda_{12})(1-\lambda_{13})(1-\lambda_{23})(1-\lambda_{123})}\\
&= p_{000} \cdot \frac{p_{+00}}{p_{000}} \cdot \frac{p_{0+0}}{p_{000}} \cdot \frac{p_{00+}}{p_{000}} \cdot \frac{p_{000}p_{++0}}{p_{+00}p_{0+0}} \cdot \frac{p_{000}p_{+0+}}{p_{+00}p_{00+}} \cdot \frac{p_{000}p_{0++}}{p_{0+0}p_{00+}}\\
&= \frac{p_{000}p_{++0}p_{+0+}p_{0++}}{p_{+00}p_{0+0}p_{00+}},
\end{aligned}$$

which gives formula (34).

Thus for the step from n to $n+1$, which will be analogous, we need the following notations:

1. Let equation (33) be true for n neurons, i.e. the system of equations containing the formulas for all π_M, $M \subseteq N = \{1, ..., n\}$ is solved when replacing all factors $(1 - \lambda_M)$ with the right side of equation (33). Note that the numbers $1, ..., n$ are just labels of the neurons indicating that a set of n neurons is analyzed. One could as well replace these numbers by any kind of labeling.

2. For $n+1$ neurons, let $N^+ := \{1, ..., n, n+1\}$. The system of equations for all π_{M^+}, $M^+ \subseteq N^+$ will be shown to lead to equation (33).

3. In the third step we want to show that all parameters that were already existent in the model with n neurons can be estimated with the same formulas as before by replacing all used probabilities with those of the model with $n+1$ neurons, where the additional neuron is restricted to non-firing (compare to 3. in the step from two to three neurons).
   Let now $i \in N^+$ be one fixed element of N^+. We define $N^{(i)} := N^+ \setminus \{i\}$. For $M^{(i)} \subseteq N^{(i)}$, let $\pi_{M^{(i)}} := P(\{O_j = 0,\ j \in M^{(i)}\})$ be defined in the model with n neurons with labels in $N^{(i)}$. Let further $\pi^+_{M^{(i)}} := P(\{O_j = 0,\ j \in M^{(i)} \cup \{i\}\})$ be defined in the model with $n+1$ neurons with labels in N^+. For example for $n = 2$, $i = 3$, $M^{(i)} = \{1\}$: $\pi_{M^{(i)}} = p_{0+}$ and $\pi^+_{M^{(i)}} = p_{0+0}$. For step 3 it is enough to show that for all $M^{(i)}$ the following holds true

$$\pi^+_{M^{(i)}} = \pi_{M^{(i)}} \prod_{M^+ \subseteq N^+, i \in M^+} (1 - \lambda_{M^+}).$$

   This is obvious, as in the model that additionally contains the neuron with label i, no process must fire that would lead to a firing of neuron i.
   For symmetry reasons, this holds true for all $i \in N^+$and thus equation (33) has been shown to be true for all $(1 - \lambda_M)$ with $0 < |M| \leq n$.

4. It remains to show equation (33) for λ_{N^+}: Again, the formula for π_{N^+} will be used. By

replacing all factors $(1-\lambda_M)$ with the right hand side of equation (33), one gets

$$\begin{aligned}\pi_{N^+} &= \prod_{M\subseteq N^+,\, M\neq\emptyset} (1-\lambda_M) \\ &= \pi_{N^+}^{\sum_{k=1}^{n+1}\binom{n+1}{k}(-1)^{k-1}} \cdot \prod_{|M|=n} \pi_M^{\sum_{k=0}^{n}\binom{n}{k}(-1)^{k+1}} \cdot \prod_{|M|=n-1} \pi_M^{\sum_{k=0}^{n-1}\binom{n-1}{k}(-1)^{k+1}} \cdot \ldots \cdot \\ &\quad \cdot \prod_{|M|=1} \pi_M^{1-1} \cdot 1,\end{aligned}$$

which is true because

$$\sum_{k=0}^{m}\binom{m}{k}(-1)^k = (1-1)^m = 0, \quad \text{and thus} \quad \sum_{k=1}^{m}\binom{m}{k}(-1)^{k-1} = 1.$$

As all $(1-\lambda_M)$ with $0 < |M| \leq n$ have been shown in 3. to fulfill equation (33), the same formula for $(1-\lambda_{N^+})$ has thus been proven. □

Remark

With the help of the Propositions 1 and 2, one can see that the maximum-likelihood estimates of the MIIP have asymptotic normal distribution for all $n \geq 2$. The asymptotic variance can be derived as in subsection 3.2.1. For $n = 3$ it will be computed explicitly in the following subsection.

4.3 Asymptotic Variance of $\widehat{\lambda}_{123}$

The asymptotic normality of the estimate as well as its asymptotic variance will be shown by application of the multidimensional δ-method as in subsection 3.2.1 for $n = 2$. For the definition of the function f recall equation (34) on page 43:

$$1-\lambda_{123} = \frac{p_{000}p_{++0}p_{+0+}p_{0++}}{p_{+00}p_{0+0}p_{00+}}$$

For the application of Proposition 1, we use $d = 7$,

$$\begin{aligned}\widehat{\theta}_T &:= (\widehat{\theta}_{T,1}, ..., \widehat{\theta}_{T,7}) := \left(\left(\frac{1}{T}\right) \cdot (S_{000}, S_{+00}, S_{0+0}, S_{00+}, S_{++0}, S_{+0+}, S_{0++})\right)_T \\ \theta &:= (\theta_1, \theta_2, ..., \theta_7) := (p_{000}, p_{+00}, p_{0+0}, p_{00+}, p_{++0}, p_{+0+}, p_{0++})\end{aligned}$$

With the same argument as for $n = 2$ one gets

$$\mathcal{L}\left[\widehat{\theta}_T\right] \xrightarrow[T\to\infty]{} \mathcal{N}(\theta, T^{-1}\Sigma(\theta)),$$

with

$$\Sigma(\theta) = \begin{pmatrix} p_{000}(1-p_{000}) & p_{000}(1-p_{+00}) & p_{000}(1-p_{0+0}) \\ p_{000}(1-p_{+00}) & p_{+00}(1-p_{+00}) & p_{000}-p_{+00}p_{0+0} \\ p_{000}(1-p_{0+0}) & p_{000}-p_{+00}p_{0+0} & p_{0+0}(1-p_{0+0}) \\ p_{000}(1-p_{00+}) & p_{000}-p_{+00}p_{00+} & p_{000}-p_{0+0}p_{00+} \\ p_{000}(1-p_{++0}) & p_{+00}(1-p_{++0}) & p_{0+0}(1-p_{++0}) \\ p_{000}(1-p_{+0+}) & p_{+00}(1-p_{+0+}) & p_{000}-p_{0+0}p_{+0+} \\ p_{000}(1-p_{0++}) & p_{000}-p_{+00}p_{0++} & p_{0+0}(1-p_{0++}) \end{pmatrix}$$

$$\cdots \begin{pmatrix} p_{000}(1-p_{00+}) & p_{000}(1-p_{++0}) & p_{000}(1-p_{+0+}) & p_{000}(1-p_{0++}) \\ p_{000}-p_{+00}p_{00+} & p_{+00}(1-p_{++0}) & p_{+00}(1-p_{+0+}) & p_{000}-p_{+00}p_{0++} \\ p_{000}-p_{0+0}p_{00+} & p_{0+0}(1-p_{++0}) & p_{000}-p_{0+0}p_{+0+} & p_{0+0}(1-p_{0++}) \\ p_{00+}(1-p_{00+}) & p_{000}-p_{00+}p_{++0} & p_{00+}(1-p_{+0+}) & p_{00+}(1-p_{0++}) \\ p_{000}-p_{00+}p_{++0} & p_{++0}(1-p_{++0}) & p_{+00}-p_{++0}p_{+0+} & p_{0+0}-p_{++0}p_{0++} \\ p_{00+}(1-p_{+0+}) & p_{+00}-p_{++0}p_{+0+} & p_{+0+}(1-p_{+0+}) & p_{00+}-p_{+0+}p_{0++} \\ p_{00+}(1-p_{0++}) & p_{0+0}-p_{++0}p_{0++} & p_{00+}-p_{+0+}p_{0++} & p_{0++}(1-p_{0++}) \end{pmatrix}$$

The formulas for the covariances in $\Sigma(\theta)$ can be derived as in subsection 3.2.1.
Take now $f : \mathbb{R}^7 \longrightarrow \mathbb{R}$ with $f(x) := f(x_1, ..., x_7) := \frac{x_1x_5x_6x_7}{x_2x_3x_4}$. f has a differential at θ for p_{+00}, p_{0+0}, $p_{00+} > 0$ and $T > 0$. Again, $f(\theta) = (1 - \lambda_{123})$. It follows that $f(\widehat{\theta}_T)$ has

asymptotic normal distribution with asymptotic variance

$$\mathrm{Var}(f(\widehat{\theta}_T)) \xrightarrow[T\to\infty]{} \frac{1}{T}\sum_{k,k'=1}^{7}\sigma_{kk'}(\theta)\left(\frac{\partial f}{\partial\theta_k}\right)\left(\frac{\partial f}{\partial\theta_{k'}}\right)$$

$$\begin{aligned}
= \quad & \frac{p_{000}p_{++0}p_{+0+}p_{0++}}{Tp_{+00}^2p_{0+0}^2p_{00+}^2}\cdot \\
& \Bigg[(1-p_{000})p_{++0}p_{+0+}p_{0++}+\frac{(1-p_{+00})p_{000}p_{++0}p_{+0+}p_{0++}}{p_{+00}} \\
& +\frac{(1-p_{0+0})p_{000}p_{++0}p_{+0+}p_{0++}}{p_{0+0}}+\frac{(1-p_{00+})p_{000}p_{++0}p_{+0+}p_{0++}}{p_{00+}} \\
& +(1-p_{++0})p_{000}p_{+0+}p_{0++}+(1-p_{+0+})p_{000}p_{++0}p_{0++} \\
& +(1-p_{0++})p_{000}p_{++0}p_{+0+}-2\frac{(1-p_{+00})p_{000}p_{++0}p_{+0+}p_{0++}}{p_{+00}} \\
& -2\frac{(1-p_{0+0})p_{000}p_{++0}p_{+0+}p_{0++}}{p_{0+0}}-2\frac{(1-p_{00+})p_{000}p_{++0}p_{+0+}p_{0++}}{p_{00+}} \\
& +2(1-p_{++0})p_{000}p_{+0+}p_{0++}+2(1-p_{+0+})p_{000}p_{++0}p_{0++} \\
& +2(1-p_{0++})p_{000}p_{++0}p_{+0+}+2\frac{(p_{000}-p_{+00}p_{0+0})p_{000}p_{++0}p_{+0+}p_{0++}}{p_{+00}p_{0+0}} \\
& +2\frac{(p_{000}-p_{+00}p_{00+})p_{000}p_{++0}p_{+0+}p_{0++}}{p_{+00}p_{00+}}-2(1-p_{++0})p_{000}p_{+0+}p_{0++} \\
& -2(1-p_{+0+})p_{000}p_{++0}p_{0++}-2\frac{(p_{000}-p_{+00}p_{0++})p_{000}p_{++0}p_{+0+}}{p_{+00}} \\
& +2\frac{(p_{000}-p_{0+0}p_{00+})p_{000}p_{++0}p_{+0+}p_{0++}}{p_{0+0}p_{00+}}-2(1-p_{++0})p_{000}p_{+0+}p_{0++} \\
& -2\frac{(p_{000}-p_{0+0}p_{+0+})p_{000}p_{++0}p_{0++}}{p_{0+0}}-2(1-p_{0++})p_{000}p_{++0}p_{+0+} \\
& +2(p_{0+0}-p_{++0}p_{0++})p_{000}p_{+0+}+2(p_{00+}-p_{+0+}p_{0++})p_{000}p_{++0} \\
& -2(1-p_{0++})p_{000}p_{++0}p_{+0+}+2(p_{+00}-p_{+0+}p_{++0})p_{000}p_{0++} \\
& -2\frac{(p_{000}-p_{00+}p_{++0})p_{000}p_{+0+}p_{0++}}{p_{00+}}-2(1-p_{+0+})p_{000}p_{++0}p_{0++}\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{p_{000}p_{++0}p_{+0+}p_{0++}}{Tp_{+00}^2p_{0+0}^2p_{00+}^2}\cdot\Bigg[p_{000}p_{++0}p_{+0+}p_{0++}\left\{-\frac{1}{p_{00+}}-\frac{1}{p_{0+0}}-\frac{1}{p_{+00}}\right\}\\
&+2p_{000}^2p_{++0}p_{+0+}p_{0++}\left\{\frac{1}{p_{00+}p_{0+0}}+\frac{1}{p_{+00}p_{00+}}+\frac{1}{p_{0+0}p_{00+}}\right\}\\
&+2p_{000}\left\{p_{++0}p_{00+}+p_{0++}p_{+00}+p_{+0+}p_{0+0}\right\}-p_{000}p_{++0}p_{+0+}p_{0++}\\
&-p_{000}\left\{p_{+0+}p_{++0}+p_{+0+}p_{0++}+p_{++0}p_{0++}\right\}+p_{++0}p_{+0+}p_{0++}\\
&-2p_{000}^2\left\{\frac{p_{0++}p_{++0}}{p_{0+0}}+\frac{p_{+0+}p_{++0}}{p_{+00}}+\frac{p_{+0+}p_{0++}}{p_{00+}}\right\}\Bigg]\\
&= \frac{p_{000}p_{++0}p_{+0+}p_{0++}}{Tp_{+00}^2p_{0+0}^2p_{00+}^2}\cdot\Bigg[\frac{p_{000}p_{++0}p_{+0+}p_{0++}}{p_{+00}p_{0+0}p_{00+}}\cdot\\
&\cdot\left\{-p_{00+}p_{0+0}p_{+00}+2p_{000}(p_{+00}+p_{0+0}+p_{00+})-p_{00+}p_{0+0}-p_{00+}p_{+00}-p_{+00}p_{0+0}\right\}\\
&+p_{000}\left\{-p_{+0+}p_{++0}-p_{+0+}p_{0++}-p_{++0}p_{0++}\right\}\\
&+2p_{000}\left\{p_{++0}p_{00+}+p_{+0+}p_{0+0}+p_{0++}p_{+00}\right\}+p_{++0}p_{+0+}p_{0++}\\
&-2p_{000}^2\left\{\frac{p_{0++}p_{++0}}{p_{0+0}}+\frac{p_{+0+}p_{++0}}{p_{+00}}+\frac{p_{+0+}p_{0++}}{p_{00+}}\right\}\Bigg]\\
\sigma^2_{\widehat{\lambda}_{123}} &= \frac{(1-\lambda_{123})}{T(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)(1-\lambda_{12})(1-\lambda_{13})(1-\lambda_{23})}\cdot\\
&[-p_{000}+3-3\left\{\lambda_1+\lambda_2+\lambda_3\right\}+2\lambda_1\lambda_2+2\lambda_1\lambda_3+2\lambda_2\lambda_3\\
&-(1-\lambda_1)(1-\lambda_2)(1-\lambda_{12})-(1-\lambda_1)(1-\lambda_3)(1-\lambda_{13})\\
&-(1-\lambda_2)(1-\lambda_3)(1-\lambda_{23})+2(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)\cdot\\
&\cdot\left\{(1-\lambda_{12})(1-\lambda_{13})+(1-\lambda_{12})(1-\lambda_{23})+(1-\lambda_{13})(1-\lambda_{23})\right\}\\
&-2(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)\left\{1-\lambda_{13}+1-\lambda_{23}+1-\lambda_{12}\right\}+1]\\
&= \frac{(1-\lambda_{123})}{T(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)(1-\lambda_{12})(1-\lambda_{13})(1-\lambda_{23})}\cdot\\
&[(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)(1-\lambda_{12})(1-\lambda_{13})(1-\lambda_{23})\lambda_{123}\\
&+(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)(\lambda_{12}\lambda_{13}+\lambda_{12}\lambda_{23}+\lambda_{13}\lambda_{23}+\lambda_{12}\lambda_{13}\lambda_{23})\\
&+(1-\lambda_1)(1-\lambda_2)\lambda_{12}\lambda_3+(1-\lambda_1)(1-\lambda_3)\lambda_{13}\lambda_2\\
&+(1-\lambda_2)(1-\lambda_3)\lambda_{23}\lambda_1+\lambda_1\lambda_2\lambda_3] \qquad (36)
\end{aligned}
$$

This formula will be used to construct the test statistics $Z := \frac{\widehat{\lambda}_{123}}{\sigma_{\widehat{\lambda}_{123}}}$, which will be compared to $z^* := \Phi^{-1}(1-\alpha)$ just like in the two-neurons-case.

For the lower-order correlations the formula for the variance changes slightly to account for the additional parameters introduced in the model. By analogous computation, one gets

$$\mathrm{Var}(1-\widehat{\lambda}_{12}) = \frac{(1-\lambda_{12})(\lambda_{12}(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2)}{T(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)(1-\lambda_{13})(1-\lambda_{23})(1-\lambda_{123})}.$$

Compared to equation (19) on page 21, the numerator stays the same, whereas the denominator now additionally contains all factors $(1-\lambda_M)$, where $M \subseteq \{1,2,3\}$ and $3 \in M$.

4.4 Performance of the Test

Just like in the two-neurons-case, the following questions need to be approached:

1. What happens as long as the piece of data is not long enough to grant the applicability of the asymptotics?
2. What requirements regarding the size of T need to be met to use the asymptotics?
3. Given the asymptotics are applicable, what do we learn about the relation between the different parameters T, α, β, $\lambda_1, \lambda_2, \lambda_3, \lambda_{12}, \lambda_{13}, \lambda_{23}$ and λ_{123}? Not all interdependencies between the variables can be discussed here. So mainly the same considerations as for $n = 2$ will be made.

4.4.1 Applicability to Finite Strings of Data

As in subsection 3.3.1, the empirical significance level and test power derived from simulations will be compared with their asymptotic correspondents for different firing probabilities and lengths T of the data set. As the number of parameters has increased, typical values need to be inspected in order to reduce the computational effort necessary for the simulations in this subsection. Therefore, situations with equal background firing probabilities as well as equal pairwise coincidence probabilities were examined. Furthermore, the pairwise coincidence probabilities are chosen to take on the values $0, 0.002$ and 0.004, and the triplet firing probability

λ_{123} was chosen to be either 0 or 0.002. For T, the values 10000 and 50000 were inspected. As in general, the results are analogous for other values of either firing probabilities or T, the chosen values are considered sufficient to show the effects of a change in those variables.

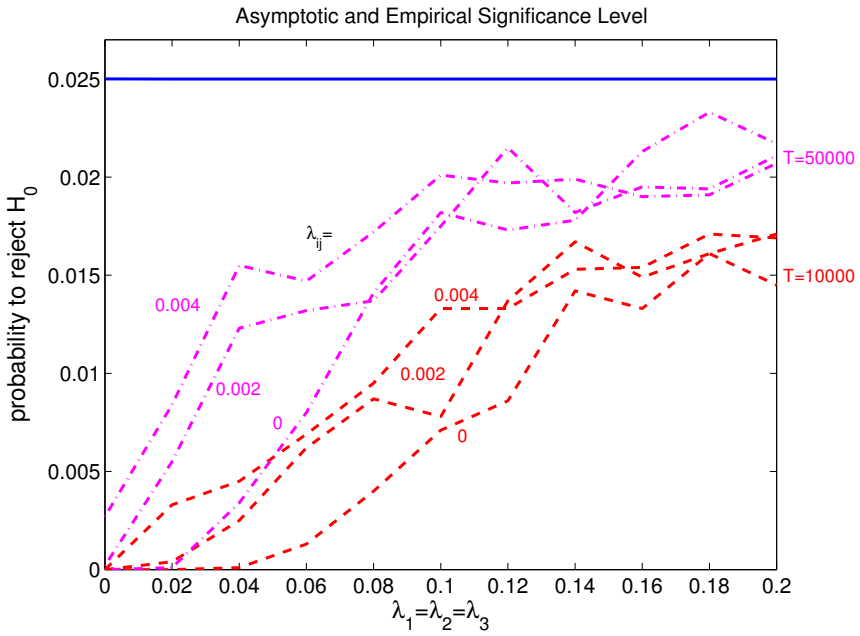

Figure 15: Comparison of empirical and asymptotic (solid) significance level for the test and the MIIP for $n = 3$. $\alpha = 0.025$, $\lambda_1 = \lambda_2 = \lambda_3 \in \{0, 0.2, 0.4, ..., 0.2\}$, $\lambda_{12} = \lambda_{13} = \lambda_{23} \in \{0, 0.002, 0.004\}$, $\lambda_{123} = 0$, 10000 trials simulated per data point. All simulations done for $T = 10000$ (dashed) and $T = 50000$ (dash-dotted).

Figure 15 compares the asymptotic significance level (chosen to be 0.025, solid curve) with the empirical significance level derived from simulations. The dashed curves show the relative number of significant experiments for $T = 10000$ depending on equal background rates, split up into three different curves corresponding to the three different pairwise firing probabilities. The dash-dotted curves show the analogous for $T = 50000$. Per parameter set, 10000 experiments were simulated, i.e. 10000 data sets of length T, consisting of the three background and the three pairwise correlation processes, were simulated and merged onto the three observable processes O_1, O_2 and O_3 to calculate the test statistics $Z = \frac{\widehat{\lambda}_{123}}{\sigma_{\widehat{\lambda}_{123}}}$, which was then compared to the value $z^* = 1.96$. The number of experiments with $Z > z^*$ is plotted. One can observe the

same as in the analogous figure for $n = 2$:

- The test does not reach the pre-defined significance level for "small" T.
- The smaller T and the smaller the background and pairwise firing probabilities, the larger the difference between the empirical and the asymptotic significance level.
- Still, the test stays conservative, meaning that the probability to falsely reject $H_0 : \lambda_{123} = 0$ is smaller than required.
- One additional observation is that the empirical significance level for both T in question is even smaller than its analogous in the two-neurons-case (compare to figure 3), meaning that the probability to falsely find $\lambda_{123} > 0$ is even smaller. This is considered to be a border effect in the used parameter ranges, as the background firing probabilities and the pairwise firing probabilities are so small that too few events are produced by chance.

In figure 16, empirical and asymptotic test power are compared for $\lambda_{123} = 0.002$ and two different T. As for $n = 2$, one can see that

- the matching of asymptotic and empirical test power improves with increasing T,
- both empirical and asymptotic test power grow with T,
- and the higher the pairwise correlation firing probabilities and the background firing probabilities, the smaller is the test power.
- Additionally, the dashed and the dotted lines represent the analogous values for the same parameter sets and $n = 2$. One can see that the test power has increased, which will be discussed in subsection 4.4.2.2.

4.4.2 Discussion of the Asymptotic Test Power

We will now discuss theoretical relations between the parameters. Just as for $n = 2$, the minimal T and minimal λ_{123} to reach a pre-given test power at given background and pairwise coincidence firing probabilities will be examined. One additional aspect - background reduction - will be introduced to explain some of the results.

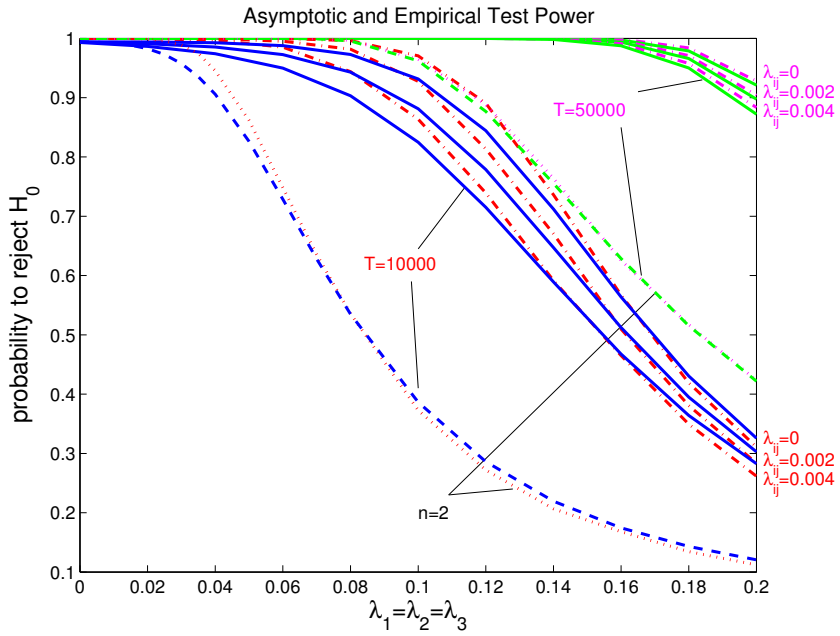

Figure 16: Comparison of empirical and asymptotic test power for the test and the model for $n = 3$.
solid: asymptotic test power for $n = 3$ and $\lambda_{123} = 0.002$, all other parameters as in figure 15. $T = 10000$, $T = 50000$ as indicated in the figure.
dash-dotted: empirical test power for corresponding parameter sets ($n = 3$, $\lambda_{123} = 0.002$).
dashed: asymptotic test power for $n = 2$, $\lambda_{12} = 0.002$ and $T = 10000$, $T = 50000$ as indicated.
dotted: empirical test power for $n = 2$, $\lambda_{12} = 0.002$ and $T = 10000$, $T = 50000$ as indicated.

4.4.2.1 Minimal Required Number of Time Steps T

In figure 17, the minimal number of time steps to reach a test power of $1 - \beta = 0.975$ is plotted depending on the background firing probabilities ($\lambda_{123} = 0.002$ and $\lambda_{12} = \lambda_{13} = \lambda_{23} \in \{0, 0.001, ..., 0.005\}$). Like in the two-neurons-case,

- the minimal required T grows with increasing background and pairwise firing probabilities, because they disturb the detection of genuine triplets.
- Moreover, the required T is smaller than for $n = 2$ when comparing the data points at equal background firing probabilities (compare with figure 7). This is considered to be due to the same effect as the increase in test power for the same parameter sets, and will be discussed in the following subsection.

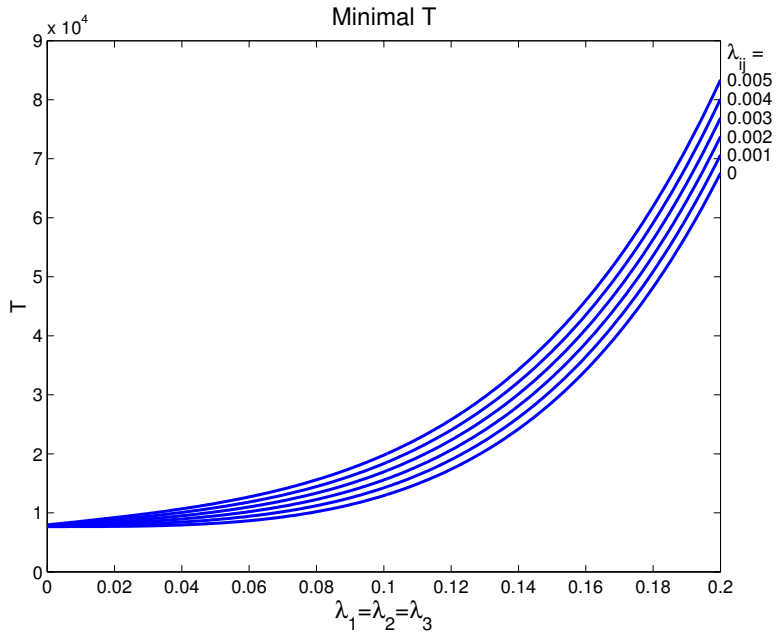

Figure 17: Minimal required number of time steps to reach a test power of 0.975 for $\lambda_{123} = 0.002$, $\lambda_1 = \lambda_2 = \lambda_3 \in [0, 0.2]$, $\lambda_{12} = \lambda_{13} = \lambda_{23} \in \{0, 0.001, ..., 0.005\}$.

4.4.2.2 About probabilities of chance coincidences - "Background Reduction"

We have seen that for corresponding parameter sets, the empirical and asymptotic test power at the tested parameters are higher for $n = 3$. Also we have seen that the required T to reach a pre-defined test power is smaller for $n = 3$ and equal background firing probabilities. In this subsection, an effect called "background reduction" will be introduced on the basis of which one can understand the mentioned results. In this place, only the symmetrical case will be discussed, i.e. equality of all background firing probabilities and equality of all pairwise coincidence probabilities. The last subsection of 4.4 will deal with deviations from this assumption. "Background reduction" means that for the same parameter set, the probability of a chance triplet (for $n = 3$) is (for the used parameter ranges and symmetry) smaller than the probability of a chance doublet (for $n = 2$). This may at first seem counter-intuitive because of the many ways to get a chance triplet. But it is plausible when considering two facts: First, the probability to get any chance triplet is considerably smaller than the probability of a chance doublet, which is not astonishing, because - roughly speaking - more spikes need to be produced, and the probability of a spike is smaller than that of a non-spike. Secondly, the biggest part of the probability of a chance triplet is due to only four configurations, namely the non-overlapping chance triplets as illustrated in figure 14 on page 41. The effect of the background reduction will now be stated and proven. Let

1. p_{11}^c be the probability to get a chance doublet for $n = 2$, conditioned on the non-firing of the doublet process,

2. p_{111}^c be the probability to get a chance triplet for $n = 3$, conditioned on the non-firing of the triplet process,

3. $\lambda_1^{(2)}, \lambda_2^{(2)}, \lambda_{12}^{(2)}$ be the firing probabilities for $n = 2$,

4. $\lambda_1^{(3)}, \lambda_2^{(3)}, \lambda_3^{(3)}, \lambda_{12}^{(3)}, \lambda_{13}^{(3)}, \lambda_{23}^{(3)}, \lambda_{123}^{(3)}$ be the firing probabilities for $n = 3$, where $\lambda_1^{(2)} = \lambda_1^{(3)} = \lambda_2^{(2)} = \lambda_2^{(3)} = \lambda_3^{(3)} =: \lambda_b$ and $\lambda_{12}^{(3)} = \lambda_{13}^{(3)} = \lambda_{23}^{(3)} =: \lambda_c$ and $\lambda_b \in (0, 0.2]$, $\lambda_c \in [0, 0.01]$,

5. and let the higher-order firing probabilities be of the same size: $\lambda_{\text{ho}} = \lambda_{12}^{(2)}$ and $\lambda_{\text{ho}} = \lambda_{123}^{(3)}$, respectively (where 'ho' stands for 'higher order').

6. It will further be assumed $\lambda_b > 0$ and $\lambda_b \geq 5 \cdot \lambda_c$. This condition is sufficient, as will be seen in the proof on page 57. Because of the usually very small size of λ_c, this is not a strong restriction, meaning that for relatively large values of pairwise coincidence rates, the background rate needs to be "large enough", i.e. for high $\lambda_c = 0.01$ we require $\lambda_b \geq 0.05$. The normal experimental situation is confronted with a much larger quotient than $\frac{\lambda_b}{\lambda_c} = 5$, which is not astonishing, as the whole debate about coding by coincidence would not arise if about one sixth of a neuron's spikes took part in coincident firing. Note however that $\lambda_b < 5\lambda_c$ does not affect the general procedure of the data analysis at all. Only the existence of background reduction cannot be guaranteed any more.

Then we get

$$p_{11}^c = \lambda_b^2 \cdot (1 - \lambda_{\text{ho}}) \tag{37}$$

and

$$p_{111}^c = (1 - \lambda_{\text{ho}}) \cdot [NO + O], \tag{38}$$

where

$$NO := \lambda_b^3 \cdot (1 - \lambda_c)^3 + 3 \cdot \lambda_b \cdot \lambda_c \cdot (1 - \lambda_b)^2 \cdot (1 - \lambda_c)^2 \tag{39}$$

is the probability of a non-overlapping chance triplet, given the process C_{123} did not fire, and

$$O := O(1) + O(2) + ... + O(9) \tag{40}$$

is the probability to get an overlapping chance triplet, given the process C_{123} did not fire, with

$$\begin{aligned}
O(1) &= 3 \cdot (1 - \lambda_b)^3 \cdot \lambda_c^2 \cdot (1 - \lambda_c) + (1 - \lambda_b)^3 \cdot \lambda_c^3 \\
O(2) &= 9 \cdot \lambda_b \cdot (1 - \lambda_b)^2 \cdot \lambda_c^2 \cdot (1 - \lambda_c) \\
O(3) &= 3 \cdot \lambda_b \cdot (1 - \lambda_b)^2 \cdot \lambda_c^3 \\
O(4) &= 9 \cdot \lambda_b^2 \cdot (1 - \lambda_b) \cdot \lambda_c^2 \cdot (1 - \lambda_c) \\
O(5) &= 3 \cdot \lambda_b^2 \cdot (1 - \lambda_b) \cdot \lambda_c^3 \\
O(6) &= 3 \cdot \lambda_b^3 \cdot \lambda_c^2 \cdot (1 - \lambda_c) \\
O(7) &= \lambda_b^3 \cdot \lambda_c^3 \\
O(8) &= 6 \cdot \lambda_b^2 \cdot (1 - \lambda_b) \cdot \lambda_c \cdot (1 - \lambda_c)^2,
\end{aligned}$$

and

$$O(9) \quad = \quad 3 \cdot \lambda_b^3 \cdot \lambda_c \cdot (1-\lambda_c)^2$$

(compare with figure 14). Thus

$$O = 3 \cdot \lambda_b \lambda_c (1-\lambda_c)(1-(1-\lambda_b)^2(1-\lambda_c)) + \lambda_b \lambda_c^3 + 2 \cdot \lambda_c^2 (1-\lambda_b)(1-\lambda_c) + \lambda_c^2 (1-\lambda_b)$$

We will now show the main

Claim (background reduction):

$$p_{11}^c > p_{111}^c \tag{41}$$

Proof

Recalling equations (38) and (39) we obtain

$$\frac{p_{111}^c}{(1-\lambda_{\text{ho}})} = \lambda_b^3(1-\lambda_c)^3 + \lambda_b \lambda_c^3 + \lambda_c^2(1-\lambda_b) + 3 \cdot \lambda_b \lambda_c (1-\lambda_c) + 2 \cdot \lambda_c^2 (1-\lambda_b)(1-\lambda_c) \tag{42}$$

and with assumption no. 6:

$$< \lambda_b^3 + \frac{3}{5}\lambda_b^2 + \frac{1}{125}\lambda_b^4 + \frac{3}{25}(1-\lambda_b)\lambda_b^2 \tag{43}$$

It thus suffices to show

$$\begin{aligned} 1 &> \frac{1}{125}\lambda_b^2 + \frac{22}{25}\lambda_b + \frac{18}{25} \\ \iff \quad & \lambda_b^2 + 110\lambda_b - 35 < 0, \end{aligned}$$

which is true for the whole parameter range $0 < \lambda_b \leq 0.2$. □

4.4.2.3 Minimal Required λ_{123}

In this subsection, figure 18 (which is similar to figure 10) will be discussed. One should learn something about the relation between the minimal required higher-order firing probability

and all other firing probabilities for fixed test power (and fixed T). In figure 10 one could observe that the minimal required higher-order firing probability was in principle a function of the product of the background rates, which is the probability $\frac{p_{11}^c}{(1-\lambda_{12})}$ of a chance doublet, conditioned on the non-firing of the higher-order process. For three neurons, the formula for p_{111}^c is definitely more complex. It can be calculated with the equations (38)-(40). In order to use it for the desired figure, a short approximation of the formula needs to be developed. In the following paragraph it will turn out that (with $\lambda_1 = \lambda_2 = \lambda_3 =: \lambda_b$ and $\lambda_{12} = \lambda_{13} = \lambda_{23} =: \lambda_c$) $p_{111}^{ap} := \lambda_b^3 + 3 \cdot \lambda_b \lambda_c$ is a very good approximation for $\frac{p_{111}^c}{(1-\lambda_{123})}$. It is thus used in figure 18, which compares the level lines of equal minimal required λ_{123} (solid) with those curves of equal p_{111}^{ap} (dashed).

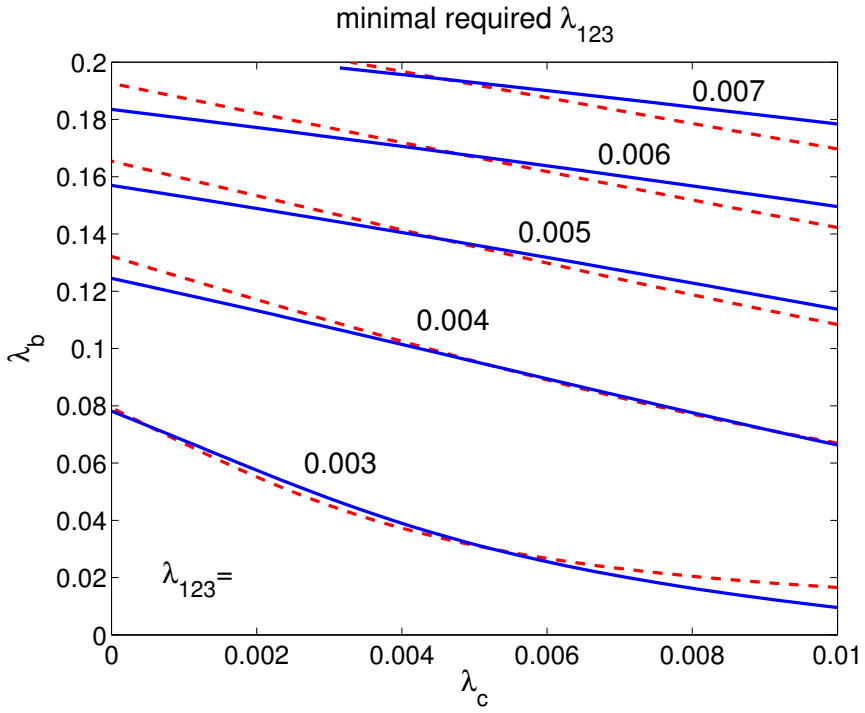

Figure 18: Comparison of the level lines of minimal required λ_{123} with the curves of constant p_{111}^{ap}. Solid: minimal required λ_{123} to reach a test power of 0.975 at $T = 10000$, given $\lambda_1 = \lambda_2 = \lambda_3 =: \lambda_b \in [0, 0.2]$ and $\lambda_{12} = \lambda_{13} = \lambda_{23} =: \lambda_c \in [0, 0.01]$. Dashed: Curves of equal $p_{111}^{\text{ap}} := \lambda_b^3 + 3\lambda_b\lambda_c$. The value of p_{111}^{ap} is chosen such that the corresponding curves intersect at $\lambda_c = 0.005$.

The value of p_{111}^{ap} is chosen such that the curves intersect at $\lambda_c = 0.005$. The parameter T is chosen to be 10000, and $1 - \beta := 0.975$. One can see that just like in the two-neurons-case, for constant T and constant test power, the minimal required λ_{123} is essentially a function of the probability of a chance triplet, given the triplet-process did not fire: $\frac{p_{111}^c}{(1-\lambda_{123})}$.

4.4.2.4 An Approximation for p_{111}^c

We propose to use $p_{111}^{\text{ap}} := \lambda_b^3 + 3\lambda_b\lambda_c$ as an approximation for $\frac{p_{111}^c}{(1-\lambda_{123})}$.

The heuristic argument is as follows: With the help of figure 14 on page 41, one can figure out that p_{111}^{ap} contains the probabilities of most of the overlapping events additionally to those represented by NO. Dividing up p_{111}^{ap} into sums of probabilities, one sees that it correctly contains 34 non-overlapping events. Thus, only 7 non-overlapping events are left out. Additionally, 18 non-overlapping events are included a second time. Still, those incorrectly represented events play a minor role for those parameter ranges in the experimental practice.

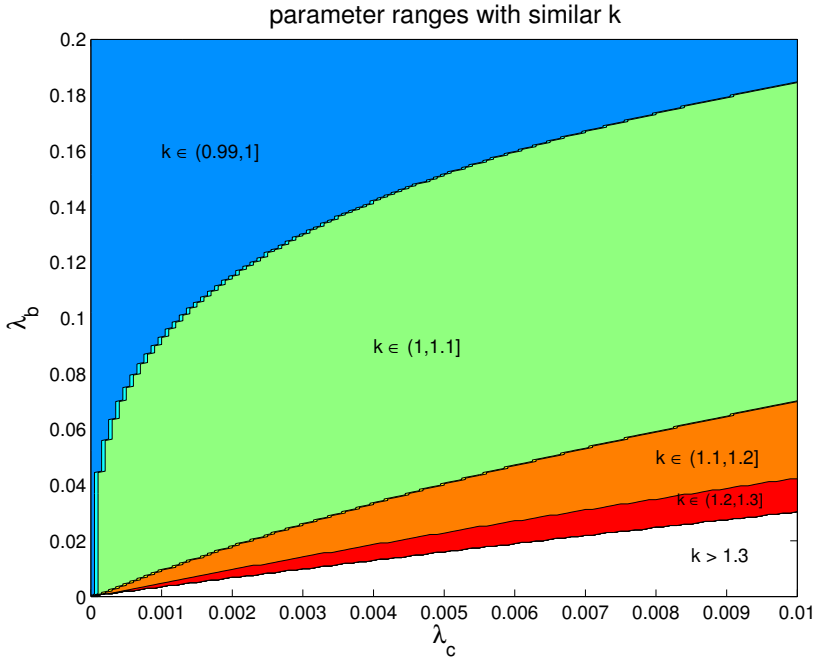

Figure 19: Parameter ranges where the quotient $k := \frac{p_{111}^c}{p_{111}^{\text{ap}}} \cdot \frac{1}{(1-\lambda_{123})}$ can be found in the given intervals. For $\lambda_b \geq 5\lambda_c$, the quotient is not larger than 1.2 (mostly even not larger than 1.1).

In figure 19 one can see that $k := \frac{p^c_{111}}{p^{\text{ap}}_{111}} \cdot \frac{1}{(1-\lambda_{123})}$ is never smaller than 0.98, and that for $\lambda_b \geq 5\lambda_c$, $k \leq 1.2$. This means that for the parameter ranges considered essential for the experimental situation, p^{ap}_{111} is quite a good approximation of $\frac{p^c_{111}}{(1-\lambda_{123})}$.

4.4.2.5 Accounting for Inequalities

The results of subsection 4.4 are restricted to $\lambda_1 = \lambda_2 = \lambda_3$ and $\lambda_{12} = \lambda_{13} = \lambda_{23}$. Some of them (as e.g. the increase of the required T for growing background and coincidence probabilities) should remain valid also under inequality of the subgroup firing probabilities. Also the fact that the minimal λ_{123} depends essentially on the probability of a chance triplet (for fixed T and constant test power) is plausible and will thus remain unproved here.

Instead, very asymmetrical parameter combinations can prevent the effect of the background reduction discussed above. An extreme case like $\lambda_1 = \lambda_2 = 0.05$, $\lambda_{12} = 0.01$, $\lambda_3 = 0.2$ would lead to

$$0.0025 = \frac{p^c_{11}}{(1-\lambda_{12})} < NO \approx 0.003,$$

because by $\frac{p^c_{11}}{(1-\lambda_{12})} = \lambda_1\lambda_2$, the parameter λ_3 is left unspecified and all λ_{ij} are only restricted to $\lambda_{ij} \leq \frac{1}{5}\min(\lambda_1, \lambda_2)$. Thus, if λ_1 and λ_2 are relatively small, but large enough to allow for a very large λ_{12}, and if additionally the third background rate is very high, then the effect of the background reduction is even inverted.

But one can develop a very rough guideline for the relation between λ_1, λ_2 and λ_3 that is sufficient for the validity of background reduction. The requirement will be a modification of the previously used condition $\lambda_b \geq 5 \cdot \lambda_c$:

$$\lambda_i \geq 5 \cdot \lambda_{jk} \ \forall i, j, k \in \{1, 2, 3\},\ j \neq k.$$

With a numerical computation, one can show that on the grid $\lambda_1, \lambda_2, \lambda_3 \in \{0.01, 0.02, ..., 0.2\}$, $\lambda_{12}, \lambda_{13}, \lambda_{23} \in \{0, 0.001, ..., 0.1\}$,

$$0.99 < \frac{p^c_{111}}{p^{\text{ap}}_{111}} \cdot \frac{1}{(1-\lambda_{123})} < 1.2$$

for

$$p_{111}^{\text{ap}} = \lambda_1\lambda_2\lambda_3 + \lambda_1\lambda_{23} + \lambda_2\lambda_{13} + \lambda_3\lambda_{12}.$$

This means that just like under equality of background and pairwise probabilities, respectively, p_{111}^{ap} is a good approximation for $NO + O$, i.e. the probability of a chance triplet, conditioned on the non-firing of the triplet process. This means that background reduction is present if the parameters fulfill

$$\begin{aligned} \lambda_1\lambda_2 &\geq (\lambda_1\lambda_2\lambda_3 + \lambda_1\lambda_{23} + \lambda_2\lambda_{13} + \lambda_3\lambda_{12}) \cdot 1.2 \\ \iff \frac{5}{6} &\geq \frac{2}{5} + \lambda_3 + \frac{\lambda_3}{5\text{max}(\lambda_1, \lambda_2)} \geq \lambda_3 + \frac{\lambda_{23}}{\lambda_2} + \frac{\lambda_{13}}{\lambda_1} + \frac{\lambda_3\lambda_{12}}{\lambda_1\lambda_2} \\ \iff \lambda_3 &\leq \frac{65\text{max}(\lambda_1, \lambda_2)}{30(1 + 5\text{max}(\lambda_1, \lambda_2))} \end{aligned}$$

It must be said that for large $\text{max}(\lambda_1, \lambda_2)$, λ_3 can even be larger than allowed by the last equation, because then, $\frac{\lambda_{23}}{\lambda_2}$ as well as $\frac{\lambda_{13}}{\lambda_1}$ are smaller than $\frac{1}{5}$.

In those parameter ranges where background reduction is not present, the observed effects - the higher test power, the smaller required T or lower significance level - can be assumed as being contrary, because the probability of a chance triplet is higher than that of a chance doublet for the subgroup $\{1, 2\}$. Still, for symmetry reasons, one could also compare the probability of a chance triplet with a chance doublet of the subgroup $\{1, 3\}$. Thus one can say that

- the background reduction is present for $\lambda_1 = \lambda_2 = \lambda_3 =: \lambda_b$ and $\lambda_{12} = \lambda_{13} = \lambda_{23} =: \lambda_c$ and $\lambda_b \geq 5 \cdot \lambda_c$,
- and for arbitrary $\lambda_1, \lambda_2, \lambda_3 \in (0, 0.2]$, $\lambda_{12}, \lambda_{13}, \lambda_{23} \in [0, 0.01]$, as long as $\lambda_i \geq 5 \cdot \lambda_{jk}$ $\forall i, j, k \in \{1, 2, 3\}$, $j \neq k$ and $\lambda_3 \leq \frac{65\text{max}(\lambda_1, \lambda_2)}{30(1+5\text{max}(\lambda_1, \lambda_2))}$.
- For symmetry reasons, one could always speak of background reduction when comparing three neurons with one of the subgroups that contains the neuron with the highest background firing probability.

5 Conclusions of Part I

1. In order to cope with the experimentally derived hypotheses about information processing in the cortex, a model has been developed. It has been shown to allow for the analysis of higher-order coincident firing in two and more parallel processes, including the distinction between genuine and chance higher-order correlations.

2. With maximum-likelihood estimation, the parameters representing the genuine correlations between all subgroups of observed neurons can be estimated. For any number of neurons n, a formula for the maximum-likelihood estimates has been derived.

3. By use of the estimates' asymptotic normality and asymptotic variance, a test has been developed for two and three processes in order to decide whether a genuine correlation of highest order is existent or not.

4. Empirical test power and significance level have been derived from simulations and compared to the asymptotic values, thereby showing that the test stays conservative in ranges where the asymptotics are not applicable.

5. Discussions of the asymptotic test power have demonstrated that for the examined parameter range and $\lambda_1 = \lambda_2 = \lambda_2$ and $\lambda_{12} = \lambda_{13} = \lambda_{23}$,
   - for constant test power and fixed higher-order firing probability, the required T grows with the background and subgroup firing probabilities,
   - for constant T, the required minimal higher-order firing probability to reach a given test power grows with the subgroup firing probabilities,
   - the higher the higher-order firing probability at constant T and constant subgroup firing probabilities, the higher is the test power,
   - the probability of a chance higher-order coincidence essentially determines the size of either the minimal required higher-order firing probability at constant T and test power or the test power at fixed T and constant higher-order firing probability.

6. The transition from two to three processes has two main effects:
   The complexity is increased, meaning that the number of parameters grows like 2^n, and that the formulas for e.g. the asymptotic variance or for the probability of a chance triplet

are more complex.

Background reduction: The properties of the test improve for $\lambda_b \geq 5\lambda_c$, $\lambda_1 = \lambda_2 = \lambda_2$ and $\lambda_{12} = \lambda_{13} = \lambda_{23}$. The test power is higher, and the number of time steps required to reach a pre-defined test power is smaller than for two neurons. All this is due to the fact that the probability of a chance higher-order coincidence decreases. If the condition of equality of background and equality of pairwise firing probabilities is not met, background reduction is guaranteed when comparing all three neurons with one subgroup that contains the neuron with the highest background rate. Note that background reduction is no general effect. For $\lambda_b < 5\lambda_c$, it is not discussed. However, absence of background reduction has no impact on the data-analysis procedure, which is generally applicable.

When developing an analysis method, one needs to keep in mind the structure of the data. As any model can only be a compromise between the complexity of reality on the one hand and analytical clarity on the other, the gap between the experimental situation and its mathematical simplification must lead to an ongoing dialogue. As described in the introduction, correlations found between neuronal pairs show an imprecision up to a few milliseconds. This implies that when applying the model, only those spikes are found to be coincident whose time of occurrence does not differ by more than one time-unit *and* which - due to an unlucky discretization of time - are not distributed into adjacent bins. Hence, spikes that are not found in the same bin but which would conceptually be considered to belong to a coincidence because of a small time difference both reduce the estimated coincidence firing probability and additionally increase the background. Thus the correlation is less easily detected (compare to Grün et al., 1999).

Unfortunately, this is not the only problem. Experimental data often show highly non-stationary firing rates. This is approached by reducing the length T of the window to a size where stationarity is assumed. But as the "real" firing rate can only be measured by averaging the number of spikes in a time period, this measure is subject to rate fluctuations. Thus one can never be sure whether the "real" firing rate has been constant in the observed period. It therefore seems best to use the minimal size of T possible. But this on the other hand is restricted by the time scale of less precise coincidences: If T is chosen too small, one could conceptually say that in the extreme case exact coincidences are compared to those coincidences with a small imprecision, which fatally changes the interpretation of the analysis's result. Therefore in the following part the analysis of less precise coincidences is dealt with in an extended model.

Part II

'Jittered' Coincidences

There is experimental evidence that the timing accuracy of spikes can be as precise as 1-5 ms (Abeles, Bergman, Margalit and Vaadia, 1993; Riehle et al., 1997). Therefore in this part, the analysis of coincidences with a small imprecision will be discussed. An extension of the described MIIP will be presented that can in principle deal with any time lag between conceptually "coincident" spikes and with any number of neurons. This part is not thought to be a complete description of the analysis method for all possible parameter sets. Instead, it intends to present the method and to show with simulations that it works well for the tested parameters.

6 The Extended Model

Let all parameters and processes be as in the MIIP, i.e.:

- Let $N := \{1, ..., n\}$ be the set of n observed neurons,
- $1, ..., T$ the indices of the time steps in the observed period of time, $t \in \{1, ..., T\}$.
- For each of the $2^n - 1$ non-empty subsets of neurons $M \subseteq N$, a basic process $P_M(t)$ is introduced.
- All P_M are assumed to be stationary Bernoulli processes with firing probabilities $\lambda_M \in [0, 1)$, and independent.
- Those P_M with $|M| = 1$ represent the independent background processes. They are called B_M ("background").
- All other P_M represent correlation processes. These basic correlation processes are called C_M ("correlation").

The MIIP is then extended to the E-MIIP by introducing additional stationary Bernoulli processes for "coincidences" whose spikes do not fall into the same bin. Those will be called "jittered" coincidences. For a description of the model, we need the following

Definition

Let n be the number of neurons, $j \in \mathbb{N}\backslash\{0\}$ the jitter-width. Let $M \subseteq N$ be a subset of observed neurons, $1 < |M| =: m \leq n$. A *configuration of M under j* is an n-tuple $(x_1, ..., x_n)^{(M,j)}$, where

a) $x_i \in \{0, 1, ..., j, \infty\}\ \forall i = 1, ..., n$

b) $x_i = \infty \quad \Leftrightarrow \quad i \in N - M$

c) $\exists\, i_1 \in M : x_{i_1} = 0$

d) $\exists\, i_2 \in M : x_{i_2} = j$

A configuration can be thought of as a jittered coincidence of specified neurons out of M, where the first and the last spike are exactly j time steps away from each other. Additionally, for every neuron $i \in M$, the exact timing of its spike in the window of length $j + 1$ is specified by x_i. Condition b) means that the neurons that are not part of the subgroup M do not take part in the configuration, because "their spike" is not observed in any finite piece of data.

The number of configurations $c(m, j)$ for m neurons and fixed jitter-size j can be computed recursively:

$$c(m, 0) := 1$$

and

$$c(m, j) = (j + 1)^m - \sum_{i=2}^{j+1} [i \cdot c(m, j - i + 1)] .$$

There are $(j+1)^m$ possibilities for the distribution of the spikes of m neurons in $j+1$ bins. Those configurations with smaller maximal distance between two spikes are part of the configurations with smaller j.

The model-extension is as follows:

- Let J be the maximally allowed "jitter" for a coincidence, n be the number of neurons.
- For every $j \in \{1, ..., J\}$ and $M \subseteq N$ with $|M| > 1$, one stationary Bernoulli process $P_{(x_1,...,x_n)^{(M,j)}}$ with firing probability $\mu_{(x_1,...,x_n)^{(M,j)}}$ is added per configuration $(x_1, ..., x_n)^{(M,j)}$. A success $P_{(x_1,...,x_n)^{(M,j)}}(t) = 1$ at time t implies $O_i(t + x_i) = 1$. Thus all neurons i with $x_i = 0$ fire at time t, whereas the others from the subgroup M fire at specified times in the time window between time t and $t + j$.

- The additional processes belong to the basic processes, all of which are independent.
- The processes $P_{(x_1,\ldots,x_n)^{(M,j)}}(t)$ are defined for $t \in \{-j, T\}$ in order to make sure that the same conditions are valid for the whole data piece.
- Note that for $j = 0$ we recover the processes C_M, $|M| > 1$ discussed in part I.

6.1 n=2

The model will now be presented for $n = 2$. Recall that the MIIP for $n = 2$ contained three basic processes B_1, B_2, C_{12}. The only subgroup M of N with $|M| > 1$ is N itself. We thus get $2J$ additional processes

$$P_{(0,1)^{(N,1)}}; P_{(1,0)^{(N,1)}}; P_{(0,2)^{(N,2)}}; P_{(2,0)^{(N,2)}}; \ldots; P_{(0,J)^{(N,J)}}; P_{(J,0)^{(N,J)}}$$

Each of them produces spikes as shown in figure 20. As per j only two processes are introduced, the firing probabilities μ are denoted shortly as μ_j and μ'_j.

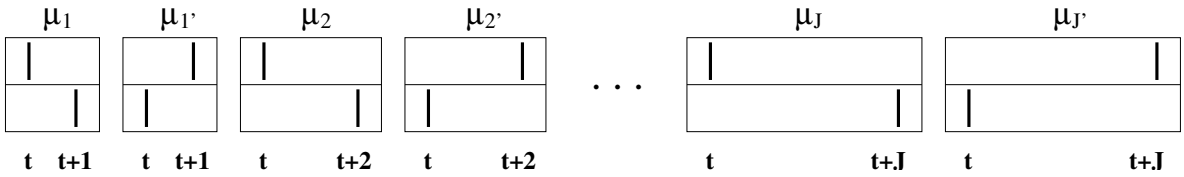

Figure 20: Schematic view of the configurations belonging to all additionally introduced processes. The two rows represent the neurons 1 and 2. A success of process $P_{(x_1,x_2)^{(N,j)}}$ at time t produces the corresponding spike-pattern.

6.1.1 Symmetrical Case

To start with, it will be assumed that the jitter is symmetrical, i.e. $\mu_j = \mu_{j'} \; \forall \, j = 1, \ldots, J$. This is in agreement with experimental findings, as most cross-correlograms are symmetrically shaped. The parameter λ_{12} will be called μ_0, as it produces coincidences with zero time lag.

In the jitter-model, no maximum-likelihood estimation will be used, because the jittered coincidences lead to dependencies between the time steps, such that the maximum-likelihood

estimates of the probabilities are no longer the events' relative frequencies. In the first step, formulas will be developed where the parameters depend on probabilities of data pieces. Then, a discussion of the estimation of those probabilities will lead to the estimates used.

The first step proves to be no more complicated than in the reduced MIIP: With

$$\begin{aligned}
E_j(t) &:= \{O_1(t) = ... = O_1(t+j-1) = O_2(t) = ... = O_2(t+j-1) = 0\}, \\
P_j &:= P(E_j(t)) = [(1-\lambda_1)(1-\lambda_2)(1-\mu_0)]^j \prod_{i=1}^{j-1}(1-\mu_i)^{2(i+j)} \prod_{i=j}^{J}(1-\mu_i)^{4j} \quad (44) \\
P_1 &= (1-\lambda_1)(1-\lambda_2)(1-\mu_0)(1-\mu_1)^4(1-\mu_2)^4 \cdot ... \cdot (1-\mu_J)^4 \quad (45) \\
P_2 &= [(1-\lambda_1)(1-\lambda_2)(1-\mu_0)]^2 (1-\mu_1)^6(1-\mu_2)^8(1-\mu_3)^8 \cdot ... \cdot (1-\mu_J)^8 \quad (46) \\
p_{0+} &:= P(\{O_1(t) = 0\}) = (1-\lambda_1)(1-\mu_0)(1-\mu_1)^2(1-\mu_2)^2 \cdot ... \cdot (1-\mu_J)^2 \quad (47) \\
p_{+0} &:= P(\{O_2(t) = 0\}) = (1-\lambda_2)(1-\mu_0)(1-\mu_1)^2(1-\mu_2)^2 \cdot ... \cdot (1-\mu_J)^2, \quad (48)
\end{aligned}$$

which implies

$$(1-\lambda_1) = \frac{P_{J+1}}{p_{+0}P_J}, \qquad (1-\lambda_2) = \frac{P_{J+1}}{p_{0+}P_J}, \quad (49)$$

$$(1-\mu_0) = \frac{p_{+0}p_{0+}}{p_{00}} \quad (50)$$

$$(1-\mu_1) = \frac{p_{00}}{\sqrt{P_2}} \quad (51)$$

and

$$(1-\mu_j) = \frac{P_j}{\sqrt{P_{j-1}P_{j+1}}}, \quad j = 1, ..., J \text{ and } P_0 := 1. \quad (52)$$

Note that equation (50) corresponds to formula (6) on page 16 in the MIIP.

6.1.1.1 Proof of equation (44)

by induction over j:

$j = 1$:

$$P_1 = (1-\lambda_1)(1-\lambda_2)(1-\mu_0)(1-\mu_1)^4(1-\mu_2)^4 \cdot ... \cdot (1-\mu_J)^4.$$

The first three factors are the same as in the reduced model. For the rest the following holds true: The two processes belonging to each parameter μ_j must not fire at any of the two times t

or $t - j$, which leads to an exponent of 4.

Assume now that for $j \geq 1$ the following holds true:

$$P_j = [(1-\lambda_1)(1-\lambda_2)(1-\mu_0)]^j \prod_{i=1}^{j-1}(1-\mu_i)^{2(i+j)} \prod_{i=j}^{J}(1-\mu_i)^{4j}.$$

Now

$$\begin{aligned} P_{j+1} &= P_j \cdot P(\{O_1(t+j) = O_2(t+j) = 0\}|E_j(t)) \\ &= P_j \cdot (1-\lambda_1)(1-\lambda_2)(1-\mu_0)(1-\mu_1)^2 \cdot ... \cdot (1-\mu_j)^2 \cdot (1-\mu_{j+1})^4 \cdot ... \cdot (1-\mu_J)^4, \end{aligned}$$

because for μ_i, $i \leq j$, the condition that no spike occurred up to j bins before implies that at time $t + j - i$ those processes did not fire. Thus only the time $t + j$ of the two processes belonging to each μ_i needs to be restricted to non-firing. For μ_i, $i > j$, no condition is set for the point $t + j - i$. Thus, per process, this point must as well be restricted to non-firing, leading to an exponent of 4. We thus get

$$P_{j+1} = [(1-\lambda_1)(1-\lambda_2)(1-\mu_0)]^{j+1} \cdot \prod_{i=1}^{j}(1-\mu_i)^{2(i+j+1)} \cdot \prod_{i=j+1}^{J}(1-\mu_i)^{4(j+1)}$$

□

6.1.2 Asymmetrical Model

In an analogous way, formulas can be shown for the general, asymmetrical model for $n = 2$. For this, the following notation is needed:

$$P_{d,j} := P(D_j) := P(\{O_1(t+1) = ... = O_1(t+j-1) = O_2(t) = ... = O_2(t+j-2) = 0\})$$

$$P_{d',j} := P(D'_j) = P(\{O_1(t) = ... = O_1(t+j-2) = O_2(t+1) = ... = O_2(t+j-1) = 0\}).$$

$P_{d,j}$ is the probability to find a piece of data of length j, where all data points are restricted to zero except for the first bin of neuron 1 and the last of neuron 2. The letter 'd' stands for "diagonal". $P_{d',j}$ is the mirrored correspondent of $P_{d,j}$. All formulas from the symmetrical case

need to be changed slightly:

$$P_j = [(1-\lambda_1)(1-\lambda_2)(1-\mu_0)]^j \prod_{i=1}^{j-1}[(1-\mu_i)(1-\mu_i')]^{(i+j)} \prod_{i=j}^{J}[(1-\mu_i)(1-\mu_i')]^{2j} \quad (53)$$

$$P_1 = (1-\lambda_1)(1-\lambda_2)(1-\mu_0)(1-\mu_1)^2(1-\mu_1')^2 \cdot ... \cdot (1-\mu_J)^2(1-\mu_J')^2$$

$$p_{0+} = (1-\lambda_1)(1-\mu_0)(1-\mu_1)(1-\mu_1') \cdot ... \cdot (1-\mu_J)(1-\mu_J')$$

$$p_{+0} = (1-\lambda_2)(1-\mu_0)(1-\mu_1)(1-\mu_1') \cdot ... \cdot (1-\mu_J)(1-\mu_J')$$

$$P_{d,j} = [(1-\lambda_1)(1-\lambda_2)]^{j-1}(1-\mu_0)^j \cdot \cdot \prod_{i=1}^{j-2}(1-\mu_i)^{i+j} \cdot \prod_{i=j-1}^{J}(1-\mu_i)^{2(j-1)} \cdot \prod_{i=1}^{j-1}(1-\mu_i')^{i+j-2} \cdot \prod_{i=j}^{J}(1-\mu_i')^{2(j-1)} \quad (54)$$

For $P_{d',j}$, we need to replace every μ_j by μ_j' and vice versa. From this it follows

$$(1-\lambda_1) = \frac{P_{J+1}}{p_{+0}P_J}, \qquad (1-\lambda_2) = \frac{P_{J+1}}{p_{0+}P_J}, \quad (55)$$

$$(1-\mu_0) = \frac{p_{+0}p_{0+}}{p_{00}}, \quad (56)$$

and

$$(1-\mu_j) = \frac{P_j}{\sqrt{P_{j-1}P_{j+1}}} \cdot \sqrt{\frac{P_{d,j+1}}{P_{d',j+1}}} \cdot \frac{P_{d',j}}{P_{d,j}} \cdot \frac{P_{d,j-1}}{P_{d',j-1}} \cdot \frac{P_{d',j-2}}{P_{d,j-2}} \cdot ... \quad (57)$$

for $j = 1, ..., J$ and $P_0 = P_{d,1} = P_{d',1} = P_{d,2} = P_{d',2} = 1$. Equations (55) and (56) are the same as in the symmetrical case. Equation (57) follows from equation (54) and its analogous for $P_{d',j}$. Due to symmetry reasons, only equation (54) will be proven. Equation (53) follows directly from the symmetrical case and the corresponding equation.

6.1.2.1 Proof of equation (54)

by induction over j.
$j = 2$:

$$P_{d,2} = (1-\lambda_1)(1-\lambda_2)(1-\mu_0)^2(1-\mu_1)^2(1-\mu_1')(1-\mu_2)^2(1-\mu_2')^2 \cdot ... \cdot (1-\mu_J)^2(1-\mu_J')^2,$$

because no jitter-process may produce spikes in $O_1(t+1)$ and in $O_2(t)$. Only for the one with parameter μ_1', this leads to only one restricted firing-time.

Let equation (54) be true for any $j > 1$. It holds

$$P(D_{j+1}) = P(D_j) \cdot P(H_j|D_j) \cdot P(D_{j+1}|H_j),$$

where $H_j := \{O_1(t+1) = ... = O_1(t+j-1) = O_2(t) = ... = O_2(t+j-1) = 0\}$ describes the event that out of j adjacent bins, all except the first of neuron one are restricted to zero (see table 2).

$H_j \mid D_j$			
t			t+j
+	0	···	0
0	0	···	0|+

$D_{j+1} \mid H_j$				
t			t+j	t+j +1
+	0	···	0	0
0	0	···	0	+

Table 2: Visualization of the events needed for the proof. The black letters represent given data from the conditioning event, whereas the green letters stand for the conditioned event.

$$\begin{aligned} P(H_j|D_j) &= (1-\lambda_2)(1-\mu_{j-1}) \cdot ... \cdot (1-\mu_J)(1-\mu_1') \cdot ... \cdot (1-\mu_J') \\ P(D_{j+1}|H_j) &= (1-\lambda_1)(1-\mu_0)(1-\mu_1) \cdot ... \cdot (1-\mu_J)(1-\mu_{j+1}') \cdot ... \cdot (1-\mu_J') \end{aligned}$$

Thus

$$\begin{aligned} P_{d,j+1} &= P_{d,j} \cdot (1-\lambda_1)(1-\lambda_2)(1-\mu_0) \cdot \\ &\quad \cdot (1-\mu_1) \cdot ... \cdot (1-\mu_{j-2})(1-\mu_{j-1})^2 \cdot ... \cdot (1-\mu_J)^2 \cdot \\ &\quad \cdot (1-\mu_1') \cdot ... \cdot (1-\mu_j')(1-\mu_{j+1}')^2 \cdot ... \cdot (1-\mu_J')^2 \\ &= [(1-\lambda_1)(1-\lambda_2)]^j (1-\mu_0)^{j+1} \cdot \\ &\quad \cdot \prod_{i=1}^{j-1} (1-\mu_i)^{i+j+1} \prod_{i=j}^{J} (1-\mu_i)^{2j} \prod_{i=1}^{j} (1-\mu_i')^{i+j-1} \prod_{i=j+1}^{J} (1-\mu_i')^{2j} \end{aligned}$$

□

Now the used probabilities of data pieces need to be estimated, which will be presented in section 7. Before, the model's extensibility onto more than two neurons will be demonstrated.

6.2 n=3

In this subsection, some considerations will be made to show that the model is extensible onto more than two neurons. In its most general form, the increase in the number of parameters is not very practical to handle. Still, as in the last section, one can often reduce the number of parameters considerably. Figure 21 shows all possible configurations for three neurons and maximal introduced jitter $J = 1$. In the figure, the parameters for the same subsets of neurons and constant jitter-size are set equal.

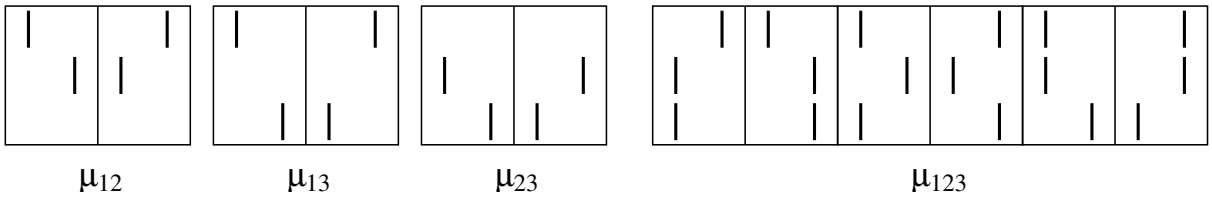

Figure 21: Schematic view of the configurations belonging to all jitter-processes for $n = 3$ and $J = 1$. Equality of the firing probabilities for fixed M and j is assumed.

In the model given in figure 21, one can easily compute corresponding formulas for the parameters and the probabilities of data pieces.

Let

$$\begin{aligned} P_{j_1 j_2 j_3}(t) &:= P(\{O_i(t) = j_i,\ i \in \{i | j_i \neq' +'\}\}), \\ P_{j_1 j_2 j_3, k_1 k_2 k_3}(t) &:= P(\{O_i(t) = j_i,\ i \in \{i | j_i \neq' +'\},\ O_i(t+1) = k_i,\ i \in \{i | k_i \neq' +'\}\}) \end{aligned}$$

Table 3 visualizes some examples of this notation.

	$P_{000}(t)$	$P_{++0}(t)$	$P_{000,+00}(t)$	$P_{+00,0+0}(t)$
$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} + \\ + \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 & + \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} + & 0 \\ 0 & + \\ 0 & 0 \end{pmatrix}$
Time	t	t	t t+1	t t+1

Table 3: Examples of $P_{j_1 j_2 j_3}$ and $P_{j_1 j_2 j_3, k_1 k_2 k_3}$. The notations from above are the probabilities of the data pieces from below. A '+' means that this data point can contain both a spike or a non-spike, i.e. it is not restricted.

One thus gets

$$\begin{aligned}
1-\lambda_1 &= \frac{P_{000,000}^2 P_{+00}}{P_{000,+00}^2 P_{000}} \\
1-\lambda_{12} &= \frac{P_{000} P_{++0} P_{000,+00}^2 P_{000,0+0}^2}{P_{+00} P_{0+0} P_{000,000}^2 P_{000,++0}^2} \\
1-\lambda_{123} &= \frac{P_{++0} P_{+0+} P_{0++} P_{000}}{P_{+00} P_{0+0} P_{00+}} \\
1-\mu_{12} &= \frac{P_{000,+00} P_{000,0+0}}{P_{000,000} P_{+00,0+0}} \\
1-\mu_{123} &= \frac{P_{0+0} P_{00+} P_{000,000} P_{000,0++}}{P_{0++} P_{000} P_{000,0+0} P_{000,00+}}
\end{aligned}$$

Thus for $n = 3$, $J = 1$ and symmetry, one can estimate the underlying parameters λ_1, λ_2, λ_3, λ_{12}, λ_{13}, λ_{23}, λ_{123}, μ_{12}, μ_{13}, μ_{23}, and μ_{123} with the help of the formulas derived. This shows that the model is extensible onto $n > 2$. We will now restrict the discussion onto $n = 2$ and symmetry and go on with the estimation of the probabilities.

7 Some Estimates

In the last section, formulas have been developed that express the model's parameters in terms of probabilities of data pieces. For the MIIP from part I that only contained exact coincidences, the maximum-likelihood estimates of all required probabilities were the relative frequencies of the corresponding events. This is no longer the case due to the introduced dependencies between

time steps. It is thus necessary to evaluate other methods of estimation of the probabilities of data pieces as well as the parameters.

The considerations will be reduced to a small range of parameters to show principal properties. A discussion of the estimation of the required probabilities will be presented for $n = 2$, $J = 1$, $\lambda_1 = \lambda_2$ and symmetry of jitter: $\mu_1 = \mu_1'$.

7.1 Estimating Probabilities of Data Pieces

In the previously defined model, the parameters λ_1, μ_0 and μ_1 need to be estimated. The following probabilities play a role: $p_{00} = P_1, p_{+0}, p_{0+}$ and P_2 (defined in equation (46) on page 67). Exemplarily for all probabilities of data pieces of length one and two, P_1 and P_2 will be estimated here.

7.1.1 Estimation of P_1

7.1.1.1 Disjoint Intervals As in the model without jitter, one could in principle count the number of bins where no spike occurred in O_1 nor in O_2. This estimation is subject to two influences:

1. Many data points contribute to the estimation, reducing its variance.
2. Dependencies between adjacent bins increase the variance of the estimation, as the probability to get no spikes in one bin is increased if no spike occurred in the preceding one.

Let now I_t be the indicator function for the event $\{O_1(t) = O_2(t) = 0\}$, meaning that

$$I_t = \begin{cases} 1 & \text{if } O_1(t) = O_2(t) = 0 \\ 0 & \text{else} \end{cases}.$$

Then the variance of the estimate

$$\widehat{P}_{1,d} := \frac{1}{T} \sum_{i=1}^{T} I_i$$

of P_1 is

$$\begin{aligned}\sigma^2_{\widehat{P}_{1,d}} &:= \text{Var}\left(\frac{1}{T}\sum_{i=1}^{T} I_i\right) = \frac{1}{T^2}\left\{\sum_{i=1}^{T}\text{Var}I_i + 2\sum_{i=1}^{T-1}\text{Cov}(I_i, I_{i+1})\right\} \\ &= \frac{1}{T^2}\left\{TP_1(1-P_1) + 2(T-1)(P_2 - P_1^2)\right\},\end{aligned}$$

because for $J = 1$, only adjacent bins are correlated.

7.1.1.2 Independent Intervals A method that could be favored if there was a truly large number of data points is the use of independent intervals. For $J = 1$, only adjacent bins are not independent. One could thus use only every second bin for the estimation. This eliminates the variance-increasing factor of interdependencies but increases the variance due to the fact that only half of the intervals are used.

The variance of this estimate $\widehat{P}_{1,i}$ is (for T a multiple of 2):

$$\text{Var}(\widehat{P}_{1,i}) = \frac{P_1(1-P_1)}{T/2} =: \sigma^2_{\widehat{P}_{1,i}}$$

Remark

For $0 \leq \mu_1 \leq 0.01$: $\sigma^2_{\widehat{P}_{1,d}} \leq \sigma^2_{\widehat{P}_{1,i}}$, which implies that the use of disjoint intervals always leads to a smaller variance of the estimate.
Indeed:

$$\begin{aligned}\frac{2P_1(1-P_1)}{T} &\geq \frac{1}{T^2}\left\{TP_1(1-P_1) + 2(T-1)(P_2 - P_1^2)\right\} \\ \iff \quad P_1(1-P_1) &\geq \frac{2(T-1)(P_2 - P_1^2)}{T} \\ \iff \quad P_1 &\leq \frac{T(1-\mu_1)^2}{T(1-\mu_1)^2 + 2(T-1)(1-(1-\mu_1)^2)}\end{aligned}$$

with $z := (1-\mu_1)^2 \in [0.9801, 1]$ and $P_1 = (1-\lambda_1)(1-\lambda_2)(1-\mu_0)(1-\mu_1)^4 \leq (1-\mu_1)^4$ it

remains to be shown

$$\begin{aligned} & & z^2 &\leq \frac{Tz}{Tz + 2(T-1)(1-z)} \\ \iff & Tz^2 + 2(T-1)z(1-z) - T &\leq 0 \\ \iff & T(z-1)^2 + 2z(1-z) &\geq 0 \end{aligned}$$

which is true, because $1 - z \geq 0$ and $(z-1)^2 \geq 0$. □

7.1.2 Estimation of P_2

7.1.2.1 Overlapping Intervals For the estimation of P_2, whose corresponding event contains two adjacent bins, there is another obvious possibility: the use of overlapping intervals. Let I_t be the indicator function for the event $\{O_1(t) = O_2(t) = O_1(t+1) = O_2(t+1) = 0\}$, meaning that

$$I_t = \begin{cases} 1 & \text{if } O_1(t) = O_2(t) = O_1(t+1) = O_2(t+1) = 0 \\ 0 & \text{else} \end{cases} . \tag{58}$$

The estimation

$$\widehat{P}_{2,o} := \frac{1}{T-1} \sum_{i=1}^{T-1} I_i$$

contains $T - 1$ data points, which keeps the variance small, but does not account for dependencies between (near-)adjacent indicator variables. The estimate's variance is the following:

$$\begin{aligned} \sigma^2_{\widehat{P}_{2,o}} &:= \text{Var}\left(\frac{1}{T-1} \sum_{i=1}^{T-1} I_i \right) \\ &= \frac{1}{(T-1)^2} \left\{ \sum_{i=1}^{T-1} \text{Var} I_i + 2 \sum_{i=1}^{T-2} \text{Cov}(I_i, I_{i+1}) + 2 \sum_{i=1}^{T-3} \text{Cov}(I_i, I_{i+2}) \right\} \\ &= \frac{1}{(T-1)^2} \left\{ (T-1)P_2(1-P_2) + 2(T-2)(P_3 - P_2^2) + 2(T-3)(P_4 - P_2^2) \right\} \end{aligned}$$

7.1.2.2 Disjoint Intervals The use of disjoint intervals on the one hand reduces the variance of the estimate by accounting for a part of the interdependencies and on the other enhances it by reducing the number of contributing data points. Let I_t be as in equation (58). Then with

$$\widehat{P}_{2,d} := \frac{2}{T} \sum_{i=1, i \text{ odd}}^{T-1} I_i$$

(for T a multiple of 2)

$$\begin{aligned} \sigma^2_{\widehat{P}_{2,d}} &:= \text{Var}\left(\frac{2}{T} \sum_{i=1, i \text{ odd}}^{T-1} I_i\right) \\ &= \frac{4}{T^2}\left\{\sum_{i=1, i \text{ odd}}^{T-1} \text{Var} I_i + 2 \sum_{i=1, i \text{ odd}}^{T-3} \text{Cov}(I_i, I_{i+2})\right\} \\ &= \frac{4}{T^2}\left\{\frac{T}{2} P_2(1-P_2) + 2(\frac{T}{2}-1)(P_4 - P_2^2)\right\} \end{aligned}$$

7.1.2.3 Independent Intervals Independent intervals seem at first to be the most clear way, because all possible dependencies are eliminated. But of course, this is done at the expense of the number of used data points. For

$$\widehat{P}_{2,i} := \frac{3}{T} \sum_{i=1, i \equiv 1 \mod 3}^{T-1} I_i$$

(for T a multiple of 3)

$$\text{Var}\left(\frac{3}{T} \sum_{i=1, 3|(i-1)}^{T-1} I_i\right) = \frac{P_2(1-P_2)}{T/3} =: \sigma^2_{\widehat{P}_{2,i}}$$

Remark

As for P_1: $\sigma^2_{\widehat{P}_{2,i}} \geq \sigma^2_{\widehat{P}_{2,d}} \geq \sigma^2_{\widehat{P}_{2,o}}$ holds in usual parameter ranges. That means that the dependencies between adjacent bins are so small that the change in the number of contributing data

points is of higher importance.

- Proof of $\sigma^2_{\widehat{P}_{2,i}} \geq \sigma^2_{\widehat{P}_{2,d}}$ (in usual parameter ranges):

$$\begin{aligned}
\frac{4}{T^2}\left\{\frac{T}{2}P_2(1-P_2)+2(\frac{T}{2}-1)(P_4-P_2^2)\right\} &\leq \frac{3P_2(1-P_2)}{T} \\
\iff 4(T-2)(P_4-P_2^2) &\leq TP_2(1-P_2) \\
\iff 4(T-2)P_2(\frac{1}{(1-\mu_1)^2}-1) &\leq T(1-P_2) \\
\iff P_2 &\leq \frac{T(1-\mu_1)^2}{T(1-\mu_1)^2+4(T-2)(1-(1-\mu_1)^2)}
\end{aligned}$$

  It remains to show

$$z^4 \leq \frac{Tz}{Tz+4(T-2)(1-z)}$$

  with $z := (1-\mu_1)^2 \in [0.9801, 1]$ and $\mu_0 \geq \mu_1$, thus
  $P_2 = (1-\lambda_1)^2(1-\lambda_2)^2(1-\mu_0)^2(1-\mu_1)^6 \leq (1-\mu_1)^8$. The assumption that there are at least as many exact coincidences as coincidences jittered to one direction by one bin is no strong restriction, as in almost all cross-correlograms, center-peaks are observed.

$$\iff \quad T(3z^4-4z^3+1)+8z^3(1-z) \geq 0$$

  The function $f(x) = 3x^4 - 4x^3 + 1$ on $\mathbb{R}$ only disappears at $x = 1$. This is a local minimum, as $f'(1) = 12 - 12 = 0$ and $f''(1) = 36 - 24 > 0$. Thus
  $T(3z^4 - 4z^3 + 1) + 8z^3(1-z) \geq 0$ with "=" $\Leftrightarrow\ z = 1\ \Leftrightarrow\ \mu_1 = 0$. □

- As for the comparison of the estimation with overlapping and disjoint intervals, it can not be generally shown that $\sigma^2_{\widehat{P}_{2,d}} \geq \sigma^2_{\widehat{P}_{2,o}}$. This depends on all used parameters, including the background and the zero-jitter-coincidence probability. For our considerations, $\mu_0 = \mu_1$ and $\lambda_1 = \lambda_2$ shall be assumed, as this is especially important for the further sections. In figure 22, one can see those parameter ranges with $\sigma^2_{\widehat{P}_{2,d}} \geq \sigma^2_{\widehat{P}_{2,o}}$. The values for T say which minimal T is sufficient for $\sigma^2_{\widehat{P}_{2,d}} \geq \sigma^2_{\widehat{P}_{2,o}}$ in the area with the same color. E.g., in the light blue area, $T \geq 20$ is sufficient for $\sigma^2_{\widehat{P}_{2,d}} \geq \sigma^2_{\widehat{P}_{2,o}}$. Thus the increase in variance

that originates in the dependencies between overlapping intervals must be compensated for over a sufficiently large time period. As the considerations made here concentrate on asymptotic behavior, T is assumed to be large. Thus $\sigma^2_{\widehat{P}_{2,d}} \geq \sigma^2_{\widehat{P}_{2,o}}$ holds true "almost always".

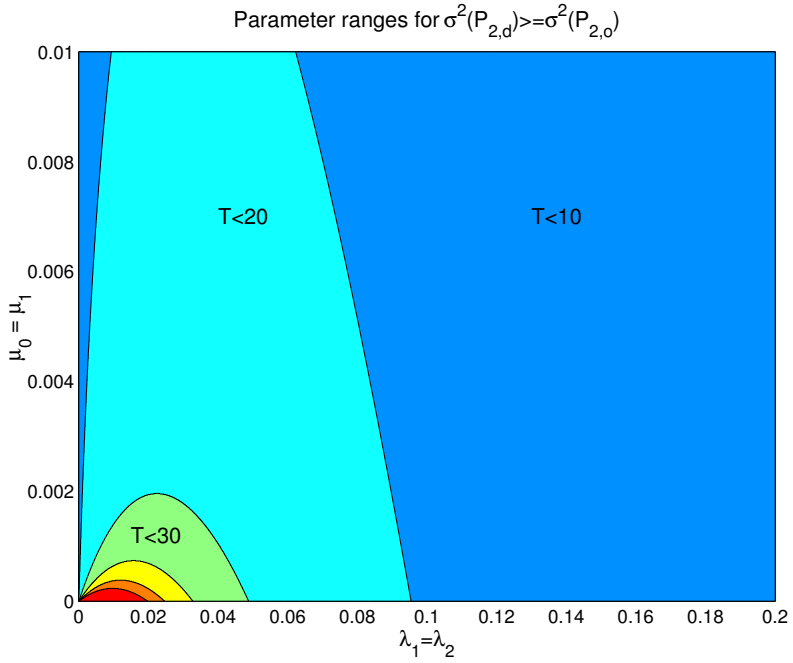

Figure 22: Parameter ranges with $\sigma^2_{\widehat{P}_{2,d}} \geq \sigma^2_{\widehat{P}_{2,o}}$ for minimal T as indicated. Yellow area: $T = 40$, Orange area: $T = 50$. For the red area, the required T is larger.

7.1.3 Implications

What has been shown is that for $n = 2$, $J = 1$ and symmetry, in usual parameter ranges

- P_1 can be estimated with smaller variance when using disjoint intervals, and
- P_2 can be estimated with smaller variance when using overlapping intervals.

In the following subsections, some simulations will be shown, all of which are restricted to $n = 2$, $J \leq 5$, and symmetry. For all estimated probabilities, the estimation with the higher

number of contributing data points (i.e. overlapping or disjoint, respectively) will be used without proving its advantages. Still, as the coincidence firing probabilities and J are very small, it seems plausible that the interdependencies between adjacent data points remain small.

7.2 Estimating the Parameters

Now the parameters from section 6.1 can be estimated with the formulas (49) to (52) on page 67. Only the formula for the background rates depends on J:

$$(1-\lambda_1) = \frac{P_{J+1}}{p_{+0}P_J}, \qquad (1-\lambda_2) = \frac{P_{J+1}}{p_{0+}P_J}$$

This is a problem, as the real maximal jitter J in the data is not known. So the estimates will be

$$(1-\lambda_1) = \frac{P_{a+1}}{p_{+0}P_a}, \qquad (1-\lambda_2) = \frac{P_{a+1}}{p_{0+}P_a}. \tag{59}$$

The letter 'a' means that for the computation, the real J in the data is 'assumed' to be a. The formula is still right when using $a > J$, only the variance of the estimation increases. On the other hand, the use of $a < J$ does not account for the existing parameters $\mu_{a+1}, ..., \mu_J$ and thus leads to a biased estimate, meanwhile keeping its variance low. In figure 23, the estimation methods for λ_1 are evaluated with the help of the standard error of the mean. Per data point, 10000 trials of length $T = 10000$ were simulated. The average deviation of the estimate from the real parameter was then computed. The x-axis represents the parameter J. Both background rates were set equal, and the overall coincidence rate was kept at $\mu := \mu_0 + 2\sum_{i=1}^{J}\mu_i = 0.006$, with $\mu_0 = \mu_1 = ... = \mu_J$. For each curve, the used a to compute the background rate is constant. One can see that for constant J

- the error is minimal for $a = J$,
- the use of $a < J$ produces a much larger error than $a > J$ for any of the plotted combinations.

Hence, the rise in variance for $a > J$ is outweighed by the bias for $a < J$. It is thus proposed to use the maximal plausible a.

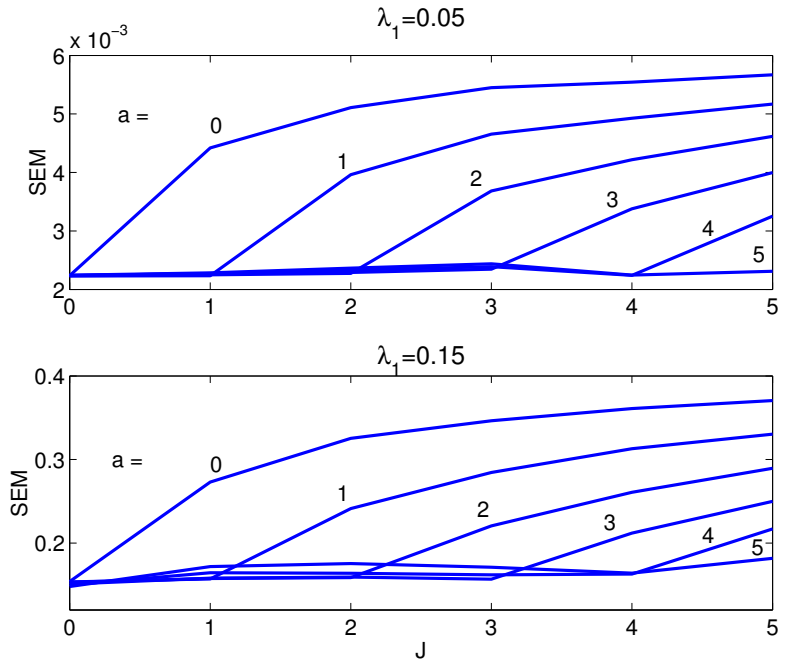

Figure 23: Standard error of the mean of the estimation of λ_1. Per curve, one fixed a has been used in formula (59). On the x-axis, the real J in the data is plotted. Per data point, 10000 trials of length $T = 10000$ were simulated. $\mu := \mu_0 + 2\sum_{i=1}^{J} \mu_i := 0.006$, $\lambda_1 = \lambda_2$.

8 Test Statistics and Underlying Model

After having estimated the model's parameters, one needs to decide whether $H_0 : \mu_0 + \sum_{i=1}^{J} \mu_i = 0$ is kept or whether there is enough evidence to reject H_0 in favor of a hypothesis that includes (near-)coincident firing. For the exemplary computations made here, the model is further specified by assuming $\mu_0 = \mu_1 = ... = \mu_J$. This would lead to a rectangular-shaped cross-correlogram. One could as well assume any other interdependency or even keep the full generality of parameters.

The proposed test statistics is

$$Z'_a := S_a / \sigma'_{S_a}$$

with

$$S_a := \widehat{\mu}_0 + \sum_{i=1}^{a} \widehat{\mu}_i$$

Again, the letter 'a' represents the assumed value of the parameter J. The variance $(\sigma'_{S_a})^2$ can be derived from simulations for the estimated parameters. For the discussion of this method's significance level and test power, the variance per parameter set will be derived before in a separate step in order to reduce the computational effort: In 10000 trials with $\mu_0 = \mu_1 = ... = \mu_J$ and all other parameters as specified, the variance of S is estimated empirically. Thus the used test statistics is modified to

$$Z_a := S_a / \sigma_{S_a}$$

where $\sigma^2_{S_a}$ is the variance of $\widehat{\mu}_0 + \sum_{i=1}^{a} \widehat{\mu}_i$ under the real parameters $\lambda_1, \lambda_2, \mu_0, \mu_1$ and J.

To present this more clearly, one can concentrate on the two values J and a.

- If $J = a$, then the analysis uses the 'right' test statistics, meaning that all positive parameters μ_j are included in the test statistics. It will turn out that this leads to the highest test power for all a under fixed J.
- If $J < a$, then all parameters $\mu_{J+1}, ..., \mu_a$ are zero. But as all estimates are subject to fluctuations, the estimates of these parameters will influence the test statistics and thus reduce the test power.
- If $J > a$, then the positive parameters $\mu_{a+1}, ..., \mu_J$ are not included into the test statistics, whose value is thus reduced.

Before the examination of the significance level, remark that all test statistics Z_a have asymptotic normal distribution:

Note that to estimate P_{J+1}, intervals of length $J + 1$ need to be examined. The use of overlapping intervals for the estimation yields a distance of $J + 1 + (J - 1) = 2J$ to grant independence of the corresponding intervals. All other probabilities that need to be estimated use smaller intervals. This implies that all S_a are functions of indicator variables I_t that are at most $2J$-dependent in the sense that $\forall n > 2J$ and $\forall l \geq 0$: $(I_1, ..., I_k)$ and $(I_{k+n}, ..., I_{k+n+l})$ are independent. By centering the indicator variables, one can apply a generalization of the classical

central limit theorem which can be found e.g. in Billingsley (1986, p. 376). It says that for an m-dependent and stationary sequence of bounded random variables X_n with $E[X_n] = 0$ and $S_n = X_1 + ... + X_n$, if $\sigma(S_n) > 0 \Longrightarrow \frac{S_n}{\sigma\sqrt{n}} \longrightarrow \mathcal{N}(0,1)$.

Thus, all estimates of probabilities can be shown to have asymptotic normal distribution. The δ-method then implies asymptotic normal distribution of the Z_a. We will therefore use $z^* = 1.96$ and compare the empirical significance level with the asymptotic $\alpha = 0.025$.

8.1 Significance Level

The new aspect in the jitter-model is the variety of test statistics Z_a, $a = 0, 1, 2, ...$ on the one hand and the number of possible underlying J, $J = 0, 1, 2, ...$ in the data on the other. In the following subsection, we will therefore write $Z_{a,J}$ for the test statistics Z_a when applied onto data with underlying parameter J. This is not necessary in this subsection, because for the significance level all coincidence firing probabilities are zero and thus the two processes are independent and we do not need to distinguish between different underlying J.

Whereas in part I, only one test statistics needed to be investigated, there are now many possibilities to choose a for the derivation of Z_a. At first we need to make sure that all of them lead to a reasonable significance level in order to control the probability to falsely reject H_0. For this, the significance level for a typical set of parameters is shown in figure 24 (upper part) for six different test statistics Z_a, $a = 0, 1, ..., 5$. The relative number out of 10000 trials with $Z_a > 1.96$ is plotted. For $a = 0$, the variance was computed per trial with the asymptotic formula from part I. The plotted significance level corresponds to the empirical value for $T = 10000$ at $\lambda_1 = \lambda_2 = 0.05$ in figure 3 on page 27. One can see that for the chosen parameters, all test statistics for $a > 0$ lead to a significance level of about 0.025 when compared to $z = 1.96$, while most of the tests stay conservative.

8.2 Test Power

To evaluate the test power of the different test statistics, different models with varying J must be considered additionally, which leads to all possible combinations for $Z_{a,J}$, $a = 0, 1, 2, ...$, $J = 0, 1, 2, ...$ The following parameter set is considered exemplarily: $n = 2$, $J, a \in \{0, 1, ..., 5\}$, $\lambda_1 = \lambda_2 = 0.05$, $\mu_0 + \sum_{i=1}^{J} \mu_i = 0.006$, $\mu_0 = \mu_1 = ... = \mu_J$, $T = 10000$. In the lower part of

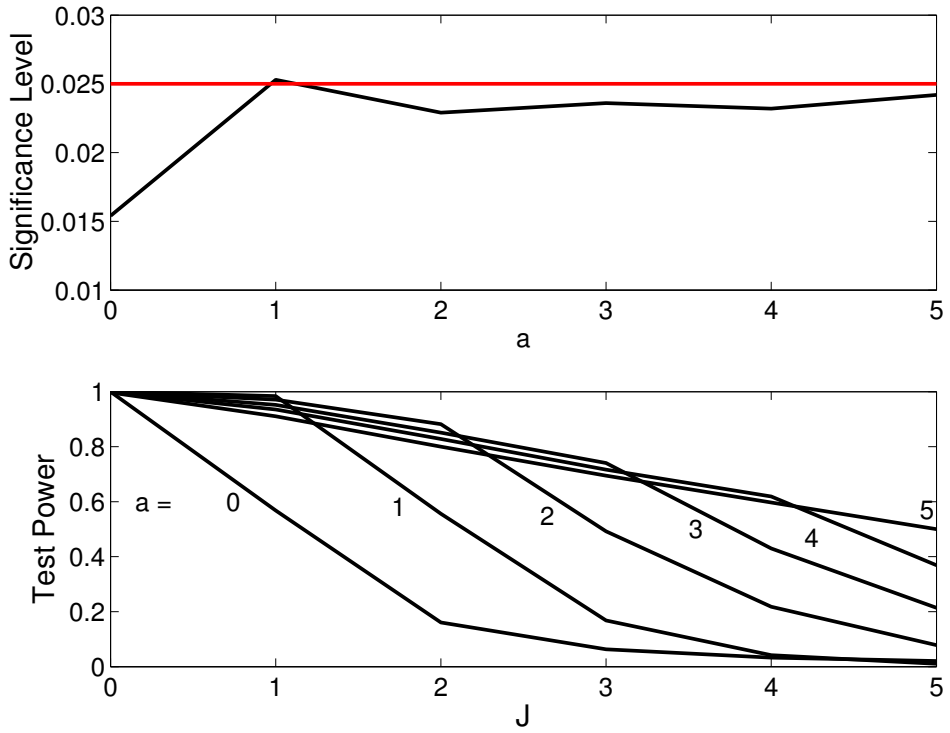

Figure 24: Evaluation of test statistics for $\lambda_1 = \lambda_2 = 0.05$ and $T = 10000$. Per data point, 10000 trials were simulated. The relative number of trials with $Z_a > 1.96$ is plotted. $\sigma^2_{S_a}$ was previously derived from simulations (in 10000 trials). For $a = 0$, the variance was computed per trial with the asymptotic formula from part I. Upper part: empirical significance level for different test statistics Z_a, $a = 0, ..., 5$. Below: empirical test power for $\mu := \mu_0 + 2\sum_{i=1}^{J} \mu_i = 0.006$, $\mu_0 = ... = \mu_J$ and varying maximal introduced jitter J. Every curve represents the test power for one special test statistics with constant a.

figure 24, the test power of the different tests is plotted depending on the J used to simulate the spike trains. Per curve, the test statistics $Z_{a,J}$ uses constant a. One can see that

- the maximal possible test power decreases with growing J. This seems plausible, because the overall coincidence rate, which stays the same, is "broadened" over a wider range.
- Moreover, for constant J, the test power is maximal for $a = J$. This is not astonishing, as the use of smaller a leaves out some positive μ_j in S_a, which reduces the test statistics. Using a larger a enhances the variance σ_{S_a}, because more parameters are included

in S_a. This reduces the test statistics. Analogous results can be found in Grün et al. (1999), where two different methods were examined to detect jittered coincidences and their coincidence width in a model for two parallel processes.

- Still, the use of $a > J$ leads to a smaller decrease in test power than for $a < J$. This seems plausible, because the variance of S_a should not increase considerably with a for $a \geq J$, because all μ_j with $j > J$ are zero. Only the fact that their estimates are included into S_a slightly enhances σ_{S_a}.

9 Conclusions of Part II

Note that this section was not meant to be a complete presentation and discussion of the extended model and test. Instead, exemplary considerations should show its usefulness and point towards further required research. The results are thus presented together with a critical evaluation and implications for future work.

1. The MIIP from part I has been extended in order to allow for coincidences that are jittered in time up to a delay of a few milliseconds. This is done by including one additional independent and stationary Bernoulli process per subgroup of neurons M ($|M| > 1$), jitter $j \leq J$ and configuration. The extension is applicable for more than two parallel processes. In its most general form, it contains a large number of parameters that can be reduced by further assumptions originating in experimental findings.

2. The method of moments was used to develop formulas for the estimation of the parameters for $\{n = 2\}$ and $\{n = 3, J = 1\}$. These formulas use probabilities of data pieces.

3. The maximum-likelihood estimates of the probabilities are no longer the events' relative frequencies. For $n = 2$, $J = 1$, and symmetry, a discussion of the estimation of P_1 and P_2 showed that the estimation with the higher number of contributing data points (i.e. overlapping or disjoint, respectively) leads to a smaller variance as compared to the estimation with the smaller interdependencies between the data pieces used for the estimation (for usual parameter ranges). This motivated the use of the former estimation

method for all other probabilities, although it might be possible that for large J and large jittered coincidence probability, this leads to a higher variance.

4. The only parameter whose estimation depends on J is the background firing probability. An exemplary study showed for typical parameter-values that the rise in variance for $a > J$ is outweighed by the bias for $a < J$, such that the maximal plausible a should be used to estimate λ_1 and λ_2. We believe that this behavior also carries over to the non-examined parameter values in the parameter range $(0, 0.2]$ for the background rates.

5. A test was developed to decide whether coincidences have been introduced at all. The used test statistics is
$$Z_a = \frac{S_a}{\sigma_{S_a}}$$
with
$$S_a = \widehat{\mu}_0 + \sum_{i=1}^{a} \widehat{\mu}_i,$$
$a = 0, 1, ..., 5$. This has been done for the parameters: $n = 2$, $J \leq 5$, $\mu_0 = \mu_1 = ... = \mu_J$, $\mu := \mu_0 + 2\sum_{i=1}^{J} \mu_i$, $\mu \in \{0, 0.006\}$, $T = 10000$, $\lambda_1 = \lambda_2 = 0.05$, and $z^* = 1.96$ as threshold value for the rejection of H_0. The significance level was satisfying in the sense that it was very close to the pre-defined $\alpha = 0.025$. In order to reduce the computational effort, the variance used for the test statistics has been derived before for the same parameters. Thus the variance is assumed to be known, which does not happen in the experiment where one has to deal with the results of the estimation. The results show that the test power is maximal for $a = J$ and that the use of a smaller a leads to a considerably larger decrease in test power than the application of higher a. It thus seems to be reasonable to use a large but plausible a for the test statistics.

To sum up the above statements: The extended model is useful for the analysis of jittered coincidences. It allows to conceptually discriminate between non-stationarity and jittered coincidences. The proposed test has a significance level of about 2.5% for the tested parameters. The model in its most general form opens up a wide range of possible estimation methods and tests that cannot be fully discussed in this work. General considerations need to be made in order to restrict the parameter range and to reduce the number of possible hypotheses.

Part III

Discussion

In this work we presented a model for the analysis of parallel stationary binary processes. It was developed to very directly represent "genuine" interactions, expressed by coincident activity, between subgroups of neurons that cannot be reduced to subgroup interactions. The introduced Model of Independent Interaction Processes (MIIP) for exact coincidences contains one parameter per subgroup of neurons that indicates whether or not this subset tends to fire in synchrony as a consequence of a genuine correlation between its members. The formulas derived for the maximum-likelihood estimation of those parameters allow to analyze all subgroup interactions independently from each other. One can thus identify any combination of correlations that is possible among all specified neuronal subsets. As all used estimates have asymptotic normal distribution with mean zero under the hypothesis that the correlation of interest does not exist, one can build up asymptotically standard normally distributed test statistics by dividing the estimates by their standard deviation.

The test was applied onto sets of two and three parallel processes for different parameter sets. With a chosen time scale of one millisecond per time step, background firing rates were inspected up to 200 Hz, and coincidence firing rates up to 10 Hz, which is thought to be the physiologically relevant range of frequencies (White et al., 1998). The use of a smaller time scale leads to an increase of time steps and a decrease in firing probabilities per time step, which does not impose any problems, because the results are valid for all studied background firing probabilities up to 0.2 (corresponding to 200 Hz for b=1ms). The application of a larger time scale implies a possible increase in firing probability up to values not inspected in our analysis. Secondly, it poses binning problems, of which the increased probability for two spikes to fall into one bin is only one. Further binning problems will be discussed in a later paragraph.

Following experimental evidence that the timing accuracy of spikes can amount to about 5 ms (Abeles et al., 1993; Riehle et al., 1997), the model was extended. The extended E-MIIP additionally includes several processes that represent near-coincident activity of a given jitter for all possible subgroups of neurons. One can directly estimate the parameters and thus get a means to analyze the given spike trains concerning near-coincident activity up to the maximal

jitter J of interest. Formulas for the parameters were presented for two parallel processes and for symmetrical jitter for three neurons and $J = 1$. Also with the help of exemplary considerations we argue in favor of the use of overlapping intervals (instead of disjunct or independent ones) for the estimation of the parameters and in favor of the application of the maximal plausible J for the estimation of the background firing probabilities. For typical parameter sets in the case of two neurons, the performance of the proposed test statistics was evaluated for a model where the probability of any coincidence of jitter j up to $J = 5$ was set equal. We suggest to use the following procedure when applying the model onto experimental data: When given parallel, stationary binary processes, one should at first ask for the maximal jitter J that can be supposed to be given by the data. This can be derived from either general considerations and other experimental findings as e.g. in Abeles et al. (1993) or Riehle et al. (1997), whose results indicate highly precise (1-5 ms) coincident spiking activity between simultaneously recorded neurons, or from a measure for the width of the peak of the data's cross-correlograms as e.g. in Toyama et al. (1981b) or Nelson et al. (1992). The chosen J should be "large enough" in the sense that the effect when J is chosen too large is negligible when compared to the application of a J that is smaller than in the data. With the given J one can estimate all probabilities of coincidences up to the jitter J with the help of the formulas derived. We assume that the direct formulas presented exemplarily carry over to higher numbers of neurons and larger J. We propose to use the sum of all coincidence probabilities divided by its standard deviation as test statistics. One can apply the δ-method (e.g. Bishop et al., 1991) or derive the standard deviation from simulations that use the estimated parameters.

Concerning the application of the analysis method onto experimental data, there are several limitations. First, the model was developed to deal with stationary data. The performance of the test when applied onto non-stationary data has not been discussed. Moreover binning problems arise. The model works on the basis of (discrete) Bernoulli processes. However, the original spikes that are emitted by the neurons exist on a continuous time scale that is divided artificially into bins of a given length b. Thus a part of all "exact coincident" spikes with delay $d < b$ is detected as coincident, and the rest is divided into two adjacent bins (Grün et al., 1999). The analogous is true for near-coincident spikes with jitter $j < J$. This reduces the number of found exact coincidences.

The MIIP for the analysis of exact coincidences is highly related to approaches that use

log-linear models for binary random variables to characterize the interaction structure between n neurons (Martignon, von Hasseln, Grün, Aertsen and Palm (1995); Martignon, Deco, Lasky, Diamond, Freiwald, Vaadia (2000); Amari, 2001). They are based on the same assumptions as independence of adjacent bins and stationarity of the binary processes. Like the MIIP, they can distinguish between the correlations of all neuronal subsets. But in contrast to the MIIP, they are not directly extensible for the analysis of near-coincident spikes, because the requirement of independence of adjacent bins is essential. Furthermore, log-linear models apply likelihood-ratio tests to assess the statistical significance, whereas in the MIIP, the parameters' estimates can in a straightforward manner be transformed into test statistics for the null-hypothesis.

The analysis of a significant "lack" of coincidences of a given order is not included in the MIIP, as it considers only superpositions of processes. Under the null-hypothesis, the parameters have normal distribution with expectation zero, so the estimates can be negative, which could be interpreted as a lack of coincidences. However, a negative parameter cannot be interpreted in terms of the MIIP, because it lacks a mechanism to reduce the number of coincidences that occur by chance. Future studies should extend the model to allow for the analysis of both lacking and excess coincidences.

Moreover, we have seen that in the MIIP for the analysis of exact coincidences, the number of parameters grows like 2^n. For the extended E-MIIP including jittered coincidences, this increase happens even faster. Therefore, screening methods need to be developed to previously find out the orders of interaction of interest as has been proposed by Nakahara and Amari (2002) and Gütig, Aertsen and Rotter (2002), who dispense with the distinction between neuronal subgroups of the same order. One could as well try to reduce the interesting size of the jitter and the full generality of the model by assuming special relations between some parameters, as has been done in sections 6 to 8.

It remains to test the proposed method's practicability by careful application onto experimental data, as only then its actual significance will show up. This closes the loop back to the original question about the function and the significance of coincident assembly activity for information processing in the cortex. A genuine correlation between a specific set of neurons could be interpreted to reflect common membership in a cell assembly. Thus, the analysis method proposed in this work may help to identify assemblies and to describe their differentiated activity in the context of the experimental situation. Its usefulness is dependent on the

experimental design that must promote the specification of the relation between joint assembly activity and external events.

This work represents only a tiny piece in the world of computational neuroscience. Much more research - theoretical as well as experimental - will be needed before we may finally start to realize how sophisticated neuronal interaction leads to perception and behavior. Only in the intense cooperation between theoretical analysis and empirical experience can we hope to reveal the mysteries that nature holds ready for us.

References

[1] Abeles M. (1982a): Role of the cortical neuron: Integrator or coincidence detector? Israel Journal of Medical Sciences vol. 18, pp. 83-92

[2] Abeles M. (1982b): Local cortical circuits: An electrophysiological study. Berlin, Heidelberg, New York: Springer-Verlag.

[3] Abeles M., Bergman H., Margalit E., Vaadia E. (1993): Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. Journal of Neurophysiology vol. 70, pp. 1629-1638

[4] Aertsen A. M., Gerstein G. L. (1985): Evaluation of neuronal connectivity: Sensitivity of cross-correlation. Brain Research vol. 340, pp. 341-354

[5] Amari S. (2001): Information geometry on hierarchy of probability distributions. IEEE Transactions on Information Theory vol. 47 No. 5., pp. 1701-1711

[6] Barlow H. (1972): Single units and sensation: a neuron doctrine for perceptual psychology? Perception vol. 1, pp. 371-394

[7] Barlow H. (1992a): The biological role of neocortex. In A. Aertsen and V. Braitenberg (eds.), Information processing in the cortex, pp. 53-80. Berlin, Heidelberg, New York: Springer-Verlag.

[8] Barlow H. (1992b): Independence. In A. Aertsen and V. Braitenberg (eds.), Information processing in the cortex, pp. 167-168. Berlin, Heidelberg, New York: Springer-Verlag.

[9] Barlow H. (1992c): Neurons as computational elements. In A. Aertsen and V. Braitenberg (eds.), Information processing in the cortex, pp. 175-178. Berlin, Heidelberg, New York: Springer-Verlag.

[10] Barlow H. (1992d): Single cells versus neuronal assemblies. In A. Aertsen and V. Braitenberg (eds.), Information processing in the cortex, pp. 169-173. Berlin, Heidelberg, New York: Springer-Verlag.

[11] Billingsley, P. (1986): Probability and measure. New York: John Wiley & Sons

[12] Bishop Y., Fienberg S. E., Holland P. W. (1991): Discrete multivariate analysis. Theory and practice. Cambridge, Massachusetts, London: MIT Press

[13] Braitenberg V., Schüz A. (1991): Anatomy of the cortex. Statistics and geometry. Berlin, Heidelberg, New York: Springer-Verlag

[14] DeYoe E. A., von Essen D. C. (1988): Concurrent processing streams in the monkey visual cortex. Trends in Neuroscience vol. 115, pp. 219-226

[15] Eggermont J. J. (1990): The correlative brain. Theory and experiment in neural interaction. Berlin, Heidelberg, New York: Springer-Verlag

[16] Engel A. K., König P., Kreiter A. K., Schillen T. B., Singer W. (1992): Temporal coding in the visual cortex: new vistas on integration in the nervous system. Trends in Neuroscience vol. 15 No. 6, pp. 218-226

[17] Engel A. K., Fries P., Singer W. (2001): Dynamic predictions: Oscillations and synchrony in top-down processing. Nature reviews vol. 2, pp. 704-716

[18] Epping W., Eggermont J. J. (1987): Coherent neuronal activity in the auditory midbrain of the grassfrog. Journal of Neurophysiology vol. 57, pp. 1464-1483

[19] Fries P., Roelfsema P. R., Engel A. K., König P., Singer W. (1997): Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. Proceedings of the National Academy of Sciences USA vol. 94, pp. 12699-12704

[20] Fries P., Schröder J.-H., Singer W., Engel A. K. (2001): Conditions of perceptual selection and suppression during interocular rivalry in strabismic cats. Vision Research vol. 41, pp. 771-783

[21] Gerstein G. L., Aertsen A. M. (1985): Representation of cooperative firing activity among simultaneously recorded neurons. Journal of Neurophysiology vol. 54 no. 6, pp. 1513-1528

[22] Grün S. (1996): Unitary joint-events in multiple-neuron spiking activity. Detection, significance and interpretation. Thun, Frankfurt am Main: Deutsch

[23] Grün S., Diesmann M., Grammont F., Riehle A., Aertsen A. (1999): Detecting unitary events without discretization of time. Journal of Neuroscience Methods vol. 94, pp. 67-79

[24] Grün S., Diesmann M., Aertsen A. (2002a): 'Unitary Events' in multiple single-neuron activity: I. Detection and significance. Neural Computation vol. 14 No. 1, pp. 43-80

[25] Grün S., Diesmann M., Aertsen A. (2002b): 'Unitary Events' in multiple single-neuron activity: II. Non-Stationary data. Neural Computation vol. 14, No. 1, pp. 81-119

[26] Gütig R., Aertsen A., Rotter S. (2002): Analysis of higher-order neuronal interactions based on conditional inference. Submitted.

[27] Hebb D. O. (1949): Organization of behavior. A neurophysiological theory. New York: John Wiley & Sons

[28] Hubel D. H., Wiesel T. N. (1962): Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology vol. 160, pp. 106-154

[29] Kandel E. R., Schwartz J. H., Jessell T. M. (1996): Neurowissenschaften. Eine Einführung. Spektrum akademischer Verlag, Heidelberg

[30] König P., Engel A. K., Singer W. (1996): Integrator or coincidence detector? The role of the cortical neuron revisited. Trends in Neuroscience vol. 19 No. 4, pp. 130-137

[31] Krengel U. (1991): Einführung in die Wahrscheinlichkeitstheorie und Statistik, Braunschweig: Vieweg

[32] Mainen Z. F., Sejnowski T. J. (1995): Reliability of spike timing in neocortical neurons. Science vol. 268, pp. 1503-1506

[33] Makram H., Lübke J., Frotscher M., Sakmann B. (1997): Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science vol. 275, pp. 213-215

[34] Martignon L., von Hasseln H., Grün S., Aertsen A., Palm G. (1995): Detecting higher-order interactions among the spiking events in a group of neurons. Biological Cybernetics vol. 73, pp. 69-81

[35] Martignon L., Deco G., Lasky S., Diamond M. E., Freiwald W. A., Vaadia E. (2000): Neural coding: Higher-order interactions among the spiking events in a group of neurons. Biological Cybernetics vol. 73, pp. 69-81

[36] Munk M. H.-J., Nowak L. G., Nelson J. I., Bullier J. (1995): Structural basis of cortical synchronization II: Effects of cortical lesions. Journal of Neurophysiology, vol. 74 No. 6, pp. 2401-2414

[37] Nakahara H., Amari S., (2002): Information-geometric decomposition in spike analysis. In Diettrich T. G. and Tresp V. (eds.), NIPS, vol. 14, in press. MIT Press: Cambridge

[38] Nelson J. I., Salin P. A., Munk M. H.-J, Arzi M., Bullier J. (1992): Spatial and temporal coherence in cortico-cortical connections: A cross-correlation study in areas 17 and 18 in the cat. Visual Neuroscience vol. 9, pp. 21-37

[39] Nowak L. G., Munk M. H.-J., Nelson J. I., James A. C., Bullier J. (1995): Structural basis of cortical synchronization I: Three types of interhemispheric coupling. Journal of Neurophysiology vol. 74 No. 6, pp. 2379-2400

[40] Perkel D. H., Gerstein G. L., Moore G. P. (1967): Neuronal spiketrains and stochastic point processes: II: Simultaneous spike trains. Biophysical Journal vol. 7 No. 4, pp. 419-440

[41] Riehle A., Grün S., Diesmann M., Aertsen A. (1997): Spike synchronization and rate modulation differentially involved in motor cortical function. Science vol. 278, pp. 1950-1953

[42] Roy A., Steinmetz P. N., Niebur E. (2000): Rate limitations of Unitary Event analysis. Neural Computation vol. 12 No. 9, pp. 2063-2082

[43] Sachs L. (1971) Statistische Auswertungsmethoden. Berlin: Springer

[44] Schmidt K. E., Grün S., Singer W., Galuske R. A. W. (2001): Parallel multi-unit recordings in primary visual cortex of the un-restrained awake and sleeping cat. Society for Neuroscience Abstracts, p. 821:58

[45] Singer W., Gray C. (1995): Visual feature integration and the temporal correlation hypothesis. Annual Review of Neuroscience vol. 18, pp. 555-586

[46] Singer W., Engel A. K., Kreiter A. K., Munk M. H.-J., Neuenschwander S., Roelfsema P. R. (1997): Neuronal assemblies: necessity, signature and detectability. Trends in Cognitive Sciences vol. 1 No. 7, pp. 252-261

[47] Singer W. (1999): Neuronal synchrony: A versatile code for the definition of relations? Neuron vol. 24, pp. 49-65

[48] Thorpe S., Fize D., Marlot C. (1996): Speed of processing in the human visual system. Nature vol. 38, pp. 520-522

[49] Toyama K., Kimura M., Tanaka K. (1981a): Cross-Correlation analysis of interneuronal connectivity in cat visual cortex. Journal of Neurophysiology vol. 46 No. 2, pp. 191-201

[50] Toyama K., Kimura M., Tanaka K. (1981b): Organization of cat visual cortex as investigated by cross-correlation technique. Journal of Neurophysiology vol. 46 No. 2, pp. 202-214

[51] Vaadia E., Haalman I., Abeles M., Bergman H., Prut Y., Slovin H., Aertsen A. (1995): Dynamics of neuronal interactions in monkey cortex in relation to behavioral events. Nature vol. 373, pp. 515-518

[52] von der Malsburg C. (1981): The correlation theory of brain function. Internal Report 81-2. Max-Planck-Institute for Biophysical Chemistry. (online at: http://www.neuroinformatik.ruhr-uni-bochum.de/ini/VDM/PUB-LIST/1981)

[53] White J. A., Chow C. C., Ritt J., Soto-Treviño C., Kopell N. (1998): Synchronization and oscillatory dynamics in heterogeneous, mutually inhibited neurons. Journal of Computational Neuroscience vol. 5, pp. 5-16

Erklärung

Hiermit erkläre ich, daß die vorliegende Arbeit selbständig verfaßt und keine anderen als die angegebenen Hilfsmittel benutzt sowie die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, durch Angabe der Quellen kenntlich gemacht wurden.

Frankfurt am Main, Juni 2002

Gaby Schneider