Mechanisms of signal encoding and information transmission in cortical neurons

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Physik
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Tim Herfurth
aus Berlin

Frankfurt am Main, Juli 2019

vom Fachbereich Physik der

Johann Wolfgang Goethe-Universität Frankfurt

als Dissertation angenommen.

Dekan: Prof. Dr. Michael Lang

Gutachter: Dr. Tatjana Tchumatchenko
Prof. Dr. Achilleas Frangakis

Datum der Disputation:

"If the human brain were so simple that we could understand it, we would be so simple that we couldn't."

Emerson M. Pugh, *The Biological Origin of Human Values*

Deutsche Zusammenfassung

Die fundamentale Funktion des Gehirns und von Nervensystemen ist es Informationen zu übertragen und zu verarbeiten. Es ist diese Funktion, die tierischem Verhalten zugrunde liegt, von einfachen reizinduzierten Reflexen bis hin zu den komplexen kognitiven Fähigkeiten höherer Säugetiere. Dabei werden Informationen, zum Beispiel über sensorische Reize, innerhalb des Nervensystems weitergegeben und prozessiert und führen im Ergebnis zu motorischem Output und damit Verhalten. Die elementaren Konstituenten des Gehirns, durch deren Zusammenwirken und Kommunikation diese Mechanismen realisiert werden, sind Nervenzellen bzw. Neuronen. Diese wechselwirken miteinander, indem sie diskretisierte elektrische Impulse, die Aktionspotenziale oder *Spikes*, über Synapsen austauschen, die elektrische Ströme in den postsynaptischen Neuronen induzieren. Oft repräsentiert die Statistik dieser Ströme dabei einen zu enkodierenden Stimulus. Die synaptischen Ströme werden von den Neuronen elektrophysiologisch integriert und bestimmen wiederum ob und wann ein postsynaptisches Neuron Spikes emittiert.

Es wird oft angenommen, dass Sequenzen von Spikes die fundamentalen Einheiten der interneuronalen Kommunikation und somit der Enkodierung von Information in Neuronensystemen sind. Eine entsprechende Kernfrage der Neurowissenschaften ist die nach dem neuronalen Code, dem Schema durch das Information in den Spikesequenzen enkodiert ist. Eine daraus abgeleitetes Forschungsfeld versucht zu quantifizieren, wieviel Information über einen gegebenen Stimulus oder Eingangsstrom in den Spikes eines Neurons enthalten ist. In einer Reihe früherer Arbeiten wurden mathematische Methoden entwickelt, die es erlauben diese Größe des Informationsgehalt, die *mutual information* genannt wird, zu berechnen. Neben vieler Methoden, die Näherungen und Schranken dieser Größe bestimmen lassen, erlaubt es eine kürzlich entwickelte Methode die relevante Information unter bestimmten Stationaritätsannahmen exakt zu ermitteln. Diese sogenannte *Korrelationsmethode* berechnet die mutual information für einzelnen Neuronen mittels zeitlicher Spike-Korrelationen. Basierend auf der Korrelationsmethode als methodische Grundlage untersucht die vorliegende Dissertation wie verschiedene Stimuluseigenschaften die Kapazität interneuronaler Informationsübertragung beeinflussen und wie Informationsübertragung in Nervensystemen entsprechend optimiert werden

kann. Zudem trägt die Arbeit zu einem besseren allgemeinen Verständnis mathematisch-methodischer Aspekte neuronaler Informationsenkodierung bei, indem relevante informationstheoretische Methoden analysiert und verglichen werden. Dabei orientieren sich die theoretischen Modellbeschreibungen an den beobachteten Eigenschaften und Prozessen des cerebralen Cortex, also der Großhirnrinde und zielen darauf ebendiese zu beschreiben. Die spezifischen Fragestellungen, angewandten Methoden und erhaltenen Ergebnissen dieser Dissertation werden im Folgenden näher erläutert.

Kapitel 2 der Dissertation befasst sich mit der Stimulus-Enkodierung mittels dynamischer Änderungen der Spikerate, welche die Dynamik des zugrundeliegenden Stimulus widerspiegeln (*rate coding*), ohne dabei synaptische Eingangsströme explizit in Betracht zu ziehen. Im Besonderen wird dabei quantitativ die Frage untersucht, welchen Einfluss zeitliche Korrelationen innerhalb der Ratendynamik auf die durch sie übertragbare Information haben. Diese Frage ist von allgemeinem Interesse, weil die Bedeutung solcher Korrelationen für die Funktion des Gehirns nicht vollständig geklärt ist und bisher nur approximativ und/oder qualitativ behandelt wurde. Die Nervenzellen werden dabei als sogenannte Poisson-Neuronen modelliert, die Spikes stochastisch und mit einer gegebenen Spikerate so erzeugen, dass die zeitlichen Abstände zwischen zwei Spikes unabhängig sind und eine hohe statistische Dispersion haben. Das Modell stützt sich unter anderem darauf, dass dieses hohe Maß an Variabilität in den Spike-Intervallen wiederholt im cerebralen Cortex beobachtet worden ist. Die Unabhängigkeit der Spike-Intervalle bei gegebener Rate führt zudem dazu, dass Korrelationen in den Spike-Zeitpunkten nur durch Korrelationen in der Ratendynamik selbst bedingt sind und nicht anderweitig entstehen können. Der gewählte Modellrahmen eignet sich deshalb um explizit den Effekt von Ratenkorrelationen zu untersuchen.

Einen Ausdruck zur entprechenden Berechnung der Information, die nur Ratenkorrelationen und keine sonstigen Spikekorrelationen in Betracht zieht, erhalte ich durch eine Formulierung der Korrelationsmethode, die im Allgemeinen alle Typen von Spikekorrelationen berücksichtigt, für den Spezialfall ratenkorrelierter Poisson-Neuronen. Durch einen Vergleich dieses Ausdrucks mit bekannten Formeln für die Information unter der Annahme gänzlich unabhängiger Spikes ist es nun möglich die informationstheoretische Redundanz des neuronalen Codes zu quantifizieren, die durch Ratenkorrelationen entsteht. Diese Redundanz wurde zuvor qualitativ beschrieben und resultiert daraus, dass zeitlich eng beinander liegende Spikes in diesem Fall Information nicht unabhängig enkodieren. Der hergeleitete Ausdruck hängt nur von der durschnittlichen Spikerate und der Autokorrelation der Ratendynamik ab. Da diese Statistiken erster und zweiter Ordnung leicht aus relativ wenig Daten berechnet werden können, stellt der Ausdruck eine wertvolle und zugängliche Alternative zu bisherigen Methoden dar die Information in Spikesequenzen zu berechnen; entweder als Näherung unter der Annahme, dass nur

Ratenkorrelationen vorliegen oder als Referenz, um die Rolle verschiedener Typen von Spikekorrelationen unabhängig von der Poisson-Annahme zu analysieren.

Im weiteren Verlauf von Kapitel 2 untersuche ich die Eigenschaften der Ratendynamik, die die Enkodierungsredundanz bedingen, genauer. Dies geschieht analytisch sowohl für den Allgemeinfall als auch anhand von drei stochatischen Beispiel-Prozessen für die Ratendynamik. Ich stelle fest, dass die Enkodierungsredundanz sowohl durch die statistische Gesamtverteilung der Spikeraten als auch durch die oben beschriebene Autokorrelationsfunktion des Ratenprozesses beschrieben ist. Ersteres bestimmt dabei den Informationsübertrag mittels unabhägiger Spikes und Letzteres die Information unter Berücksichtigung von Ratenkorrelationen. Ich zeige außerdem erstmals analytisch, dass im Grenzfall sehr kleiner Ratenmodulierungen, kleinen Stimulusstärken entsprechend, die Informationsübertragung durch korrelierte und unabhängige Spikes in führender Ordnung identisch sind und nur von der Modulierungstiefe der Ratenynamik abhängen. Weiter mache ich die Festellung, dass die Information in ratenkorrelierten Spikes zwar wie erwartbar durch verschwindend kleine Korrleationszeiten der Ratendynamik maximiziert wird, dieser Grenzwert überraschender Weise aber nicht äquivalent zum Ausdruck für unabhänginge Spikes ist. Ich diskutiere mögliche Ursachen für diesen vermeintlichen Widerspruch und begründe meine Vermutung, dass den entsprechenden Methoden verschiedene Annahmen zugrunde liegen.

Kapitel 3 beleuchtet eine andere Fragestellung der neuronalen Informationsübertragung, abermals auf Basis der Korrelationsmethode. Ausgangspunkt sind zwei verschiedene Mechanismen der Enkodierung von dynamischen Stimuli durch Modulierung der stochastischen, synaptischen Eingangsströme. Es handelt sich dabei um die stimulusabhängige Modulierung des mittleren Eingangsstroms (*mean modulation*, MM) beziehungsweise der Varianz des Eingangsstroms (*variance modulation*, VM) bei jeweils konstanter anderer Größe. Im Wesentlichen beschreiben diese Modulierungen das Enkodieren eines Stimulus in der dynamischen Änderung der Baseline und – etwas unkonventioneller – der Fluktuationsstärke der Eingangsströme. Für die Präsenz beider Mechanismen *in vivo* bestehen experimentelle Hinweise in kortikalen Netzwerken mit *excitation-inhibition balance*. Leitend für die hier behandelten Fragestellungen sind allerdings theoretische und *in vitro* Studien denen zufolge Neuronen auf Populationsebene Änderungen in der Varianz des Eingansstroms extrem schnell in Änderungen ihrer Spikerate umsetzen können, was eine hohe Kapazität der Informationsübertragung im VM-Kanal von Stimuli mit schneller Dynamik impliziert. Im Gegensatz dazu reagieren Neuronen auf Änderungen des mittleren Eingansstroms träger. Obgleich diese Ergebnisse vorallem vor dem Hintergrund einiger Annahmen gemacht wurden, z.B. dass die Eingangsströme eine Komponente weißen Rauschens besitzen, legen sie nahe, dass Modulierung der Fluktuationsamplituden bessere Eigenschaften hinsichtlich einer optimierten Informationsübertragung zwischen

Neuronen haben als Modulierungen der Strom-Mittelwerte. Diese Hypothese war aufgrund mangelnder Methoden zur exakten Bestimmung der Information in beiden Fällen nicht umfänglich untersucht worden und ist Gegenstand dieser Dissertation.

Als mathematische Neuronenmodelle dienen hier *integrate-and-fire* Modelle, die explizit die Dynamik der Membranspannung als Funktion der synaptischen Eingangsströme bestimmen und dabei diskrete Spike-Zeitpunkte ermitteln. Im Speziellen verwende ich das *leaky integrate-and-fire* (LIF) und das *exponential integrate-and-fire* (EIF) Modell. Beide Modelle können durch die genäherte Dynamik der Membranspannung die elektrophysiologischen Antworteigenschaften von kortikalen Pyramidalzellen gut beschreiben.

Ich erläutere zunächst wie beide Enkodierungsmechanismen (MM und VM) aus der kombinierten stimulusabhängigen Aktivität präsynaptischer Neuronen entstehen können und formalisiere die entsprechenden Eingangsströme als mathematische Verknüpfungen der zeitlichen Verläufe von Stimulus- und Noise-Anteilen. Sowohl Stimulus als auch Noise sind dabei stationäre, stochastische Prozesse, deren Eigenschaften, wie etwa ihre Korrelationszeit oder mittlere Amplitude, als freie Parameter bestimmt werden können. Im Besonderen ist die dominante Frequenz der spektralen Stimulus-Leistungsdichte so parametrisiert, dass verschiedene Stimulus-Geschwindigkeiten analysiert werden können. Aus den Eingansströmen werden Spikesequenzen so generiert, dass die Korrelationsmethode durch eine geeignete Fourieranalyse angewendet werden und der Informationsübertrag berechnet werden kann. Dies geschieht mit Hilfe von Computer-Simulationen, in denen die stochastischen Zeitreihen erzeugt werden und die Differenzialgleichungen, die die Neuronenmodelle definieren, per Euler-Verfahren gelöst werden. Für den Fall von LIF-Neuronen mit weißem Hintergrundrauschen entwickele ich zudem eine analytische, störungstheoretische Näherung der *mutual information*, die auf linearen Antwortfunktionen beruht. Diese neue Approximation erlaubt es den Zusammenhang neuronaler Antwortfunktionen mit der Kapazität der Informationübertragung zu beleuchten.

Das Hauptresultat hinsichtlich des Vergleiches beider Modulierungskanäle lautet, dass Modulierungen des Strom-Mittels in biologisch plausiblen Settings das effektivere Schema der interneuronalen Kommunikation sind. Dieses Ergebnis ist nur in der künstlichen Situation mit weißem Hintergrundrauschen nicht gültig, wo extrem schnelle Stimuli besser durch VM übertragen werden können (vgl. oben). Wird eine realistische Zeit der Spike-Initiierungsdynamik angenommen, was bei EIF-Neuronen möglich ist, dann gilt das Ergebnis in jedem Fall und die frequenzabhängige Übertragung von Stimulusinformation entspricht der eines *low-pass* Filters. Für beide Enkodierungskanäle untersuche ich die Abhängigkeit der *mutual information* von veschiedenen Stimulus- und Noiseparametern und finde beispielsweise, dass eine Änderung der Stimulus-Korrelationszeit

entgegengesetze Effekte auf MM und VM hat. Des Weiteren führe ich eine Untersuchung der Informationsübertragung mittels einer kombinierten, simultanen Modulierung von Mittelwert und Varianz der Eingangsströme durch. Es zeigt sich, dass nichtlineare Kopplungseffekte zwischen beiden Modulierungen zu einer vorteilhaften, synergistischen Informationsenkodierung führen und deshalb biologisch bedeutsam sein können.

Weiterhin untersuche ich erstmals systematisch, welcher Anteil der übertragenen Information in verschiedenen Szenarien linear dekodierbar ist, also mithilfe einer linear Faltung der Antwort-Spikesequenzen. Diese Fragestellung ist von besonderer Bedeutung, weil zahlreiche vorherigen Studien sich auf die Annahme gestützt haben, dass alle relevante Information linear enkodiert ist (die entsprechenden Methoden werden verglichen), ohne die dabei vernachlässigten Anteile zu kennen. Im Allgemeinen stelle ich fest, dass der Anteil nichtlinear enkodierter Information mit der insgesamt übertragenen Information ansteigt. Eine Schlussfolgerung ist, dass Nichtlinearitäten eine signifikante Rolle im neuronalen Code spielen können. Entsprechend können lineare Antwortfunktionen, die oftmals als entscheidende Chrakteristik für Informationsenkodierung benutzt wurden, zwar in vielen Fällen eine gute Intuition für die relevanten Mechanismen liefern, es müssen aber im Allgemeinen auch Effekte höherer Ordnung in Betracht gezogen werden. Verschiedene Parameter können diesen Effekt beeinflussen. So haben z.B. höhere Spikeraten und Noiseamplituden eine linearisierende Wirkung. Weitere Einsichten erhalte ich auf allgemeinerer, methodischer Ebene. Durch eine Kombination aus simulationsbasierten Fourieranalysen und analytischen Berechnungen zeige ich, dass eine zuvor als obere Schranke der Information betrachtete Größe tatsächlich äquivalent zur exakten Information ist. Außerdem demonstriere ich die Gültigkeit einiger statistischer Eigenschaften der Fourierkomponenten der Spikesequenzen, auf deren Annahme die Anwendbarkeit der Korrelationsmethode beruht.

Zusammenfassend betrachtet untersuche ich zentrale Fragen der theoretischen Neurowissenschaften bezüglich des neuronalen Codes und neuronaler Informationsübertragung. Diese fokussieren sich vorwiegend auf die Bedeutung zeitlicher Korrelationen im Spikecode, unterschiedliche Mechanismen der Informationsenkodierung in synaptischen Strömen, nichtlineare Komponenten der Informationsenkodierung und methodische Aspekte. Dabei werden Erkenntnisse gewonnen, die einen Beitrag zum besseren Verständnis der interneuronalen Kommunikation im Gehirn leisten. Die meisten der erlangten Ergebnisse beruhen auf spezifischen und/oder vereinfachenden Annahmen. So beziehe ich mich zumeist auf dekorrelierte Neuronenpopulationen statt Netzwerke und verwende bestimmte Neuronenmodelle und Stationaritätsannahmen. Wenngleich solche Modellannahmen notwendiger Teil quantitativer Grundlagenforschung sind und hier biologisch begründet werden, ist es wichtig sich ihrer als potenzielle Einschränkungen der Allgemeinheit der

Ergebnisse bewusst zu sein. In Kapitel 4 diskutiere ich die Bedeutung meiner Ergebnisse in einem größeren neurowissenschaftlichen Kontext, spezifiziere potenzielle Einschränkungen und gebe einen Ausblick auf mögliche Erweiterungen der durchgeführten Studien.

Abstract

As its fundamental function, the brain processes and transmits information using populations of interconnected nerve cells alias neurons. The communication between these neurons occurs via discrete electric impulses called spikes. A core challenge in neuroscience has been to quantify how much information about relevant stimuli or signals a neuron transports in its spike sequences, or spike trains. The recently introduced *correlation method* allows to determine this so-called mutual information in terms of a neuron's temporal spike correlations under certain stationarity assumptions. Based on the correlation method, I address several open questions regarding neural information encoding in the cortex.

In the first part (chapter 2), I investigate the role of temporal spike correlations for neural information transmission. Temporal correlations in neuronal spike trains diminish independence in the information that is transmitted by the different spikes and hence introduce redundancy to stimulus encoding. However, exact methods to describe how such spike correlations impact information transmission quantitatively have been lacking. Here, I provide a general measure for the information carried by spike trains of neurons with correlated rate modulations only, neglecting other spike correlations, and use it to investigate the effect of rate correlations on encoding redundancy. I derive it analytically by calculating the mutual information between a time-correlated, rate-modulating signal and the resulting spikes of Poisson neurons. Whereas this information is determined by spike autocorrelations only, the redundancy in information encoding due to rate correlations depends on both the distribution and the autocorrelation of the rate histogram. I further demonstrate that, at very small signal strengths, the information carried by rate correlated spikes becomes identical to that of independent spikes, in effect measuring the rate modulation depth. In contrast, a vanishing signal correlation time maximizes information transmission but does not generally yield the information of independent spikes.

In the second part (chapter 3), I analyze the information transmission capabilities of two particular schemes of encoding stimuli in the synaptic inputs using *integrate-and-fire* neuron models. Specifically, I calculate the exact information contained in spike trains about signals which modulate either the mean or the variance of the somatic currents in neurons, as is observed experimentally. I show that the information content about mean modulating signals is generally substantially larger than about variance modulating signals for biological parameters. This result provides evidence, by means of exact calculations of the mutual information, against the potential benefit of variance encoding that had been suggested previously.

Another analysis reveals that higher information transmission is generally associated with a larger proportion of nonlinear signal encoding. Moreover, I show that a combination of signal-dependent mean and variance modulations of the input current can synergistically benefit information transmission through a nonlinear coupling of both channels. On a more general level, I identify what was previously considered an upper bound as the exact, full mutual information. Furthermore, by analyzing the statistics of the spike train Fourier coefficients, I identify the means of the Fourier coefficients as information-carrying features.

Overall, this work contributes answers to central questions of theoretical neuroscience concerning the neural code and neural information transmission. It sheds light on the role of signal-induced temporal correlations for neural coding by providing insight into how signal features shape redundancy and by establishing mathematical links between existing methods and providing new insights into the spike train statistics in stationary situations. Moreover, I determine what fraction of the mutual information is linearly decodable for two specific signal encoding schemes.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Dr. Tatjana Tchumatchenko for the continuous support of my Ph.D. study and related or unrelated research, for her patience, motivation, open-mindedness, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee and, in particular, Prof. Dr. Achilleas Frangakis for his support and efforts as co-examiner of this dissertation.

I thank my fellow labmates for stimulating discussions and creating a friendly and supportive working environment.

Moreover, I am particularly grateful to my family and friends who helped me through difficult times and made me enjoy the good times even more.

Contents

List of publications

The following publications resulted from the research that I present in this dissertation. Accordingly, many of the methods and results as well as elements of the text in this dissertation are also part of these publications.

- Herfurth, T. & Tchumatchenko, T. (2017). How linear response shaped models of neural circuits and the quest for alternatives. Current Opinion in Neurobiology, 46, 234-240. [1]

- Herfurth, T. & Tchumatchenko, T. (2019). Quantifying encoding redundancy induced by rate correlations in Poisson neurons. Physical Review E, 99(4), 042402. [2]

- Herfurth, T. & Tchumatchenko, T. (2019). Information transmission of mean and variance coding in integrate-and-fire neurons. Physical Review E, 99(3), 032420. [3]

Abbreviations

AP	**A**ction **P**otential
EPSP/IPSP	**E**xcitatory/**I**nhibitory **P**ostsynaptic **P**otential
GABA	**G**amma **A**mino **B**utyric **A**cid
LIF	**L**eaky **I**ntegrate-and-**F**ire
EIF	**e**xponential **I**ntegrate-and-**F**ire
PSTH	**P**eri-**S**timulus **T**ime **H**istogram
V1	**primary V**isual cortex
CV	**C**oefficient of **V**ariation
ISI	**I**nter-**S**pike of **I**nterval
E/I	**E**xcitation/**I**nhibition
LN	**L**inear-**N**onlinear
GLM	**G**eneralized **L**inear **M**odel
MM/VM	**M**ean/**V**ariance modulation
LRT	**L**inear **R**esponse **T**heory
FC	**F**ourier **C**oefficients
MVN	**M**ulti**v**ariate **N**ormality

Chapter 1

Biological and mathematical foundations

This introductory chapter provides the background and context for the research questions that I address in detail in the subsequent chapters. Thereby, the scientific scope is successively narrowed, starting with general neurobiological considerations moving to the specific research questions. Likewise, within each topic area matters are first discussed on a broad, qualitative level and are subsequently formalized in mathematical models.

In the first section 1.1, I introduce neurons as the basic biological units of signal transmission and processing in the brain. Thereby, their relevant electrophysiological properties are presented and formalized as Poisson and integrate-and-fire type models. Section 1.2 deals with the way in which neurons are organized in the mammalian cortex, and elaborates on how this organization gives rise to irregular activity and states of balanced excitation and inhibition.

The first research questions of this dissertation, regarding the quantification of information transmission in inter-neural communication, are derived in section 1.3. I first introduce the notion of the neural code and subsequently give an understanding of the conceptual and mathematical challenges of determining the mutual information of stimuli and spike trains in neurons, from which relevant questions are derived. On a different note, two specific means to encode a given signal via modulations of the input current mean and variance are presented in section 1.4, based on the balance of excitation and

inhibition. Linking the two encoding schemes to the preceding section, I clarify why the information-theoretic investigation of mean and variance encoding helps to address important questions in neuroscience. The scope and the outline of the rest of this dissertation, including a flow diagram of its organization, are provided in section 1.5.

1.1 Neurons, the computational units of the brain

Nervous systems, such as the brain, set animals apart from other living beings and today it is uncontested that all animal behavior as well as inner mental states – including the human mind – result from processes in the brain [4, 5]. These processes, in turn, emerge from specific interactions of brain cells that integrate sensory information, compute relevant features, produce behavior, etc. I discuss these interactions in this chapter.

The cells of the brain, and nervous systems in general are divided into two classes: nerve cells, or neurons, and glial cells, or glia [5][1]. For example, an average human brain contains around 85 billion neurons and about the same number of non-neuronal cells (mostly glia) with varying ratios in different brain regions [6]. Glial cells fulfill a number of mostly supportive functions in the brain and are usually not considered essential for information processing [7]. The most abundant type of glial cells in the brain are astrocytes. These cells separate and insulate neurons and neural groups, they maintain the concentrations of different ions and neurotransmitters in the extracellular space, they guide neuronal growth and help nourish neurons. Another glia type called microglia functions as immune system cells. Furthermore, oligodendrocytes and Schwann cells comprise insulating sheaths around neurons (see figure 1.2) that serve important functions for signal enhancing and the electrophysiological properties of the neurons [5] (see section 1.1.3). It needs to be emphasized that the role of glia for signal processing is not entirely clear and neuron-glial signal processing through astrocytes has been reported [5, 7]. Moreover, a number of other modulatory effects of glial cells on neural computation has been reported [8]. However, these findings do not provide conclusive evidence. Therefore, I do not consider glial cells explicitly in the present work.

In contrast to glia, it is commonly accepted that *neurons are the fundamental, computational and processing units in the brain* [4, 5, 7, 9]. Neurons are therefore the primary

[1] Note that also small amounts of blood vessel cells, mesenchymal stem cells, and other cell types are found in the brain [6].

biological subject of the present work that deals with the mechanisms of information transmission in neural systems. In the remainder of this section I introduce the basics of neurobiology. Next to a general introduction (sections 1.1.1 and 1.1.2), this includes neuroelectrophysiology (section 1.1.3), the mathematical formulations thereof in terms of neuron models (section 1.1.4) and the theory of synaptic transmission (section 1.1.5).

1.1.1 The neural doctrine

The neuron has a rather young scientific history even though humankind had investigated the brain for thousands of years. A major reason for that is technical: as most brain cells are around 0.01 to 0.05 millimeters in diameter, it required microscopes to visualize them. Moreover, the soft consistency impeded the preparation of thin tissue slices that were necessary for microscopy. Another challenge was given by the fact that untreated slices of brain tissue do not exhibit salient visible structure. In 1873, a staining technique invented by Camillo Golgi eventually allowed to identify neurons under the microscope in considerable detail [7]. An example of a Golgi stain is shown in figure 1.1. Despite having the staining method at hand, the interpretation of the obtained images was not unequivocal. The preliminary view that resulted from brain tissue imaging was that the nerve cells are anatomically separable and form a continuous reticulum [7, 10, 11].

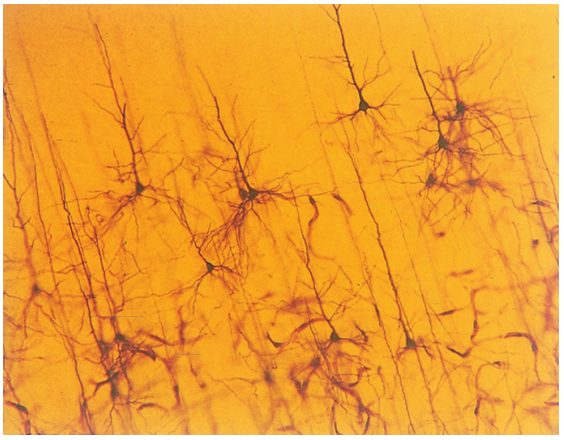


FIGURE 1.1: **Golgi-stained neurons.** From [12].

This view was incompatible with the established cell theory according to which all biological organisms are composed of individual cells [13]. It was Santiago Ramón y Cajal who stated that the cell theory applies to the brain and that neurons form contacts with each other instead of a continuum [4, 7, 10]. Cajal's idea is known as the *neural doctrine*

and nicely manifested in his work *Textura del Sistema Nervioso del Hombre y de los Vertebrados* [14]. The neural doctrine is following the notion of neurons as fundamental processing units.

1.1.2 Neuron biology and structure

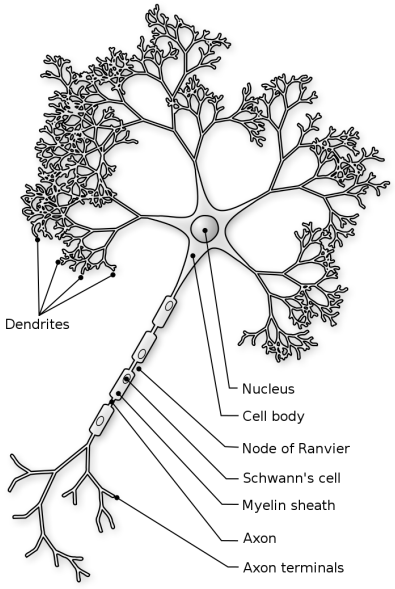


FIGURE 1.2: **Schematic of a neuron with important sites annotated.** Synapses are not shown. Image source: work by Nicolas Rougier, via Wikimedia Commons (Creative Commons).

Neurons are electrically and chemically excitable cells and develop from neuroepithelial cells of the embryonic nervous system. The cells are bounded and separated from the outside by the cell membrane which as all biological membranes is a bilayer structure and provides a hydrophobic barrier impermeable to most water-soluble substances [5]. The membrane is about 5 nm thick and is studded with proteins. It is the distinctive features of membrane proteins that give rise to the information processing capabilities of neurons. Some of the membrane-associated proteins pump substances from the inside to the outside. Others form pores that regulate which substances can gain access to the inside of the neuron [7]. The importance of membrane proteins becomes clear in section 1.1.3 where I discuss the role of channel proteins for the electrophysiological properties of neurons.

The morphology of neurons varies considerably across species and location in the nervous system (see below). However, neurons share several cellular features, and a typical neuron structurally consists of a cell body (soma), dendrites, and an axon [7]. A schematic of a neuron is shown in figure 1.2.

1.1.2.1 Basic structure of a neuron

The soma. The soma is a neuron's central compartment. It has an approximately spherical shape and about 20 µm in diameter. The inside of the cell (cytoplasm) is comprised of cytosol and membranous organelles. The cytosol is a salty solution containing a number of ions and molecules that are crucial for the cell's excitability (see section 1.1.3). The organelles found in neurons are the same as in all animal cells. The most important among these are the nucleus, the rough endoplasmic reticulum (ER), the smooth ER, the Golgi aparatus, and the mitochondria [7, 15]. These organelles store and express genetic information, provide the cell's energy supply, and fulfill other essential functions[2].

The cytoskeleton. The cytoskeleton functions as a scaffold for the neuron. It further determines the neuron's shape and the distribution of organelles within the cell. The cytoskeleton is comprised of three main structures: *microtubules, microfilaments* and *neurofilaments*. These filaments account for approximately 25% of the total proteins of a neuron [5].

The axon. Contrary to the elements introduced so far, the axon is unique to neurons and not part of other cell types. The axon is a slender process emerging from the soma that, similarly to a telegraph wire, conducts electric pulses (action potentials) away from the soma along the axon's processes (see figure 1.3). As a signal conductor, the axon is essential for interneural communication. It is the particular set of proteins in the axon membrane that facilitates the propagation of action electric signals in the axon as is explained in section 1.1.3. The axon process starts on the soma in a region called axon hillock, and the main process may branch several times along the way. These branches

[2]Despite being very important, the details of the processes involving the cell organelles are not of particular interest in the present work, and I refer to textbooks such as Alberts *et al.* [15] for a detailed discussion of cell biology.

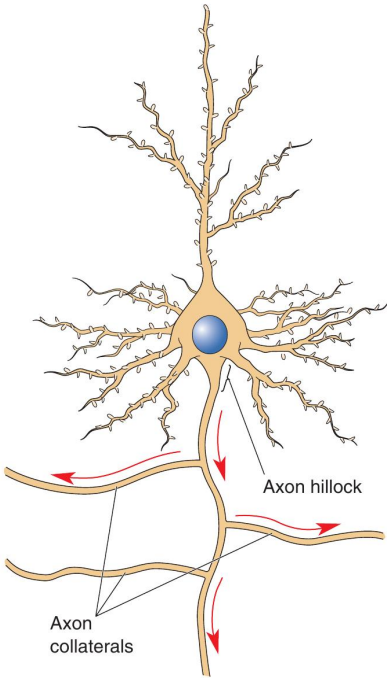


FIGURE 1.3: **Signal propagation in an axon.** Action potentials are elicited at the axon hillock and then travel away from the soma down the axon and its collaterals, much like current in a branched wire. From [7].

are called axon collaterals (see figure 1.3). The length of axons varies considerably between less than a millimeter and over a meter. The diameter ranges from less than a millimeter to 25 mm [7].

The end parts of axons (axon collaterals) are called axon terminals or axon boutons. The axon terminal is the place where a neuron makes contact to other neurons (in some cases the axon contacts the cell it belongs to) [7]. This contact is known as *synapse* and is the major site of biochemical communication between two neurons (see section 1.1.5).

Dendrites. The term dendrite stems from the Greek word for tree owing to the branches and tree-like structure of these neurites (see figures 1.2 and 1.4). If one visualizes the axon as the sending wire of a neuron the dendrites can be considered the antennae or receiver. Accordingly, the dendrites host the receiving ("postsynaptic") parts of the synapses that the axons connect to [5, 7].

The structure, or morphology, of the dendritic trees is often used to classify different neuron types [16], and relations between a neuron's dendritic structure and its function exist [17–20]. One way to categorize neurons is by the number of neurites that extend

from their soma (cf. figure 1.4). Other classification criteria for neuron types are molecular composition (proteome), connection properties, axon length, or neurotransmitter types. Often traits in these categories are correlated and, overall, there are around 100 different neuron types [5, 7, 16]. In the present work, numerical values for neuron properties, when required, are chosen to be representative for pyramidal cells, the principal excitatory neurons in many regions such as cortex, hippocampus, and the amygdala [5] (see section 1.1.4).

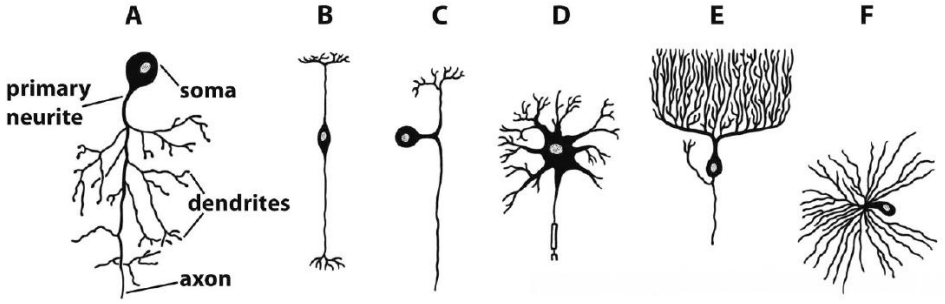


FIGURE 1.4: **Neuron types classified by dendritic tree structure. A.** Unipolar neuron. **B.** Bipolar neuron. **C.** Pseudo-unipolar neuron. **D, E, F.** Multipolar neurons with different morphologies. **D** shows a motor neuron, **E** shows a Purkinje cell of the cerebellum, **F** shows a stellate cell. Adapted from [16] (modified from various sources).

Whereas it has long been thought that dendrites collect the signals arriving at the synapses and simply pass them to the soma for further processing, the functional role of dendrites for neural computation and information processing is, in fact, crucial and very diverse [5, 18]. The computational function of dendrites emerges from both passive electric properties and active mechanisms such as dendritic action potentials. Examples for dendritic computation are low-pass filtering, logical operations realized by a combination of excitation and shunting inhibition, signal amplification through nonlinear summation, coincidence detection activated by active dendritic mechanisms, and the mediation of synaptic plasticity through dendritic action potentials [5, 19]. It is important to note that even in the face of the diversity of dendritic computations, they mostly and effectively influence the current/voltage changes arriving at the soma where further signal processing occurs. On this account, I do not explicitly model the processes occurring at the synapses or in dendrites but their effects on the site of action potential initiation.

1.1.3 Electrophysiology of the neuron

In order for the neuron to be a "computational unit", it needs to be able to receive, process, and distribute information. In a neuron, these functions are implemented electrically, and computations in the brain thus often correspond to transformations of electric voltages and currents. In some regards axons can be compared to the workings of a telephone wire: axons, similarly to cables, have a cylindrical shape and also have the function to carry electrical signals. However, in axons signals are not conducted by free electrons but by ions and a neuron cannot be seen as an electrically well-insulated system as is a cable [4, 7]. The electrophysiology of living organisms is based on exploiting the electric power stored in concentration gradients and voltage differences. Evolutionary, this has been made possible by the "invention" of transmembrane ion channels [4] whose details have been discovered rather recently [21]. These channels enable cells to a specific form of electric pulse called *action potential* (AP) and hence make cells excitable. This section is explains the basic processes constituting neuronal electrophysiology, i.e. the electric properties of nerve cells.

1.1.3.1 The membrane resting potential

If one measures the internal electric potential of a neuron at rest, i.e., if it is not electrically interacting with its environment, one typically finds values of –40 to –90 mV. This negative potential reflects the excess of negative charge inside the cell relative to the extracellular fluid and is called *resting potential*. The role of the resting potential is crucial as rapid reversals of the negative voltage feature the action potentials which in turn facilitate a neuron's communication with other neurons [4, 5, 7].

The resting potential is established by the differences in concentration of charged molecules and ions in the intra- and extracellular fluid of the cell. These concentrations are comparatively similar across different cell types, tissues, and animal species. Ranges of typical concentrations of the most important ions are given in table 1.1. The particularly high inside/outside concentration differences (> 100 mM) of sodium and potassium are fundamental for the occurrence and shape of action potentials [22].

The cell membrane itself is not permeable to ions and, therefore, maintains the concentration gradients. However, some transmembrane proteins are permeable to one or

Ion	intracellular range (mM)	extracellular range (mM)
Na^+	5–20	130–160
K^+	130–160	4–8
Ca^{2+}	$0.05{\cdot}10^{-3}$–$1{\cdot}10^{-3}$	1.2–4
Mg^{2+}	10–20	1–5
Cl^-	1–60	100–140

TABLE 1.1: Ranges of concentrations of the most important ions found in the intra- and extracellular fluid of animals [22], given in millimol (mM).

multiple ion types. Now, if one assumes the membrane to be permeable to one ion type, there are two forces acting on these ions. A chemical driving force enforces the movement of ions against the concentration gradient such that the concentrations of either side of the membrane equalize. This force is related to diffusion which favors the equal distribution of particles and increases in strength with higher temperatures. The second force is the electrical driving force which reflects the fact that equal charges repel and opposite charges attract each other. The electric force is proportional to the charge of the ions, and the potential difference between the cell's inside and outside, called membrane voltage. The concentration of an ion type is in equilibrium when both forces equalize. The resulting membrane voltage is called *equilibrium voltage* and can be determined by the *Nernst equation* [5, 7, 22] (see [22] for a compact derivation of this equation). The equilibrium potential E_x for an ion of type x is given by [5]

$$E_x = \frac{RT}{zF} \ln \frac{[x]_\mathrm{o}}{[x]_\mathrm{i}}, \tag{1.1}$$

where R is the gas constant, T is the temperature in Kelvin, z is the valence of the respective ions, F is the Faraday constant and $[x]_\mathrm{o}$, $[x]_\mathrm{i}$ are the ion concentrations outside and inside the cell. It follows that the equilibrium potentials of sodium and potassium are $E_\mathrm{Na} = 62$ mV and $E_\mathrm{K} = -80$ mV if concentration ratios (outside:inside) of 10:1 and 1:20 are assumed.

Now, in the resting state, a neuron is permeable to not only one kind of ion but to several of them to different degree, given by a specific permeability. Here, the situation becomes more complicated. For example if the membrane is permeable to sodium and potassium one can expect the resting potential to be in between their respective equilibrium potentials. Quantitatively, the membrane resting potential can be estimated from the *Goldman equation* [23] if the ion concentrations and permeabilities are known.

Most neuron membranes are considerably permeable (in resting state) only to sodium, potassium, and chloride. The Goldman equation then predicts the membrane resting potential [5, 24],

$$V_{\text{rest}} = \frac{RT}{F} \ln \frac{P_{\text{Na}} \left[\text{Na}^+\right]_{\text{o}} + P_{\text{K}} \left[\text{K}^+\right]_{\text{o}} + P_{\text{Cl}} \left[\text{Cl}^-\right]_{\text{o}}}{P_{\text{Na}} \left[\text{Na}^+\right]_{\text{i}} + P_{\text{K}} \left[\text{K}^+\right]_{\text{i}} + P_{\text{Cl}} \left[\text{Cl}^-\right]_{\text{i}}}, \tag{1.2}$$

where P_x are the permeabilities of the membrane to ion x. The current of each ion type x through the membrane at a given membrane voltage V is given by the product of the driving force $(V - E_x)$ and the membrane conductance g_x for that ion:

$$I_x = g_x(V - E_x). \tag{1.3}$$

It follows that especially sodium and potassium ions have a constant, directed net movement across the membrane in the resting state. This would eventually lead to a breakdown of their concentration gradients and hence the resting potential. However, ion pumps in the neuron membrane actively transport ions against their electrochemical gradient to maintain the resting potential and therefore the cell's excitability [5]. A more formal description of the passive membrane properties is given in section 1.1.4.

1.1.3.2 The action potential

The action potential can be considered the basic unit of neural communication and computation [5, 7, 25, 26] (see section 1.3). As mentioned earlier, action potentials are very short reversals (approximately 2 ms) of the membrane potential such that the membrane voltage temporarily switches from negative to positive. Initiated in the axon hillock, they travel along the axon and eventually are communicated to other neurons through electrical or chemical synapses. Action potentials are also informally called *spikes*, a term that will be used frequently in this work.

Action potentials are caused by the active properties of the cell membrane among which the most important is the (region-specific) presence of voltage-gated sodium and potassium channels. These channels can change their configuration as a function of the membrane voltage and thereby change their membrane permeability for the respective ions. According to the Goldman equation (1.2), changes in the permeabilities mean that the steady-state of the membrane is changed. Similarly, Eq. (1.3) predicts that changes in

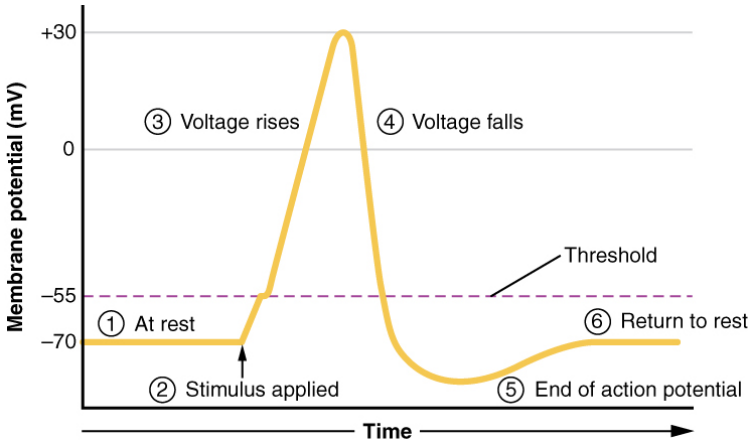


FIGURE 1.5: **Stereotypical shape and phases of an action potential.** At **(1)** the membrane voltage is at resting potential where the passive membrane properties determine the voltage dynamics, until the neuron is excited at **(2)**. If the excitation causes the membrane voltage to cross a specific threshold (here at -55 mV) the activation of voltage-gated sodium channels leads to a rapid increase of the sodium conductances. This results in a rapid influx of sodium ions and a fast rise and peaking of the membrane voltage in phase **(3)**. The inactivation of sodium channels lets the voltage drop back towards the resting potential in phase **(4)**. Additionally, the delayed opening of voltage-gated potassium channels pulls the membrane voltage to even more negative values (after-hyperpolarization). This causes voltage-gated sodium channels to close temporarily and no action potentials can be elicited during that time (refractory period). Eventually, the membrane potential returns to its resting state **(6)**. Image source: work by OpenStax Anatomy and Physiology, via Wikimedia Commons (Creative Commons).

the conductance result in changes of ionic currents. Based on this rationale the stereotypical shape of an action potential can be explained by the kinetics of voltage-gated sodium and potassium channels [5, 7]. The stereotypical course of an action potential is summarized in figure 1.5.

An action potential is typically generated at the axon hillock in a region with a high number of voltage-gated sodium channels ("spike-initiation zone"). Initially, the action potential is a very local membrane depolarization which then induces voltage differences along the membrane in the vicinity of the action potential. In a purely passive membrane, this leads to a slow and exponentially attenuated spread of the voltage along both directions of the axon. The presence of voltage-gated channels, however, has two effects: neighboring regions of the membrane can become depolarized enough to open sodium channels, and the action potential is "renewed". Moreover, the refractory period of recently depolarized membrane patches effectively allows only for forward propagation of the action potential. Larger axon diameter and axon myelination increase the velocity of the propagation of an action potential by decreasing the leak current through the

membrane [5, 7].

In most neurons, and in the scope of this work, spikes are elicited at the axon hillock close to the soma. In general, action potentials can also be generated at other parts of the neuron – for example, in most sensory neurons spikes are generated at the nerve endings [7]. Moreover, action potentials that are not based on sodium conductances are known to be important in neural systems. For example, calcium spikes play a crucial role in plasticity and are also found in dendrites [18].

1.1.4 Neurons as electrical circuits: integrate-and-fire models

In this section, I formalize the discussed neuroelectrophysiology and introduce the mathematical neuron models that I use in the remainder of this work.

In 1952, Hodgkin and Huxley published their seminal work on the squid giant axon in which they fully described the electrical signals inside the axon [27]. Without actually knowing about the presence of ion-selective channels they formulated a set of partial differential equations that successfully described the dynamics of sodium and potassium ionic currents. From then on, mathematical models of neuroelectrophysiology have been developed further in two different directions. On one hand, more detailed models with explicit channel and molecular dynamics and spatio-temporal variables have been used to describe neurobiological processes on very small scales [28]. On the other hand, various kinds of simplifications and generalizations of the Hodgkin-Huxley model have been devised that are mathematically more tractable or allow for the computational investigation of entire neuron circuits/populations (an overview of different effective neuron models can be found in [9, 29, 30]). One of the most popular classes among these models are *integrate-and-fire models*.

1.1.4.1 Leaky integrate-and-fire model

The Hodgkin-Huxley model incorporates both the constant ion conductances as known from the derivation of the resting potential equation (1.2) and the kinetics of voltage-gated ion channels that underlie the generation of action potentials. Taken together, these building blocks can be represented as the circuit diagram shown in figure 1.6. Here, the conductance g_{leak} comprises all passive conductances for sodium, potassium

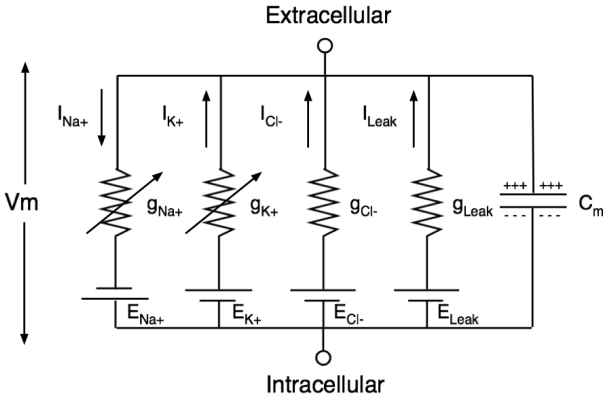


FIGURE 1.6: **Circuit diagram of neural membrane according to Hodgkin-Huxley model [27].** In a simplified picture the neural membrane can be seen as a capacitor with constant and voltage-dependent (denoted by an arrow) conductances g_x through which different ions can pass. Each ion type has its equilibrium potential E_x. As long as the membrane voltage V_m is small enough only the leak conductance and chloride conductance are relevant. The sodium and potassium conductances are voltage activated and determine the dynamics of the action potential. Image source: work by Nrets, via Wikimedia Commons (Creative Commons).

and other ions including the effect of ion pumps. The membrane itself is an insulator and therefore acts as a capacitor [22].

This picture can be further simplified by separating the passive and active membrane properties. In the passive *subthreshold* regime, where no voltage-gated ion channels are activated, the passive conductances g_{leak} and g_{Cl} can be summarized in one conductance. In practice, often the inverse of this conductance, the resistance R, is used. In this setting, the neuron can be seen as a parallel RC-circuit, an electric circuit comprised of a resistor with resistance R, a capacitor with capacity C_m, and a battery at V_{rest} (see figure 1.7). This model of the passive membrane properties is core to many integrate-and-fire neurons and is well known from electronic engineering [9]. In the so-called *leaky integrate-and-fire* (LIF) model, the kinetics of an action potential are approximated to be instantaneous and are elicited when the membrane voltage $V(t)$ reaches a threshold value Θ (the threshold for the opening of voltage-gated sodium channels). Formally, if a current $I(t)$ enters the neuron, either as the summed effect of synaptic activity (see section 1.1.5) or as direct stimulation, the dynamics of the membrane voltage are given by the following differential equation:

$$\tau_m \frac{\mathrm{d}V(t)}{\mathrm{d}t} = -(V(t) - V_{\text{rest}}) + R\ I(t), \tag{1.4}$$

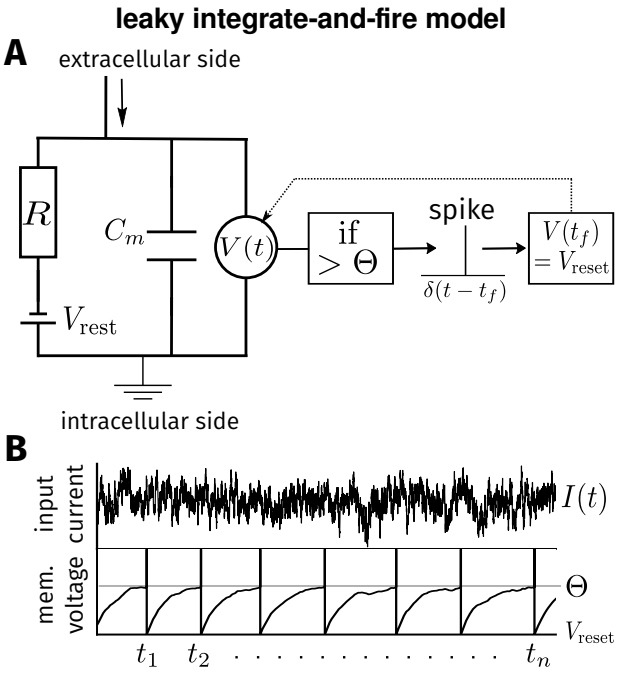


FIGURE 1.7: **Visualization of the leaky integrate-and-fire model. (A)** The circuit diagram underlying the model. In contrast to the more detailed model of figure 1.6, all passive conductances are summarized through the input resistance R (inverse of total conductance) and the resting potential V_{rest} is applied across the membrane. Input currents $I(t)$ that are generated at the synapses (see section 1.1.5) are integrated according to the dynamics of a RC-circuit [Eq. (1.4)]. In a leaky integrate-and-fire neuron the action potential is approximated by an instantaneous leap of the membrane voltage (due to sodium currents) whenever the membrane voltage $V(t)$ crosses the threshold Θ. Spikes are saved as δ-peaks at firing times t_f and the voltage is set to its reset potential V_{reset} after each spike. **(B)** Two example traces of input current (top) and resulting membrane voltage (bottom). In this work, the spike times t_i are the most relevant features of the voltage traces.

where $\tau_m = RC_m$ is the membrane time constant. The neuron is assumed to emit a spike as soon as $V(t_f) \geq \Theta$, where t_f is a spike time. Formally, spike trains are expressed as a sum of infinitesimally narrow, idealized spikes in the form of Dirac δ-functions [29]. A spike train $\rho(t)$ is thus given by

$$\rho(t) = \sum_i \delta(t - t_i), \tag{1.5}$$

where the sum is taken over all spike times enumerated by i. After a spike the membrane voltage is reset to a voltage V_{reset} that can be considered the membrane voltage at after-spike hyperpolarization. Generally, a refractory period t_r is introduced after each spike during which the membrane is not excitable (due to the insensitivity of sodium channels). The membrane voltage is then again determined by equation (1.4) until the next spike is emitted. This algorithm of the LIF model is depicted in figure 1.7.

The LIF model simplifies the electrophysiological description of a neuron to a single differential equation and does not require morphological information (such model neurons are called *point neurons*). This brings many advantages regarding the mathematical tractability and computational complexity of analyzing these models compared to more detailed models. Despite their simplicity, LIF neuron models have been shown experimentally to capture many electrophysiological properties of real neurons considerably well, in particular when spike rate adaptation effects are small [29]. Some examples of such studies focusing on cortical pyramidal neurons (see section 1.2 for more information on the cerebral cortex) are given in the following:

- C Teeter *et al.* [31]: The LIF model can overall explain ∼70% and in some cases ∼90% of the temporal variance in peri-stimulus time histograms. [Neurons from primary visual cortex (V1) of adult mice.]
- A Rauch *et al.* [32]: The LIF model allows for good fits of transfer functions, i.e. firing rate as a function of input current. [Regular spiking pyramidal neurons from rat somatosensory cortex; adding adaptive currents improved the performance.]
- from P Dayan & L Abbott [29]: Firing rates can be well predicted by LIF neurons in the steady-state after prolonged excitation or if only little short-time adaptation takes place [33, 34]. [V1 cat pyramidal neurons *in vivo*.]

1.1.4.2 Exponential integrate-and-fire model

The leaky integrate-and-fire model assumes that the generation of an action potential occurs instantaneously as soon as the membrane voltage crosses the threshold Θ. A Fourcaud-Trocmé *et al.* proposed an extension of the LIF model that takes into account the kinetics of the voltage-gated sodium channels (see section 1.1.3.2). The basic idea of this extension is that when the threshold is reached first a few sodium channels open and therefore the cell starts depolarizing. This, in turn, leads to the activation of more sodium channels and increased sodium influx, which again results in a positive feedback of this dynamics and the voltage rises exponentially. Formally, this *exponential integrate-and-fire* (EIF) model can be written as [35]

$$\tau_m \frac{\mathrm{d}V(t)}{\mathrm{d}t} = -(V(t) - V_{\text{rest}}) + \Delta_T \exp\left[\frac{V(t) - \Theta}{\Delta_T}\right] + R\,I(t). \tag{1.6}$$

Theoretically, a spike is emitted when the membrane voltage reaches infinity. However, practically, it is sufficient and necessary to set any value that is considerably higher than Θ as the point for spike emission [35]. Otherwise, the EIF model is implemented analogously to the LIF model as shown in figure 1.7. The exponential term in Eq. (1.6) implements the described dynamics of the sodium currents when $V(t) \geq \Theta$ and is negligible at membrane voltages below the threshold. The parameter Δ_T is called spike slope factor[3], or *spike initiation time*, and determines the sharpness of the rising phase of the action potential: the smaller Δ_T, the faster the exponential rise. In the limit $\Delta_T \to 0$, the EIF and LIF model are equivalent [36]. The validity of the EIF model has been supported by a number of experimental studies among which are:

- *L Badel et al.* [37, 38]: EIF neurons can be very well fitted to real neurons' *I-V* curves.[4] [Rat layer 5 somatosensory cortex, pyramidal neurons and interneurons.]
- *M Pospischil et al.* [39]: Features of stimulus induced peri-stimulus time histogram and features thereof can be well predicted by the EIF model. [Regular spiking neurons in guinea pig visual cortex.]

Qualitative differences between the LIF and EIF model occur when the membrane voltage undergoes high frequency changes, e.g., when the neuron experiences high-frequency stimulation or high firing rates are evoked. In these cases, the properties of the EIF model are closer to those of Hodgkin-Huxley type models and variants thereof [35, 36]. Therefore, the EIF model is used in this work to complement the results obtained for LIF neurons in high frequency ranges. A more detailed discussion of the response properties of the models and real neurons is given in section 1.4.2 (see e.g. [40] for a review).

Various other nonlinear integrate-and-fire models (containing nonlinear terms of V), among which are multidimensional variants with adaptive currents, and other classes of phenomenological neuron models have been devised in the past but are outside the scope of this work (see e.g. [41] and [40] for an overview).

[3] Technically, the spike slope factor is the inverse of the $I-V$ curve's[4] curvature at the threshold Θ [35].
[4] *I-V* curves depict the voltage-dependent ionic currents [37].

1.1.4.3 Poisson neurons

Integrate-and-fire models exemplify very reduced representations of spiking neurons. Nonetheless, obtaining spike times from these models requires knowledge of the neuron's input current and solving the respective differential equations for the membrane voltage (Eq. (1.4) and (1.6)). For a given input current $I(t)$, the integrate-and-fire models are deterministic and noise is normally incorporated in a stochastic part of $I(t)$ (cf. section 3.1.1). A type of neuron models that is even more reduced than integrate-and-fire models and that takes noise into account through a stochastic spiking process is the *Poisson neuron* model. This model is based only on the *instantaneous firing rate* of a neuron without explicit modeling of the input currents and their integration.

The firing rate is generally defined as the number of emitted spikes in a given time window divided by the length of that time window. However, for very small time windows this definition has to be refined because not more than one spike can be expected to lie within that window. In that case an instantaneous time-dependent firing rate $r(t)$ is defined as a trial-average over the spike trains $\rho(t)$ from Eq. (1.5) [29]:

$$r(t) = \frac{1}{\Delta} \int_t^{t+\Delta} \mathrm{d}t' \langle \rho(t') \rangle. \tag{1.7}$$

Here, $\langle\rangle$ denotes averaging over trials with the same (experimental) condition and the window length Δ should be taken as small as possible [29].

For Poisson neurons, it is assumed that the firing rate $r(t)$ is known; usually it is an implicit or explicit, parametrized function of an underlying stimulus (see section 1.3.1). The generation of spikes in Poisson neurons occurs according to the following scheme:

- Time is discretized in bins of size $\Delta \to 0$ that contain at most one spike [29, 42].
- At a given bin at time $n\Delta$, a spike occurs with probability $p = \Delta \cdot r(t = t_n) = \Delta \cdot r(n\Delta)$ and no spike is emitted with probability $1 - p$. Spiking in a single bin hence follows a Bernoulli process with probability p.

This Poisson spiking "mechanism" implies that all time bins are independent and spiking occurs uncorrelated (except for signal-induced correlations; see chapter 2). More formally, the probability of spiking in one time bin is not conditioned on the probability of spiking in any other time bin [29]. The Poisson neuron model owes its name to the

fact that, for a constant firing rate $r(t) = r$, the spike count within any given time window follows a Poisson distribution[5].

Moreover, Poisson neurons implement extrinsic and intrinsic noise on the level of spiking variability. Two important measures for spiking irregularity are given by the Fano factor and the coefficient of variation (CV) of the interspike interval (ISI) distribution. The Fano factor measures spike count variability as the ratio of the variance and the mean of the number of spikes in a given time window. The Fano factor of Poisson processes is 1. The ISI distribution for a spike train is the probability distribution of the times τ between two consecutive spikes. The coefficient of variation of a spike train is defined as the ratio of standard deviation and mean of the ISI distribution [29]:

$$CV_{\text{ISI}} = \frac{\sigma_\tau}{\mu_\tau}. \tag{1.8}$$

The mean μ_τ is the inverse of the overall firing rate, and σ_τ is a proxy for the width of the ISI distribution. For a given firing rate Poisson neurons have a CV of 1 [29], which represents a high spiking irregularity.

Poisson neurons are popular in theoretical neuroscience because they are mathematically tractable and sufficiently characterized by the time-dependent firing rate. On the experimental side, the Poisson model is supported by the fact that cortical neurons *in vivo* have been found to possess very irregular firing pattern with values of the CV and Fano factor close to 1 [43–49]. However, this is not uniformly the case in cortical neurons, and the firing regularity varies considerably across neurons and brain regions with values of the CV between 0.3 and 1.2 [50].

1.1.5 Synapses

Synapses are the points of contact between two neurons at which electrical signals are transmitted between those neurons. Accordingly, synapses are the essential sites of interneural communication. An average neuron forms and receives 1,000 to 10,000 synaptic connections [5]. In the brain, there are two types of synapses: *electrical synapses* and *chemical synapses*.

[5]The probability of finding n spikes within a window of length T is $P_T(n) = \frac{(rT)^n}{n!}\exp(-rT)$ [29].

Electrical synapses are direct physical connections between two neurons with a gap of only about 3 nm between the cell membranes and a complex of connexin proteins bridging the gap. This connection is called *gap junction*. Electrical synapses connect two neurons in a way that they can directly exchange ions (and smaller molecules). Neurons forming electrical synapses are hence electrically coupled. A change of membrane voltage in one neuron causes a voltage difference between the two coupled neurons and therefore, according to Ohm's law, induces a current flow across the gap junction. This kind of transmission of voltage changes is fast, reliable, and importantly also applies to voltage changes that are less extreme than the depolarization during action potentials [7]. In contrast to chemical synapses, the signal transmission through gap junctions is bidirectional. Electrical synapses are found in many brain regions and the peripheral nervous system. They are known to serve important functions in brain development and in synchronizing the activity of neurons [7, 51, 52].

Most interneuronal communication in the (adult) mammalian brain occurs through synaptic transmission at chemical synapses [5], which was first demonstrated by B Katz and colleagues [53]. As the transmission at chemical synapses is unidirectional, usually from axon terminal to dendrite, the involved neurons play different roles. The sending neuron is called *presynaptic* and the receiving neuron is called *postsynaptic*. The universal structure of a chemical synapse is shown in figure 1.8. The presynaptic membrane of the axon terminal and the postsynaptic membrane are separated by the synaptic cleft (approximately 15-25 nm wide [54]). In the presynaptic terminal reside small, spherical organelles that are called vesicles and contain the neurotransmitter molecules that function as chemical messengers in the synaptic transmission [4]. Figure 1.8 summarizes the steps involved in synaptic transmission (see also [4, 5, 7]).

The postsynaptic functionality of chemical synapses is determined by their neurotransmitter receptors. There are over 100 known types of neurotransmitters and even more receptors. The latter can generally be divided into two classes: transmitter-gated ion channels and G-protein-coupled receptors [4, 7]. If neurotransmitters bind to their respective transmitter-gated channels they induce a confirmational change in the channels and thereby make them permeable to certain ions. This permeability is usually less selective than for voltage-gated ion channels. Which ions types can pass through the channel, and the postsynaptic membrane voltage determine the effect of the transmitter on the postsynaptic neuron, explained by the Goldman equation (1.2). Hence, receptors

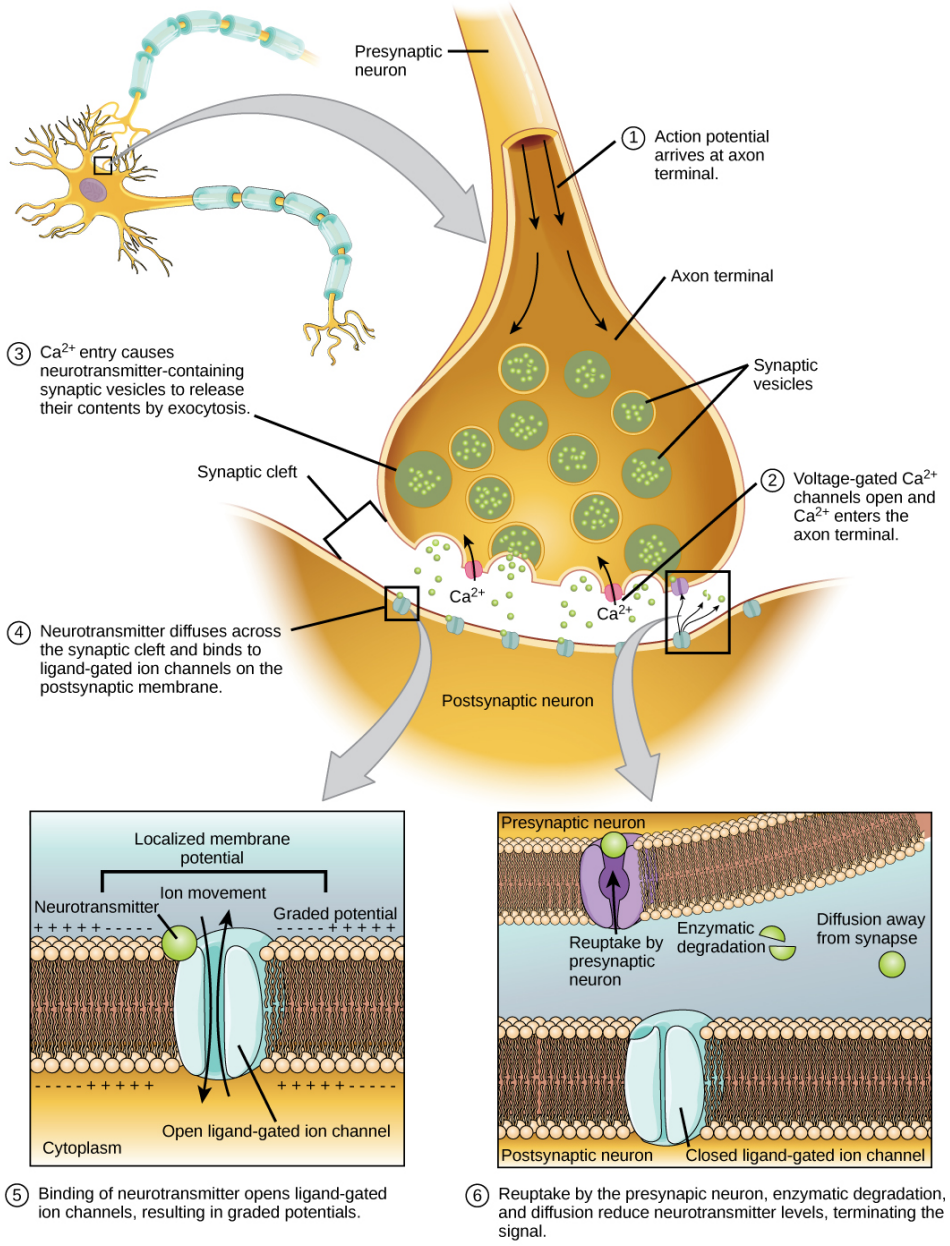


FIGURE 1.8: **The biochemical mechanisms of synaptic transmission.** The picture shows a sketch of a synapse as the contact point of presynaptic axon and postsynaptic dendrite. The components and steps involved in synaptic transmission are annotated. For a detailed discussion see text. Image source: work by CNX OpenStax, via Wikimedia Commons (Creative Commons).

that are permeable to Na^+ or Ca^+ facilitate a transient membrane depolarization, called *excitatory postsynaptic potential* (EPSP). These synapses are called excitatory. Channels that are permeable to Cl^- (or only K^+) have a hyperpolarizing effect and cause

inhibitory postsynaptic potentials (IPSP)[6]. In the central nervous system, the most important excitatory synapses are mediated by the neurotransmitter glutamate, the most important transmitters for inhibitory synapses are gamma-aminobutyric acid (GABA) and glycine [5].

The described process of generating postsynaptic potentials through transmitter-gated ion channels provides a way of fast synaptic transmission ($\geq$ 0.3 ms, but slower than electrical synapses) [5]. G-protein coupled receptors mediate slower, but long-lasting and more diverse synaptic effects. These effects include the activation of other ion channels or the initiation of more complex signaling cascades that can influence a diversity of intracellular mechanisms [7, 15].

An important feature of synapses is *synaptic plasticity*, the modulation of synaptic efficacy as a response to the activity of the pre- and postsynaptic neurons (and possibly other neurons). For example, synaptic plasticity is of high relevance for the neurobiological implementation of learning and homeostatic mechanisms in the brain [5]. In this work, synaptic plasticity is not essential; an introduction to the topic is, for example, available in [5].

1.1.5.1 Models of postsynaptic integration

Various models for the mathematical description of the process of synaptic transmission have been devised (see for example in [9] or [55] for a review). Here, the focus is on those models that are based on the resistor-capacitor circuits with gated conductances that also underlie the integrate-and-fire models (cf. Fig. 1.6). If a presynaptic terminal is activated and neurotransmitters are released the respective neurotransmitter receptors on the postsynaptic site become permeable for their respective ions (mostly sodium, calcium, and chloride). According to the Nernst equation (1.1), this means that a resting or *reversal potential* can be assigned to a synapse. Together with the postsynaptic membrane potential, the reversal potential determines the resulting current flow [Eq. (1.3)]. The effect of a single arriving action potential can be approximated to leave the difference between membrane potential and synaptic reversal potential constant. In this case we find a proportionality of synaptic conductance and synaptic current,

[6] These properties of excitation and inhibition hold around the resting potential and, according to Eq. (1.3), can be reversed at very high or low membrane potentials (see below).

$I_{\text{syn}}(t) \propto g_{\text{syn}}(t)$ [56]. The temporal evolution of synaptic conductances elicited by a presynaptic spike can well be formalized by a sum of two exponentials with different time constants, one representing a rising phase and the other representing the decaying phase [9, 56]. Often, the rise time of a synapse is very short compared to the relaxation time. Then, the postsynaptic current caused by a single action potential at time t_0 can be approximated as [56]

$$I_{\text{syn}}(t) \propto e^{-(t-t_0)/\tau_{\text{syn}}}, \tag{1.9}$$

for times $t > t_0$ and with relaxation time τ_{syn}. Sometimes, the postsynaptic currents are further simplified and represented as point-processes (δ-peaks), corresponding to the limit of infinitely short decay times:

$$I_{\text{syn}}(t) \propto \delta(t - t_0). \tag{1.10}$$

The effect of many arriving spikes at a postynaptic neuron is usually modeled as a superposition of the effects of single, independent spikes [9]. The resulting net postsynaptic currents can then be approximated as Gaussian processes, as explained in section 1.4.1.

1.2 Cerebral cortex and neural circuits

As mentioned earlier, this work focuses on cortical pyramidal neurons. In this section, I emphasize the high relevance of the cerebral cortex in neuroscience in general and thus for my work in particular. In section 1.2.1 I introduce fundamental properties of the structure and wiring of the cerebral cortex. In section 1.2.2 I show how these properties lead to states of balanced excitation and inhibition. These states provide a special setting for the mechanisms of signal processing that I analyze in this work.

The largest part of the human brain and that of some other primates are the cerebral hemispheres. They consist of the cerebral cortex, the underlying white matter, and some deep lying structures among which are the basal ganglia, amygdala, and hippocampus (sometimes the hippocampus is classified as allocortex and is then considered part of the cerebral cortex) [5, 7]. The cerebral cortex, or short *cortex*[7], is a distinct brain region

[7] Often the term cortex actually refers to the *neocortex* which is the cerebral cortex excluding hippocampus and olfactory cortex, and only exists in mammals [7]. In this work, too, cortex refers to neocortex if not stated otherwise.

and central to higher cognitive function such as consciousness or symbolic reasoning [4]. The processing of sensations, perceptions, voluntary movement, learning, and speech predominantly occurs in the cortex [7]. Arguably, the cerebral cortex distinguishes mammalian intelligence from that of other species, and understanding the computational properties and mechanisms within the cortex is thus one of the most prominent quests in neuroscience [4]. For the same reason, the cortex is the most studied brain region and also the basis for the neurobiological modeling in this work.

The cortex is about 2 to 4 mm thick and can be divided into distinct functional areas that can often be distinguished by cytoarchitecture (see e.g. [7] for a good overview on neuroanatomy). These areas are classified as primary, secondary, or tertiary areas depending on their level and function within the information-processing pathways. Many of these areas along sensory pathways are arranged topographically: neighboring groups of cells are receptive to neighboring stimuli in the respective sensory space, e.g. neighboring cells in V1 (primary visual cortex) represent inputs to neighboring patches of the visual field. Primary sensory areas receive a majority of the incoming signals from the thalamus, which in turn is only one or a few synaptic connections distant from the sensory receptors. Secondary sensory areas receive their inputs mostly from primary areas and are unimodal. They present more abstract features of the sensory input. Tertiary areas are multimodal and integrate complex information and associations from different sensory systems. The situation is different in motor areas of the cortex, here primary motor cortex is the final stage of processing and sends signals to the periphery to initiate movement [5, 7].

1.2.1 Circuit structure

Even though the different cortex areas are ascribed very different and diverse functions, the micro-structure of the cortex is remarkably similar across areas [4]. Across species and areas, the neuron density in the cortex is around 100,000 cells under a squared millimeter surface, irrespective of the cortical thickness, except for visual cortex where it is about twice as large [58, 59]. The most important, and sometimes defining, uniform property of the cortex is its layered structure. As depicted in figure 1.9, the cortex exhibits six distinct horizontal layers with characteristic neuron types and myelination.

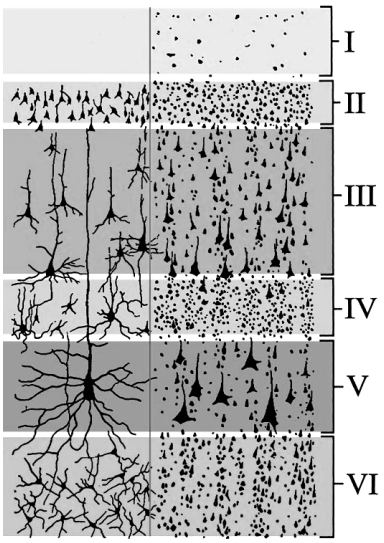


FIGURE 1.9: **Illustration of the layered structure of the cerebral cortex.** The cerebral neocortex has six horizontal layers of varying thickness and neuron density. The left column shows layer typical cell morphologies of excitatory neurons as they are seen in a Golgi stain. The right column depicts the cell body density in each layer according to a Nissl stain. The layers are by names: I = molecular layer; II = outer granular layer; III = outer pyramidal layer; IV = inner granular layer; V = inner pyramidal layer; and VI = multiform layer. Modified from [57].

The basic properties of these layers are as follows (going from pial surface down to white matter) [4, 5]:

- I. *molecular layer*: contains mostly dendrites of cells from deeper layers that make connections to neurons of other cortical areas.
- II. *outer granular layer*: contains mainly small pyramidal cells in high density.
- III. *outer pyramidal layer*: makes up almost one third of the cortex' thickness and contains large pyramidal cells; the axons of layers II and III mostly project locally to nearby neurons or to other cortical areas
- IV. *inner granular layer*: similar to layer II in that it contains small and densely packed cells; it is the major receiving layer of thalamic projections and is therefore most prominent in primary sensory areas
- V. *inner pyramidal layer*: contains pyramidal neurons that are typically larger than that in layer III; these neurons make the major outward projections to other cortical regions and subcortical structures
- VI. *multiform layer*: has a heterogeneous cell composition and makes connections to other cortical areas

The definition of the layers is based on the location of the cell bodies, however, also the location of dendrites of specific neurons is characteristic for the different layers. For example, layers I, II, III contain dendrites from neurons in layers II, III, V, VI. Accordingly, the input to neurons in the different layers is also specific. In primary visual cortex the output from the thalamus is projected to layer IV neurons [5].

The principal cell in the cortex is the pyramidal neuron[8], which make up about 70% of all neurons in the cortex and are prevalent in layers II, III, and IV. Pyramidal neurons are the major cortical output and can, therefore, be considered the relevant "readout" for the computations occurring in the cortex. At the same time, lateral intracortical connections make pyramidal neurons the most important source for excitatory input in the cortex [4]. This unique role of pyramidal neurons in cortical processing is the reason why they are the point of reference in this work whenever cell properties are concerned.

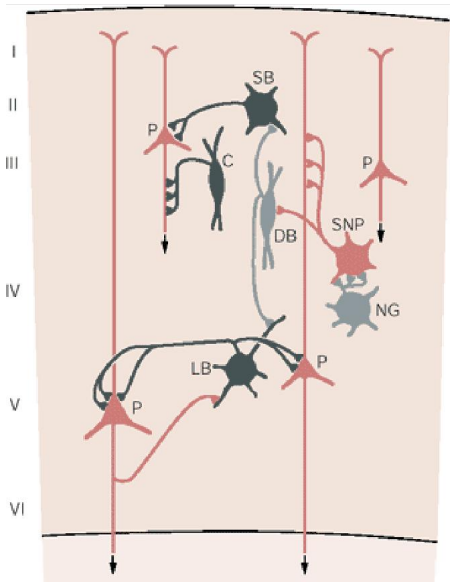


FIGURE 1.10: **Inhibitory interneurons are crucial for orchestrating the neural activity in neocortical circuits.** Here, different types of GABA-ergic interneurons (dark gray) and other putative types (light gray) are shown, including their connectivity pattern. Chandelier cells (C) target the axons of pyramidal cells (P). Small and large basket cells (SB and LB) terminate at the somata of pyramidal cells in layers II/II and IV. Double bouquet cells (DB) terminate in two distant tufts of other inhibitory cells. Moreover, excitatory spiny non-pyramidal cells (SNP) excite pyramidal neurons and are themselves targeted by neurogliaform (NG). From [5].

Nonetheless, not only pyramidal neurons are relevant in cortical signal processing, and more than 40 different types of cortical neurons have been identified [5]. Other excitatory neurons such as *spiny stellate* cells, concentrated in layer IV and abundant in

[8]Pyramidal neurons owe their name to their conical shape. They possess a basal dendritic structure and an apical dendritic tuft more distant from the soma (see figure 1.3 for a simplified sketch) [4].

visual cortex, or *bipolar* cells serve critical functions. Moreover, inhibitory neurons are crucial parts of cortical circuits. They have an obvious homeostatic purpose: by inhibiting excitatory neurons, they can prevent catastrophic runaway effects of excitatory feedback. Interneurons have other important functions, e.g., for the generation of neural oscillations and for establishing the excitation-inhibition balance discussed in the next section [60, 61]. In the cortex, GABA-ergic interneurons make up 20% of all neurons and are located in all layers. Figure 1.10 shows how different types of interneurons interact with the principal (pyramidal) neurons. Moreover, and not shown in the figure, excitatory interneurons exist as well and introduce a high degree of recurrent feedback in the wiring. Neurons in the neocortex are often organized in columns perpendicular to the surface and traversing the layers. Cortical columns are local processing modules and neurons within a column tend to have similar response properties. The resulting cortical circuits are specialized for different functions but are believed to share basic properties and to give rise to canonical computations in cortex [4, 5, 62].

1.2.2 Excitation/inhibition balance

In the previous sections, I introduced different neuron models and described how they integrate current inputs (section 1.1.4), how postsynaptic currents are generated from arriving action potentials (section 1.1.5), and how neurons are connected in cortical circuits (previous subsection). The net effect of this neuronal micro-architecture results in characteristic currents that pyramidal neurons receive in active cortical circuits, the details of which I discuss in this section.

As mentioned in the introduction of Poisson neurons in section 1.1.4.3, tonic firing cortical neurons often possess a high variability in their spiking pattern with coefficients of variation around 1 or above in different species, tasks and stimulation protocols [43] (e.g. in V1 of macaque monkeys during visual stimulation [63], or in V5/MT during a motion discrimination task [64]). This high variability could not be explained by a stochastic arrival of independent EPSPs only (EPSP driven activity had been the notion of inputs to cortical cells because the number of excitatory synapses in cortex outnumbers this of inhibitory synapses by a factor of around 5.5) [5, 43, 65]. M-N Shadlen and W-T Newsome proposed that a balance of excitatory and inhibitory inputs leads to membrane

voltages that fluctuate around a subthreshold value and only occasionally and stochastically depolarize the neuron enough to elicit an action potential. They showed that the resulting firing variability would be in agreement with the experimental findings [44, 46]. Additionally, C van Vreeswijk and H Sompolinsky demonstrated theoretically that excitation/inhibition (E/I) balance and firing irregularity emerge naturally from recurrent networks of excitatory and inhibitory neurons if these are connected randomly, sparsely and with relatively large synaptic strengths [66, 67]. This E/I balance is generated when a neuron receives a high number of irregular, uncorrelated excitatory and inhibitory inputs that cancel each other out on average. The balance is established despite the higher abundance of excitatory synapses because interneurons make stronger synapses and fire at higher rates. Recurrent/feedback connections activate interneurons in cortical networks such that inhibitory input currents are proportional to their excitatory counterparts [61, 68]. However, this balance only holds on a slow time scale at which the rates of the presynaptic neurons change. On a fast timescale, rapid fluctuations such as single presynaptic events are not balanced, and the net current to the cell is dominated by these fluctuations, leading to irregular Poisson-like firing. This type of E/I balance where fast fluctuations of the excitatory and inhibitory inputs are uncorrelated is called loose. In some cases, excitatory and inhibitory synaptic currents are strongly correlated on very fast time scales such that even fluctuations remain balanced, this kind of balance is referred to as tight [69].

Electrophysiologically, E/I balance is investigated by measuring the excitatory and inhibitory conductances $g_e(t)$ and $g_i(t)$ intracellularly. A balance of excitation and inhibition is given when the ratio of these conductances, which determines the cells membrane voltage (see e.g. in supplemental material of [70]), remains constant[9]. As it is not possible to measure both conductances simultaneously, a way to determine the ratio g_e/g_i is to record them separately in different trials by clamping the cell at different voltage levels. This type of experiments has confirmed the E/I balance hypotheses, e.g. in ferret prefrontal cortex [71, 72], rat auditory cortex [70], or mouse V1 [73]. In order to compare g_e and g_i on faster time scales trial averaging is not appropriate. Therefore, paired recordings of neighboring neurons have been used because they are known to receive highly correlated inputs [74]. These paired recordings showed evidence for a tight E/I

[9]The role of g_e and g_i becomes clear when the ionic conductances in the circuit of figure 1.6 are replaced by these synaptic conductances. The conductance-based circuit model for the membrane voltage is then given by [70] $C_m \frac{dV}{dt} = -g_{\text{leak}}(V(t) - E_{\text{leak}}) - g_e(t)(V(t) - E_e) - g_i(t)(V(t) - E_i)$.

balance in cortex [71, 72, 74–76] and hippocampus [77]. Often inhibition trails excitation by a few milliseconds and thereby opens small temporal windows for firing before inhibition suppresses the cells [68, 78–80].

1.2.2.1 Functions of balanced networks

The role of E/I balance for neural computation and coding is believed to be multifarious, and the picture is complicated by the fact that different characteristics of E/I balance (e.g. loose and tight) have been observed [68, 69]. An evident function of E/I balance is to prevent a runaway of excitation that can lead to epileptic neural firing and seizures [81]. Finely tuned inhibition and E/I balance contributes to gain control, sharpening of the tuning curves of sensory neurons, and promoting the precision of spike timing [68, 70, 79, 80, 82]. The fluctuation driven regime of E/I balance also increases the responsiveness of cortical neurons and can lead to fast linear response functions (see section 1.3.1.3) [66, 67, 83–85]. It has further been proposed that E/I balance helps to keep neurons independent (uncorrelated) and therefore increase overall information transmission (cf. section 1.3.1.5) [61, 66]. Recent theoretical work suggests that balanced networks can implement optimal, noise-robust neural selectivity [86] and neural circuits for solving differential equations via inhibitory feedback loops [87]. At a larger scale, the disruption of a precise E/I balance has been linked to psychiatric disorders such as autism and schizophrenia in mammals and humans [88–90], emphasizing the importance of maintaining the balance.

1.3 Neural coding and information theory

The brain is an information processing system that works by electrophysiological signaling between neurons. Naturally, a basic question in neuroscientific research is *how* specific information is encrypted in the electrophysiological activity of neurons. This, essentially, is the question for the *neural code*. On the other hand, information theory provides methods to investigate the amount of *neural information transmission* quantitatively without explicitly asking for the specific encoding schemes in play. In this

section I introduce the basic concepts of neural coding (section 1.3.1) and neural information theory (section 1.3.2). Thereby, I present the mathematical representation of those concepts that are relevant for the specific questions of this work in more depth.

1.3.1 The neural code

As pointed out earlier, spikes (action potentials) are the basic units of neural communication[10]. Colloquial speaking, spike sequences are the language of the brain. As a consequence, whatever a neuron is representing and sending to other cells must be ciphered in the spike sequences it emits. The question *how* specific information is *encoded* in the spiking activity of a neuron, or a neuron population, frames the problem of finding the neural code and poses one of the fundamental challenges of neuroscience [25, 26, 42] (a critique of the concept of the neural code has been brought up by R Brette recently [92]).

Historically, important early contributions to the understanding of the neural code were made by E D Adrian (summarized in [93]). He demonstrated that action potentials are stereotyped, all-or-nothing events and therefore established that the arrival times of spikes, rather than other features of an action potential, provide information [26, 94]. This is a simplification, and the shape of action potentials can carry relevant information as well [91]. However, these effects are mostly thought to be of less importance [25, 26, 42]. Adrian further provided experimental evidence for the neural code being used in stretch receptor neurons that are embedded in muscles. He found that the number of spikes that are emitted by these neurons in a given time window increases with the weight that is used to stretch the muscle. This is an example of a rate code, which is discussed in more detail below. After a while, the neural response to the stimulus (stretch) declines, a general physiological phenomenon called adaptation. Adrian showed this aspect of neural coding, too, and thereby provided a link between neural activity and mental percepts, namely the gradually decreasing awareness of a continuously applied stimulus.

Even though the fundamentals of neural coding that Adrian and colleagues established are still prevailing, many questions concerning the neural code have remained open or

[10] It has been argued that action potentials have advantages over continuous signal transmission because they can easier be transmitted over long distances by means of active membrane processes (see section 1.1.3.2). However, graded voltage responses exist, e.g. in bipolar cells of the retina where signaling occurs over very short distances [5, 25, 91].

debated. They involve the following dichotomies: Do the exact spikes times matter in the neural code, or is the number of spikes per time relevant? Do spikes carry information independently, or are spike time correlations relevant? Is the information encoded in populations of neurons the summed information carried by each neuron, or are emergent properties crucial for population encoding?

1.3.1.1 Noise requires a probabilistic concept of neural coding

Formally, the problem of finding the neural code is equivalent to finding the function $f = \rho(s(t), t)$ that describes how a given time-dependent signal/stimulus[11] $s(t)$ is encoded in the spike train $\rho(t)$. The inverse function f^{-1} then allows inferring (decoding) the stimulus from an observed spike train $\rho(t)$. However, a general experimental observation is that the spike sequences that result from a repeated presentation of the same stimulus are not identical: randomness, or noise, is essential to neural systems. Sources of noise exist at many different length scales and exist in single neurons as well as in neuron populations. Examples on the cellular level are thermodynamic noise in sensory receptors, stochastic processes involving protein production, fusion of synaptic vesicles, or binding to receptors, electric noise due to random fluctuations in ion channel opening, or synaptic noise introduced by spontaneous miniature postsynaptic currents (see A A Faisal *et al.* for a review [95]). On the population level, stimulus-unrelated changes of the global network state, adaptation and predictive coding, and high sensitivity to the initial conditions of the network dynamics have been proposed as sources for trial-to-trial variability [26, 67, 96, 97]. Due to neural noise and the associated trial-to-trial variability in the spike trains, stimulus-response functions (such as f) have to be conceptualized as probability distributions $f = p(\{t_i\}|s(t))$. This denotes the conditional probability to observe a spike train $\rho(t) = \sum_i \delta(t - t_i)$ with spike times $\{t_i\}$[12] in response to a stimulus $s(t)$. However, exact computations of the distributions $p(\{t_i\}|\cdot)$ over the full space of all possible spike times $\{t_i\}$ are not feasible in practice due to the extremely high number of possible states within that space (see section 1.3.2). Therefore, the neural code is often described in terms of other quantities that can be derived from the observed spike times $\{t_i\}$ but require fewer data to be evaluated, e.g. firing rates [25, 29, 98].

[11] In this work, the terms signal and stimulus are synonymous if not stated otherwise. Often they refer to sensory stimuli but generally denote the quantity to be encoded.

[12] The set $\{t_i\}$ contains all times $t_1, t_2, \ldots, t_N$ where spikes occurred.

1.3.1.2 Rate codes

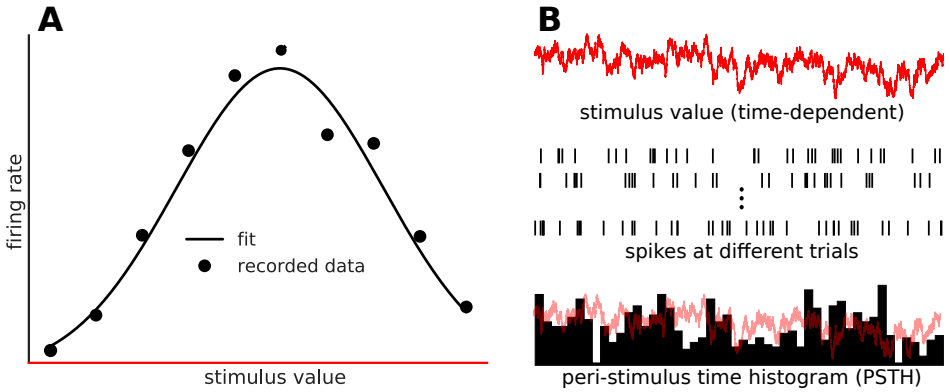


FIGURE 1.11: **The concept of firing rate codes. (A)** In tuning curves, the firing rate of a neuron is plotted as a function of the presented/evoking stimulus. Thereby, the firing rate is obtained by trial averaging of the number of spikes within a time window after stimulus onset. **(B)** A dynamic stimulus (top) is presented repeatedly and the spiking responses are recorded (center). Noise is visible as trial-to-trial variability. Binning the summed responses yields the peri-stimulus time histogram (PSTH). The stimulus is shown as overlay for comparison. The temporal width of the bins is chosen manually and can critically influence the correlation of stimulus and PSTH. The data here were arbitrarily generated for demonstration.

An intuitive and important feature of spike trains that can be computed from the spike times $\{t_i\}$ is the firing rate, or firing frequency, broadly defined as "spikes per time". The notion of a rate code is that the relevant stimulus is encoded in a neuron's firing rate. Thereby, the time scale over which a firing rate is computed and linked to a stimulus is not uniquely specified and situative (see below) [25, 29]. From the perspective of decoding, i.e. inferring the stimulus from neural activity, rate coding implies that (full or partial) knowledge of the stimulus can be gained from observed firing rates. Here, I focus on single neuron encoding. The concept is extended to neuron populations in section 1.3.1.6.

For slowly changing or static stimuli, rate codes are well established and often represented by tuning curves, which visualize the correlation of stimulus values and the resulting firing rates (see figure 1.11A). This approach is the basis of many famous results for feature selectivity in sensory and motor neurons, including the findings of Adrian [93], the discovery of orientation selective neurons in cat V1 by Hubel & Wiesel [99], and movement direction selection in neurons of monkey motor cortex [100].

Technically, these rate codes are based on determining firing rates in terms of spike

counts (number of spikes) within windows of typically hundreds of milliseconds[13]. In natural conditions, relevant stimuli often change at time scales that are shorter than these windows. Moreover, on the decoding level, it is unclear whether the biophysical properties of neurons allow for the long integration times that would be required [25, 42]. Therefore, often time-dependent firing rates are used instead and are based on using much smaller time windows that allow incorporating fast changes of the firing dynamics in response to fast changing stimuli. The firing rates within each bin are obtained by averaging the number of spikes for different trials of repeated presentation of the stimulus (equivalent to Eq. (1.7) with small but finite bin size Δ)[14]. The resulting histograms are known as peri-stimulus time histograms (PSTH, see figure 1.11B). In an experimental setting, a PSTH can always be generated if a sufficiently higher number of spikes/trials is recorded. However, the choice of Δ is not unique and may crucially influence the resulting relation of PSTH and the underlying stimulus [25, 29].

The time-dependent firing rate can be derived from a PSTH, however, in a given trial a single time bin contains a variable number of spikes (mostly zero or sometimes one spike if Δ is very small). The experimentally obtained time-dependent firing rate can be interpreted as describing the (stimulus-dependent) expected number of spikes in a given time bin. This probabilistic picture accounts for the previously discussed trial-to-trial variability as a consequence of noise. Because spike trains consist of discrete events rather than continuous firing rates, the rate coding framework needs to be extended to be able to explain how specific spike trains arise from a given stimulus. This is commonly achieved using *generative spiking models* in which a given firing rate is mapped to a probability distribution for the number of spikes in the respective time window. Spikes are then generated by sampling from this distribution. The spike count distribution of the model determines the statistical properties of the resulting spike trains beyond the average number of spikes such as the Fano factor or spike autocorrelation properties [25, 40, 101]. Popular choices are the Poisson model (see section 1.1.4.3) or γ-distribution generators that allows to realize more structured ISI distributions as found in experiments [29, 50, 102]. However, temporal structure in the spike trains is also induced by temporal

[13] In these cases, firing rates are obtained as spike count rates rather than instantaneous firing rates. The spike count rate is formally obtained when the time window Δ in Eq. (1.7) is kept at rather large values [29].

[14] Often sliding window filters are first applied on the spike trains in order to obtain smoother rate trajectories [29].

structure in the rate trajectory itself. The effect of both types of temporal structure is discussed thoroughly in chapter 2.

1.3.1.3 Linear rate models

Remark: *The figures and some elements of the text in sections 1.3.1.3 and 1.3.1.4 are also contained in this review article that Dr. Tchumatchenko and I published in 2017:*

- Herfurth, T. & Tchumatchenko, T. (2017). How linear response shaped models of neural circuits and the quest for alternatives. Current Opinion in Neurobiology, 46, 234-240. [1]

A model of a neural rate code which can make predictions beyond observed data requires knowledge of the function f_r that maps a signal to a firing rate [$f_r : s(t) \rightarrow r(t)$]. Generally, the firing rate can be a complex, nonlinear function of the input current. Following the concept of Taylor series, the time-dependent firing rate can be represented by a similar series expansion called *Volterra expansion*. In a Volterra expansion, the firing rate at a given time is assumed to be a function not only of the signal value at the same time but of signal values at earlier times, too. This makes the firing rate a functional of the signal, and the expansion is carried out accordingly. If only the first order of the signal $s(t)$ is taken into account in the Volterra series one obtains the following linear approximation to the firing rate $r(t)$ [29, 103–105]:

$$r(t) \approx \nu_0 + (L * s)(t) \equiv \nu_0 + \int_0^\infty \mathrm{d}t' L(t') s(t - t'). \tag{1.11}$$

The term ν_0 denotes the firing rate during spontaneous activity, due to noise or background activity, when no signal is present. The mathematical operation denoted by "$*$" is called convolution, L is called *Wiener kernel*, or linear response function and is a weighting factor that determines how the value of the stimulus at time $t - \tau$ affects the firing rate at time t [106]. The linear response function describes the linear filtering properties of the system under consideration. In particular, the amplitude of the Fourier transform of L specifies how strong each frequency that is contained in the signal is present in the response [30]. The nature of the linear rate approximation Eq. (1.11) implies that it becomes asymptotically better for decreasing signal strengths and is exact in the limit $s(t) \rightarrow 0$ [1].

For many neuron models, as those introduced in section 1.1.4, the linear response function can be derived analytically [107, 108] (see [40] for a review). The response functions for the LIF model are used in chapter 3. In experimental studies, linear response functions can be obtained based on recorded spikes by calculating the ratio of the correlation of the signal and rate trajectories on the one hand and the autocorrelation of the presented signal on the other hand [29, 30]; alternative methods exist, see e.g. in [109–111]. Various experimental studies have assessed the linear response functions for various neuron types [17, 112–115], as I discuss in sections 1.4.2 and 3.2. The notion of a linear relationship between stimulus and response has been used to derive receptive fields through linear reverse correlation[15] of the stimulus trace and the corresponding rate dynamics [29, 101, 116]. This has offered important insights into the receptive fields of retinal ganglion cells as well as visual and auditory cortical neurons [117–119], including the discovery of Gabor-like filters in V1 ("ON-OFF") neurons [120]. As for decoding, the same concept allows to determine the *spike-triggered average*, the linearly estimated average trajectory of the stimulus preceding a spike. The notions of spike-triggered averages and receptive fields illustrate that linear response functions can be used to understand what stimulus features a neuron filters/selects [29, 30, 101, 121, 122]. Even in situations where the assumptions of weak signals are not valid, the linear response function has been proven useful [113–115, 123]. Apart from modeling firing rates the explanatory power and interpretability of linear response functions and models has been demonstrated repeatedly, e.g. in the context of synchronous activity [104, 105, 124, 125], brain oscillations [126], and neural information transmission [127–130] (see figure 1.12 and Herfurth & Tchumatchenko 2017 [1] for a review).

1.3.1.4 Complex information processing requires nonlinear (rate) encoding

The linear response concept has apparent drawbacks: nothing in Eq. (1.11) prevents the firing rate from becoming negative or arbitrarily high if the stimulus is chosen accordingly, both cases are non-biological. Moreover, the concept is generally most useful if inputs are small relative to background activity and if inputs can be considered stationarity. Accordingly, in many situations, the linearity assumption turns out to be not

[15] These receptive fields are spatio-temporal extension of the linear response function. The underlying reversed correlation method is tightly linked to the linear response function (and spike-triggered average) calculation [29].

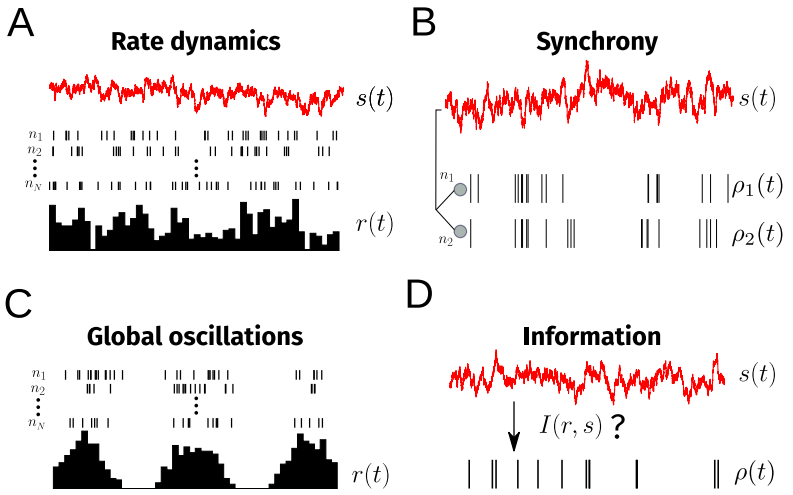


FIGURE 1.12: **Four exemplary phenomena that are addressable using linear response functions. (A)** Firing rate $r(t)$ (shown as PSTH) in response to a dynamic stimulus $s(t)$. The rate is obtained by averaging neural responses over different trial/neurons $n_1, \ldots, n_N$, and can be approximated by Eq. (1.11). **(B)** Synchronization of pairs of neurons. Two neurons receiving correlated input $s(t)$ exhibit partly synchronized spiking responses $r_1(t)$ and $r_2(t)$ (cf. equation (17) in [105]). **(C)** Emergence of self-sustained rhythmic activity in networks. The constituent neurons $n_1, \ldots n_N$ oscillate such that the population rate $r(t)$ is rhythmic, and its frequency can be predicted by linear response; cf. Eq. (6) in [126]. **(D)** Information transmission from input stimulus to spiking output. A neuron that responds with a spike train $\rho(t)$ to a signal $s(t)$ transmits information about that signal which can be characterized by the mutual information $\mathcal{I}(r, s)$ (see section 1.3.2.2). A lower bound for $\mathcal{I}(r, s)$ has been derived via linear signal reconstruction. Modified from Herfurth & Tchumatchenko 2017 [1].

accurate enough, and higher order, nonlinear aspects of the neural responses play a significant role. For example, the degree of decorrelation in the retinal output in response to correlations in the visual input has been explained by the contribution of nonlinear neural transfer functions [131]. On the cortical level, simple and complex cells in the visual cortex are most precisely described if nonlinearities are incorporated in the respective models [29, 119, 132–134]. Similar results that emphasize the importance of nonlinear input-output relations have been obtained for other cortical areas as well [135–138].

In general, different theoretical considerations suggest that any non-trivial processing operation needs nonlinearities. Functions such as contrast-invariance, classification, adaptation, or stimulus-dependent normalization all require some nonlinear components for their implementation [139, 140]. For layered networks, it has been shown that if at least one layer exhibits nonlinear transfer functions, then the network can approximate all continuous functions arbitrarily well as long as the number of units in the network is high enough [141, 142]. Furthermore, the chaotic dynamics found in neural networks can only be explained by nonlinear interactions [1, 140, 143–145].

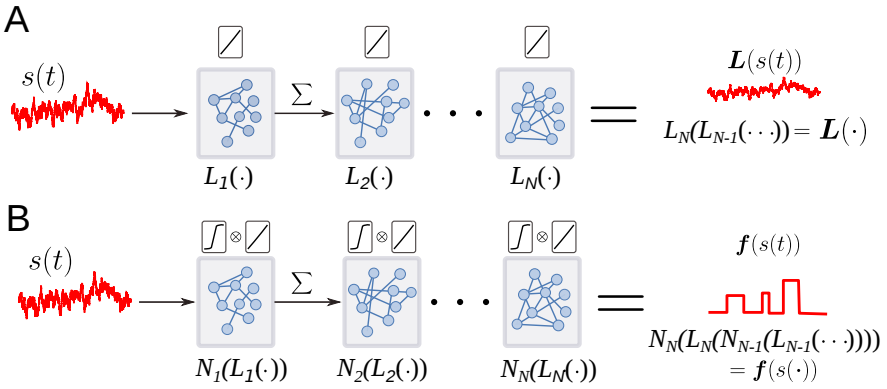


FIGURE 1.13: **The inclusion of nonlinear components is necessary for complex computations. (A)** A combination of linear transformations remains a linear operation. **(B)** Including nonlinearities, such as thresholding, in a feed-forward chain enables the representation of complex functions, e.g. classification of objects, that cannot be realized through the scheme of (A). Those nonlinearities are for example implemented in generalized linear models (GLM, see text). From Herfurth & Tchumatchenko 2017 [1].

Formally, nonlinear input-output relations are most commonly introduced by means of linear-nonlinear (LN) models [29, 40]. In these models the input/stimulus is first filtered by a linear feature detector, e.g. the linear response function, effectively reducing the relevant stimulus dimensionality [101]. The eventual firing rate of a neuron is then obtained by applying a static nonlinear function onto the filtered input that, amongst others, implements thresholding and saturation in the spiking rate. In mathematical terms, the LN models read

$$r(t) \approx \nu_0 + [N\,(L * s)](t), \tag{1.12}$$

where $N(\cdot)$ is a static nonlinearity that can be parametrical or derived from optimization procedures [101] (cf. figure 1.13). As mentioned earlier, spiking models then generate spikes stochastically from $r(t)$. LN models can be fitted based on experimental data or analytically for some neuron models. Often, they accurately match electrophysiological properties of real neurons [146] (see [40] and [101] for a review on LN models). A special variant of LN models are generalized linear models (GLM) that extend the former by incorporating spiking history, e.g. adaptation or refractory effects [147–149], as a modulator for the firing rate. GLM have been very successful in reproducing the variety of spiking patterns found in real neurons and other nonlinear phenomena [101, 150–153]. Apart from the discussed models, numerous other nonlinear rate models exist [101, 154]. Moreover, advances in modern machine learning have put forward (partly) model-free and data-driven approaches to investigate neural coding [155, 156]. I present an extended

discussion of linear and nonlinear models in neuroscience in section 4.2.

Open questions. Linear (rate) approximations have been proven to be a very helpful methods in theoretical neuroscience. However, they can be expected to lose utility if signal strengths become too large. The regimes in which such linear theories are valid are not generally known. In chapter 3, I explicitly address the following question:
How well do linear approximations of neural information transmission perform under different given parameters? And how do input and neuron properties shape the linearity of information encoding?

1.3.1.5 Temporal codes

At first sight, rate coding implies that only the number of spikes in a given time window is representative of the encoded stimulus and that the exact timing of the spikes does not matter. The notion of temporal codes is that the times of spikes (or high-frequency rate fluctuations) are the relevant observable that encodes a stimulus. A debate about whether spike rates or spike timings are the fundamental units of neural coding has been lingering in the neuroscience community [25, 26, 29].

The general importance of precise spike timing has been demonstrated in various studies [157–159]. For example, the synergetic encoding effect of precisely timed spikes has been shown in monkey cortex [160, 161], other studies showed that rates alone are not sufficiently informative about stimuli to account for observed behavior [162, 163], and coding through relative spike latencies was proven to exist in the retina [164, 165], and across sensory cortices [166–168]. The observation that the occurrence of single spikes in cortex can have significant impact on perception and motor output provides further evidence for a non-trivial role of spike timing in neural systems [169–172]; along with the high and reproducible precision of spike times in the fluctuation driven regime [173], and the importance of spike timing for synaptic plasticity (*spike-timing dependent plasticity*) [174]. Moreover, it has been emphasized that in many natural situations sensory neurons have to interpret spiking input within very small time windows with only a few spikes. This impairs the possibility of reliably estimating firing rates and supports encoding in spike times [25, 29]. More generally, neurons communicate with each other in terms of spikes rather than rates, and conceptually rates dismiss information contained

in the spike times. However, this does not principally exclude that spikes are a binary implementation of a neural code that is algorithmically based on firing rates [26, 175].

In the rate coding framework, the time scale over which rates are determined is a free parameter. For that reason, a clear distinction between rate and temporal codes is not possible: if the temporal resolution in a rate code is increased sufficiently enough, the representations of rate and temporal code converge. Similarly, a rate code can be considered a temporal code with some precision invariance. Often, a neural code is classified as temporal if the changes of the rate are fast such that only very few spikes are generated before considerable rate changes occur. Another identification of temporal codes can be made if the dynamics of the stimulus are known. Then, a code can be named temporal if the coding-relevant precision of spike times is shorter than the fastest stimulus frequencies [29]. It is apparent that these distinctions between rate codes and temporal codes are not stringent.

Of particular interest in this work are temporal codes in which the temporal correlations of spikes encode stimulus information, e.g. in the inter-spike interval (ISI) distribution or moments thereof [9, 25, 176, 177]. In many cases, second order statistics of presynaptic spiking have specific effects on the postsynaptic membrane potential [19, 178]. Yet, whereas the ISI distribution only accounts for second order correlations in spike trains, it is not *a priori* clear that higher order correlations are irrelevant for neural coding (for instance, pattern of precisely timed groups of spikes were proposed to exist as temporal codes [179]).

In this work, neural coding and information transmission are generally understood in terms of temporal codes of which rate codes can be considered a special case. The methods that I later apply to quantify the information contained in spikes about a stimulus account for precise spike timing and all temporal spike correlations (see section 1.3.2.2). Accordingly, here I use spiking neuron models rather than rate models (cf. section 1.1.4).

Another class of temporal codes that I do not considered in this work are *phase codes* in which the spike timing relative to the phase of large-scale brain oscillations (rhythms) encodes information. Phase codes have been shown to add to stimulus encoding through rates in cortex of different animals [158, 160, 161, 180].

The role of spike correlations. Temporal coding generally only implies that the precise timing of single spikes is coding-relevant. Whether the exact temporal coordination *between* two or more spikes, i.e. temporal spike correlations, is crucial for the neural code is a complementary question[16]. For example, the encoding in ISIs is a temporal code that fundamentally builds upon the correlation of consecutive spikes [176]. Generally, it has been found that spike correlations can both enhance and impair the neural code [181, 182]. Often, if the relevance of correlations is not explicitly known spikes are assumed to be independent encoding entities [29]. This enables a simpler analysis of the code, and some experimental studies have supported that the assumption of independence is sufficient [42, 183].

In theoretical neuroscience, the role of (temporal) spike correlations is often evaluated by comparing the stimulus decodability from original spike trains and that from the same spike trains that are shuffled such that spike correlations are eliminated (see [184, 185] for reviews). However, computing the respective quantities requires full knowledge of the neural code in place, which is normally lacking. Therefore, many studies addressing the role of spike correlations for neural coding have relied on methods that include correlations only approximately [25, 184, 186]. Developing an exact method to assess the role of spike correlations is part of this dissertation, and I introduce the respective information-theoretic foundations in section 1.3.2.2.

1.3.1.6 Population codes

I hitherto considered neural coding in single neurons. It is reasonable, though, to assume that neural coding in the brain occurs at the level of the simultaneous activity of populations of neurons of variable size [187, 188]. Here, I give a summary of the principal notions of neural coding on the population level.

Generally, it seems obvious that neuron populations can provide more stimulus information by distributing the stimulus space among them such that single cells only represent a small fraction of that space and can hence represent it more reliably within their dynamic range. Beyond that more complex population codes, including combinatorial codes, have been suggested (see e.g. [69, 154, 189–193]). A fundamental idea of population coding

[16] Here, I focus on the role of correlations between spikes in one neuron. Spike correlations across cells are an additional issue in population codes; see section 1.3.1.6.

is that the imprecision or trial-to-trial variability (noise) in the spikes of single neurons cancels out on the population level [25, 29]. However, this "averaging out" of noise in the population response relies on the assumption that the cross-correlations of the variability (*noise correlations*) between neurons with similar stimulus tuning are not too strong (strong noise correlations would make the cells essentially identical). Moreover, other observed types of cross-neural correlations, such as signal and pattern correlations, also influence the encoding capabilities in populations. Overall, spiking correlations in a population can both increase or decrease the quality of the neural code (in terms of decoding performance and information transmission), depending on the structure of noise correlations as well as signal-induced correlations [130, 181, 184, 188, 194–197].

In this work, populations are assumed to consist of neurons that receive identical signal-related inputs, i.e. neurons that are equally stimulus-tuned, and receive independent, uncorrelated noise[17]. The latter assumption is based on *in vivo* findings according to which noise correlations are often very weak with spike count coefficients of correlation between 0.01 and 0.2 [188, 198] and only marginally relevant for stimulus encoding [197], e.g. in different cortices [199, 200] and the retina [201] (exceptions exist [154, 202]).

1.3.2 Quantifying information in spike trains

The neural code refers to the question, "How are stimuli encoded in the neural activity?". In practice, this question can be answered conclusively only in rare cases and is especially elusive in cortical neurons. Moreover, it is not principally possible to assess whether a suggested neural code encompasses all the relevant features in the spiking activity. Likewise, comparing potential neural codes by comparing the performance of respectively designed decoders only works if the decoding is known to be optimal, which is not normally the case [197].

An alternative approach for analyzing the stimulus-response relation is guided by the question, "*How much* does the neural activity tell us about a stimulus?", i.e., by asking for the amount of information about a stimulus one can gain from observing neural activity. The mathematical framework for this approach is provided by *information theory*. Advantageously, information theory can be applied without having explicit knowledge of the neural code that is used to transmit stimulus information. As a consequence, it

[17] Notably, within this network architecture neuron and trial averages are equivalent.

allows comparing the performance of different neural codes by considering the input-output statistics in each case. Information theory is hence suitable to evaluate the relative performance of neural coding schemes and to investigate how different parameters in a model of information transmission influence its optimality [25, 29]. Information theory, therefore, provides the methodical basis of this work.

I introduce the information-theoretic concepts of *entropy* and *mutual information* in general terms in section 1.3.2.1. In section 1.3.2.2, I present different methods of applying information theory to spike trains that are relevant in this work.

1.3.2.1 Basics of information theory

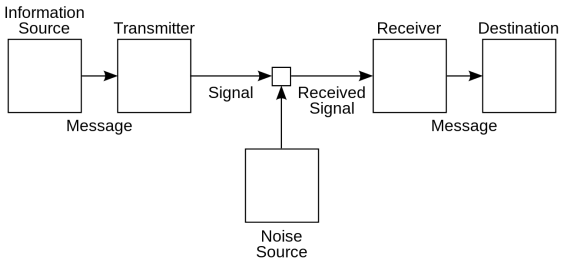


FIGURE 1.14: **Shannon's communication channel.** A message is chosen stochastically by the information source and transformed into a signal. This signal is distorted by a noise source and sent to the receiver. The receiver aims to transform the signal back into the original message. Image source: work by Wanderingstan, via Wikimedia Commons (Creative Commons); based on [203].

C E Shannon introduced information theory as a general framework to quantify the information that can be transmitted through a communication channel (illustrated in figure 1.14) [203]. The idea of this channel is tantamount to that of the noisy rate encoding described in section 1.3.1.1: a signal/stimulus x is chosen from a given distribution $p(X)$ and the communication channel (e.g. a neural circuit) produces a noise-corrupted, encoding output y (e.g. firing rate). Information theory now aims to quantify the information about X contained in Y. The respective quantity is called *mutual information*[18].

This mutual information consists of two parts. The first part describes the amount of information made available by the source via $p(X)$ and is called *entropy*. The entropy of a continuous distribution $p(X)$ is given by

$$H[X] = -\int \mathrm{d}x \; p(x) \log_2 p(x) = \langle -\log_2 p(x) \rangle_{p(X)} \,, \tag{1.13}$$

[18] Here, lower case x, y refers to single instances from ensembles (upper case) X,Y.

and has units bits; the integral is over the domain of $p(x)$. In the context of a communication channel the entropy of the source X is also called *signal entropy*.

The entropy of a distribution $p(X)$ measures the negative expected value of the *Shannon information* $\log_2 p(x)$ under this distribution. Moreover, the entropy fulfills a number of reasonable properties, such as being additive for independent signals. Loosely speaking, entropy measures information in terms of "surprise or uncertainty in the occurrence of a message x" and, therefore, is related to the spread of the distribution $p(X)$. For example, $H[X]$ is zero when $p(x_0) = \delta(x_0)$, i.e. when there is no uncertainty that x_0 will occur. On the other hand, $H[X]$ is maximized when $p(X)$ is a uniform distribution (if no further constraints are applied). Another illustration of the notion of entropy is shown

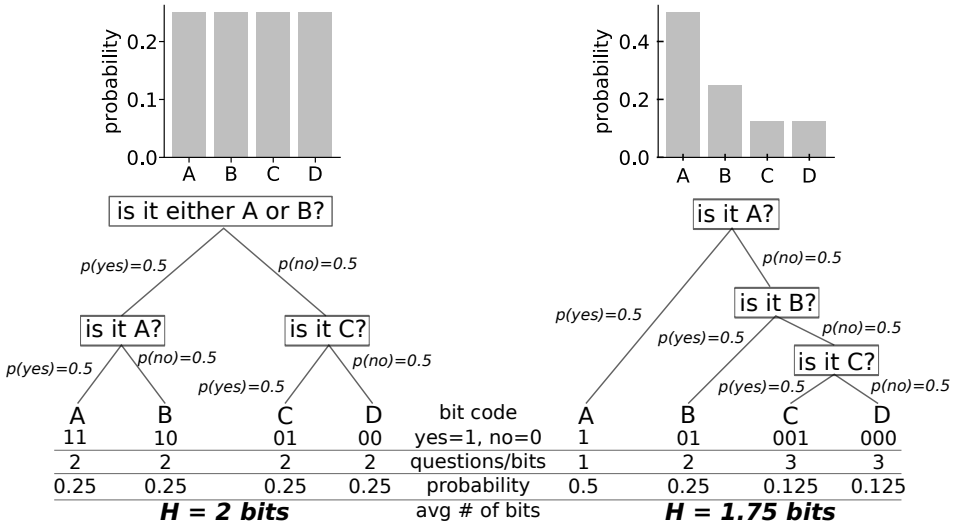


FIGURE 1.15: **Shannon entropy is equal to the average number of optimally placed yes/no-questions required to retrieve the message.** Given are a uniform (left) and a decaying (right) probability distribution for the letters A to D. The decision trees show optimal ways to retrieve an unknown letter that was drawn from the distribution above by asking binary yes/no-questions. The entropy of each distribution is then equal to the average number of required questions in bits. Conversely, the optimal retrieval strategy needs in average a number of questions equal to the entropy of the underlying distribution. An optimal binary code represents the sequence of yes/no answers. The uniform distribution possesses higher "uncertainty" and therefore has larger entropy. Here the discrete version of the entropy is used, $H = -\sum_i p(x_i) \log_2 p(x_i)$.

in figure 1.15 for a discrete distribution. It reflects Shannon's source coding theorem, according to which the average number of bits in an optimal encoding of X is equal to the entropy of $p(X)$ [203, 204].

The second part of the mutual information reflects how X and Y relate to each other. For example, Y could be purely stochastic and independent of X, or Y may be completely determined by X. The (statistical) uncertainty about x after having observed y, i.e. the

possibility of decoding x from y, is expressed in the conditional entropy $H[X|Y]$,

$$H[X|Y] = -\int \mathrm{d}y\ p(y) \int \mathrm{d}x\ p(x|y) \log_2 p(x|y). \tag{1.14}$$

Here, $p(X|Y)$ denotes the conditional probability of x after observing y. Since $H[X|Y]$ represents the variability in the mapping from a given input to the observed output it is often called *noise entropy* [29][19].

Taken together, the mutual information[20] of X and Y – the average information contained in Y about X (and vice versa) – is given by the information in X minus the uncertainty in X after observing Y. It has units of bits and can be written as

$$\mathcal{I}[X,Y] = H[X] - H[X|Y] = H[Y] - H[Y|X]. \tag{1.15}$$

The mutual information is always non-negative and invariant under homeomorphic transformations of X and Y [204]. If X and Y are independent it is $\mathcal{I}[X,Y] = 0$, and if X and Y are deterministically related the mutual information is maximized at $\mathcal{I}[X,Y] = H[X]$.

1.3.2.2 Information theory for spike trains

In the neuroscience context, the source X in Shannon's communication channel (figure 1.14) corresponds to relevant stimuli and signals, such as sensory inputs, behaviorally relevant cues, or direct stimulation of neurons. The outputs Y can generally be neural responses at various time and length-scales as, for example, local field potentials, electroencephalographic signals, or functional magnetic resonance imaging signals. As explained earlier, in this work spike trains are the information carrying units [25, 29, 197]. The central question, "*How much* does the neural activity tell us about a stimulus?", is quantitatively answered by the mutual information of the signals and the respective spike trains. Formally, following Eq. (1.15), the mutual information $\mathcal{I}$ about the stimuli $\{s\}$ conveyed by the resulting neural spikes $\{\rho\}$ is given by the difference of signal and

[19] In other words, $H[X|Y]$ represents the uncertainty about x after observing y.
[20] In this work, the shortened version *information* is often used interchangeably.

noise entropy,

$$\mathcal{I} = H_{\text{signal}} - H_{\text{noise}} = -\int \mathrm{d}\rho\; p(\rho)\log_2 p(\rho) + \left\langle \int \mathrm{d}\rho\; p(\rho|s)\log_2 p(\rho|s) \right\rangle_{\{s\}}, \quad (1.16)$$

where $p(\rho)$ is the probability of encountering a spike train ρ, and $p(\rho|s)$ is the conditional probability of encountering a spike train ρ given that the signal s is presented; $\langle\rangle_{\{s\}}$ denotes the average over different signals from the relevant set.

Here, I mostly consider the mutual information *per spike*, in contrast to the commonly used mutual information rate. This choice aligns with the *efficient coding hypotheses* by H Barlow [205]. According to this hypotheses, neural coding is optimized for maximum information transmission while simultaneously minimizing the number of required spikes because spikes are accompanied by a metabolic cost. Thus, neural systems are thought to maximize information transmission per spike. The efficient coding hypotheses has been supported repeatedly [25, 206–209].

A practical challenge in applying information theory to spike trains lies in defining the probability distributions through which H_{signal} and H_{noise} are defined. This is a corollary of the problem of finding the spike time generating probabilities $p(\rho) = p(\{t_i\}|s(t))$, as discussed in section 1.3.1.1. In particular, incorporating temporal correlations in the spike trains poses a challenge (see below). Therefore, many studies on mutual information in spike trains have built upon reduced, low-dimensional spike train features, such as spike counts, firing rates or ISI distributions [25, 104, 159, 165, 210–214]. However, according to the data-processing inequality, these reduced representations of the spike trains merely yield lower bounds for the full mutual information [197]. From the discussion on neural coding, it is obvious that high temporal precision and correlations, as well as higher order moments of the distribution, may be relevant to capture the information contained in spikes fully. In these cases, the obtained lower bounds cannot be expected to be good approximations to the full mutual information.

A method to compute the mutual information via the complete probability space of all possible discretized sequences of "spike" and "no spike" was introduced by S Strong *et al.* [215]. This method, though, depends on the chosen temporal discretization, is numerically sensitive, and requires enormous amounts of spiking data because the probability space grows exponentially with the number of bins [98, 215–217]. Another method provides an exact decomposition of the mutual information in terms of different probability

distributions over the space of possible spike times [181]. This method, too, suffers from a probability space that grows exponentially with the number of spikes and analytic expressions for these distributions are available in limited cases only (e.g. if a spike count code is assumed) [184]. I provide a more detailed discussion of these and related methods in section 2.4.3.

Under certain stationarity assumptions, the recently introduced *correlation method* overcomes many of the described issues and allows to compute the complete mutual information using spike times. In the following, I introduce this and other information-theoretic methods and related functions that are important in the remainder of this thesis.

Stationarity assumptions. Here, I assume stationarity and ergodicity in the inputs and spiking activity. Furthermore, I consider very long trial lengths T. Each trial corresponds to the presentation of a dynamic signal, which is drawn from a given ensemble, and from which the neuron generates a spike train. Stationarity in the signal implies that the trial-averaged statistics of the signal do not change over time and are equivalent to the time-averaged statistics (this generalizes to the input current if applicable). This implies signal and noise processes with finite correlation times. Static neuron models with finite memory then lead to stationary spiking statistics. This is naturally fulfilled for the Poisson and integrate-and-fire models here. These stationarity assumptions are common in the literature and supported by experimental findings [25, 213, 218].

Spike train correlation functions. Computing the mutual information largely relies on calculating different signal and spike correlation functions. The (cross-)correlation function between two time series $x(t)$ and $y(t)$ with length T is defined as

$$R_{xy}(h) = \frac{1}{T}\int_0^T \mathrm{d}t\ \langle x(t)y(t-h)\rangle, \tag{1.17}$$

where $\langle\rangle$ represents averaging over the statistical ensemble under consideration. The correlation function, loosely speaking, measures the linear similarity of the two time series if they are shifted against each other by a time h. In this work, correlation functions are normally represented in the Fourier domain, i.e. as spectral functions of

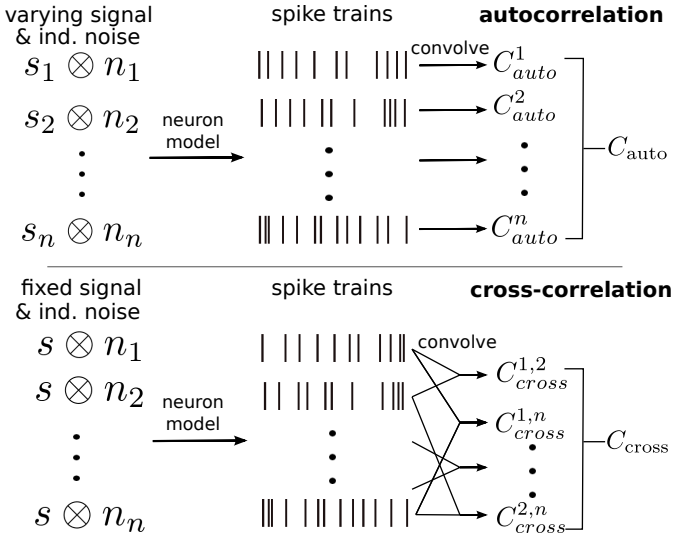


FIGURE 1.16: **Auto- and cross-correlation functions are obtained from trials with varying or fixed signals, respectively.** Top: The autocorrelation C_{auto} is computed from independent trials with varying presented signals and noise. It is obtained by averaging the single autocorrelation functions across trials (technically, one very long trial suffices). Bottom: To compute the cross-correlation C_{cross}, first, a signal is drawn and presented in each trial with independent noise. Then, the cross-correlation functions of all pairs of trials are evaluated and averaged. Another averaging over the cross-correlation functions obtained from different signals may occur (not shown, see text). The notation $s \otimes n$ denotes a general operation of signal s and noise n that is fed into the neuron as input current.

frequency. The Fourier transform $\mathcal{F}$ of a time series $x(t)$ is defined as

$$\mathcal{F}(x(t)) = \tilde{x}(f) = \int_0^T x(t) e^{-i2\pi f t} \mathrm{d}t \ . \tag{1.18}$$

Using this definition and the *convolution theorem* [30], the Fourier transformed correlation function $R_{xy}(h)$ as a function of frequency f is given by

$$C_{xy}(f) = \lim_{T\to\infty} \langle \tilde{x}(f) \tilde{y}^*(f) \rangle \,, \tag{1.19}$$

with * denoting the complex conjugate. Sometimes these functions are expressed in terms of angular frequency $\omega = 2\pi f$. Also note that $C_{xy}(f)$ is equivalent to the (cross-) power spectrum $S_{xy}(f)$ of x and y under a common convention [29]. In particular, the power spectrum of a signal $s(t)$ is denoted by $S_{ss}(f)$, and the rate power spectrum is $S_{rr}(f)$.

As before, spike trains $\rho(t)$ are represented as sequences of δ-peaks at spike times t_f, such that $\rho(t) = \sum_{t_f} \delta(t - t_f)$. Then, the spike autocorrelation function in frequency

space is given by

$$C_{\text{auto}}(f) = \lim_{T\to\infty} \langle \tilde{\rho}_n(f)\tilde{\rho}_n^*(f)\rangle_{\text{tr}_n} \ , \tag{1.20}$$

where $\tilde{\rho}_n(f)$ is the Fourier transform of the spike train in trial n. For each trial $1, \ldots, N$ a new signal and noise from their respective stochastic ensemble are presented and averaging $\langle\rangle_{\text{tr}_n}$ occurs over all trials[21] (see figure 1.16). One noteworthy property of C_{auto} is that its high-frequency limit is equal to the average firing rate,

$$\nu = \lim_{f\to\infty} C_{\text{auto}}(f) = \frac{1}{T}\int_0^T \mathrm{d}t \ \langle \rho(t)\rangle_{\text{tr}}. \tag{1.21}$$

The spike cross-correlation function is defined analogously as [98]

$$C_{\text{cross}}(f) = \lim_{T\to\infty} \langle \tilde{\rho}_m(f)\tilde{\rho}_k^*(f)\rangle_{\text{tr}_{m\neq k}}. \tag{1.22}$$

Here, the averaging occurs over pairs of trials, or neurons, m and k whose spike trains originate from presentation of the same signal but have different, independent noise traces (see figure 1.16). Notably, $C_{\text{cross}}(f)$ is equivalent to the Fourier transformed autocorrelation function of the PSTH in the limit of infinitesimally small binning [98, 196] (cf. section 2.1.2).

Correlation method. The correlation method is at the core of this work and was published by A Dettner *et al.* [98] in 2016. It makes use of the fact that the Fourier coefficients of spike trains follow Gaussian distributions under the stationarity and finite memory assumptions (see also section 3.4). Using the correlation method, the mutual information of signals and spike trains carried by each frequency mode is [98]

$$\mathcal{I}_{\text{tot}}(f) = -\frac{1}{2}\log_2\left(1 - \frac{C_{\text{cross}}(f)}{C_{\text{auto}}(f)}\right). \tag{1.23}$$

The label "tot" indicates that this expression determines the total, i.e. complete/full, information – as opposed to the approximations presented below. Importantly, even though Eq. (1.23) only depends on two-point correlation functions, it captures the full information including that of higher order correlations (due to Gaussianity). Moreover,

[21] It is noteworthy that a trial averaging is not strictly necessary to compute the autocorrelation function due to the assumed stationarity and $T \to \infty$.

Eq. (1.23) is exact under the stationarity assumption but does not require Gaussianity in the signal or inputs (see section 3.4.2).

Lower bound: linearly decodable information. An approximation of the mutual information is given by the information content that *can be decoded linearly* and reads [128, 218],

$$\mathcal{I}_{\mathrm{ld}}(f) = -\frac{1}{2}\log_2\left(1 - \frac{|C_{s\rho}(f)|^2}{S_{ss}(f)C_{\mathrm{auto}}(f)}\right) \leq \mathcal{I}_{\mathrm{tot}}(f), \tag{1.24}$$

with the signal power spectrum $S_{ss}(f)$ and the signal-response cross-correlation $S_{s\rho}(f) = C_{s\rho}(f)$.

$\mathcal{I}_{\mathrm{ld}}$ represents a *lower bound* for the total information content [218]. This lower bound results from the mutual information of the signal and a response-based linear signal reconstruction, i.e. it represents how much information a purely linear decoder could retrieve from the spike trains. Nonlinear signal-response relations cannot be captured by Eq. (1.24) and are equivalent to noise from the information transmission perspective. Nonetheless, this lower bound has often been used for information estimates [129, 206, 219–223]. Even though $\mathcal{I}_{\mathrm{ld}}$ is an approximation, it is suitable as a reference to determine how much information is encoded nonlinearly in a given system (see chapter 3).

Upper bound. An *upper bound* for the mutual information has been introduced and used in previous studies [218, 224–226]. It is given by

$$\mathcal{I}_{\mathrm{ub}}(f) = -\frac{1}{2}\log_2\left(1 - \gamma_{\rho_1\rho_2}(f)\right), \tag{1.25}$$

whereby

$$\gamma^2_{\rho_1\rho_2}(f) = \lim_{T\to\infty}\frac{|\langle\tilde{\rho}_1(f)\tilde{\rho}_2^*(f)\rangle|^2}{|\langle\tilde{\rho}_1(f)\rangle|^2|\langle\tilde{\rho}_2(f)\rangle|^2}, \tag{1.26}$$

is the response-response coherence with spiking responses ρ_1, ρ_2 to the same signal in two different trials. The validity of Eq. (1.25) as a universal upper bound has been questioned due to inconsistent results and the implied assumption that neurons encode information in their mean conditional response [218, 225, 226]. Addressing this, I discuss the exact relation between the correlation method Eq. (1.23) and the upper bound in

section 3.5 (in fact, they are equivalent). Thereby, I investigate the assumptions made for deriving the upper bound and validate them.

Information per spike. Here, I am mostly interested in the mutual information per spike (see above). The average information per spike in either case (labeled by x) is obtained by integrating Eqs. (1.23), (1.24), and (1.25) over all frequencies, and by normalizing with the average firing rate ν,

$$\mathcal{I}_x = \frac{1}{\nu} \int_{-\infty}^{\infty} \mathrm{d}f \; \mathcal{I}_x(f). \tag{1.27}$$

[$\mathcal{I}_x$ has units bits/spike.] Including negative frequencies in the integration is in line with many other studies in which "backward" frequencies are considered to contribute independent information (see [25, 128, 218]). Arguably, the negative frequencies carry information about the temporal directionality of the spike trains' modes. The results of section 2.3 support the inclusion of negative frequencies.

The inclusion of negative frequencies, however, remains equivocal. In chapter 3, the integration in Eq. (1.27) occurs only over positive frequencies, following [98] and accounting for the fact that, for any real function, the Fourier coefficients at positive frequencies determine the Fourier coefficients at negative frequencies [30], which should hence not provide additional information[22]. This choice does not influence the findings in chapter 3 since the relations of different mutual information values are unaffected, whereas absolute values are subordinate.

Information in independent spikes. The correlation method, expressed through Eq. (1.23), captures the full mutual information and fundamentally depends on spike correlation functions. As discussed earlier, often spikes are considered to transmit signal information independently [29]. Under this assumption, the information about a stimulus contained in a spike train is simply the sum of the information provided by each spike in isolation. The information carried by (assumedly) *independent* spikes that are

[22] Also note that in chapter 3 values of the mutual information per spike are indicated by superscript labels for better discriminability, whereas frequency-resolved information values have subscripts.

generated via point processes with time-dependent firing rate $r(t)$ reads [42, 227]

$$\mathcal{I}_{\text{ind}} = \frac{1}{T}\int_0^T \mathrm{d}t\ \frac{r(t)}{\nu}\log_2\left(\frac{r(t)}{\nu}\right), \tag{1.28}$$

where T denotes the length of the sequence $r(t)$ and ν is the average firing rate. Here, a rate encoding with a deterministic transfer function $r(s(t))$ is implied. Therefore, in stationary conditions and for sufficiently long T, Eq. (1.28) is equivalent to an averaging over the rate-determining signal ensemble. Hence, one can rewrite

$$\mathcal{I}_{\text{ind}} = \left\langle \frac{r(t)}{\nu}\log_2\left(\frac{r(t)}{\nu}\right)\right\rangle_s, \tag{1.29}$$

where $\langle\rangle_s$ denotes the signal averaging which, due to stationarity, effectively removes the time-dependence of the rate [42, 227]. This expression has been known for a while and provides a mathematically tractable information measurement that only requires knowledge of the signal-induced rate variations. It has been used to analyze information transmission in various brain regions such as hippocampal cells [228–232], visual cortex [233], and the fly visual system [227].

Coding synergy/redundancy. As indicated above, correlations between spikes (within a neuron or pairs of neurons) can either increase or decrease the mutual information. These effects are known as coding *synergy* and *redundancy*, respectively. Here, I quantify coding redundancy/synergy among spikes as the difference between the mutual information of spikes and signals that is carried by the spike trains as a whole ($\mathcal{I}_{\text{tot}}$) and the one carried by their individual spikes if they are treated as being independent ($\mathcal{I}_{\text{ind}}$) [25, 184, 227]:

$$Syn = \mathcal{I}_{\text{tot}} - \mathcal{I}_{\text{ind}}. \tag{1.30}$$

Positive values of Syn indicate synergy, negative values redundancy. More generally, the concept applies to the comparison of the information in compound events of spikes (e.g. spike triplets or certain ISIs) with that carried by single units of these events (e.g. spikes) [227]; cf. the definition of synergy in chapter 3.

Open questions. A long-standing challenge in neuroscience has been to understand the code that is used in sequences of stereotyped action potentials (spike trains) to transmit information about relevant signals. Earlier studies have suggested that the efficiency

of the neural code can be increased when temporal correlations in the spike trains are minimized, or equivalently, that temporal correlations in spike trains introduce redundancy to signal encoding [205, 227, 234, 235]. In particular, in rate codes correlations in the firing rate across time introduce redundancy to the information carried by consecutive spikes within intervals that are not much longer than the rate correlation length which renders $\mathcal{I}_{\text{ind}}$ an upper bound for the exact mutual information [227, 236].

Exact methods to determine $\mathcal{I}_{\text{tot}}$ were lacking until recently and therefore, quantifying the effect of temporal spike correlations on coding redundancy has remained an open challenge. In this work, I address the following questions:
How can the impact of different temporal spike correlation structures on coding redundancy be quantified? What are the properties of a rate code that determine these effects and how can such a code be optimized?

In chapter 2, I analyze how different signal/stimulus properties and temporal spike correlations shape information transmission.

1.4 Input mean and variance modulation

As described in section 1.1.4, neurons emit spikes as a result of integrating synaptic inputs. In this section, I explain how cortical structure (cf. section 1.2) can give rise to specific properties of spontaneous and stimulus-dependent input currents to cortical neurons. In section 1.4.1, I introduce the characteristics of input currents in the fluctuation driven regime during spontaneous cortical activity. In section 1.4.2, I present the notion of mean and variance modulated input currents that both produce spike trains that carry information about the modulating stimulus. These two encoding schemes are central to the present work, and I discuss the information-theoretic questions concerning mean and variance coding at the end of section 1.4.2.

1.4.1 Spontaneous activity

The characteristics of spontaneous activity in cortical circuits, i.e. neural activity which is not directly linked to sensory/external stimuli, differ for different brain states [83].

During slow-wave sleep and in some states of anesthesia, the membrane potential of cortical neurons switches between depolarized *up-states* and hyperpolarized *down-states* [237]. Experimental studies have shown that during the up-states neurons receive balanced excitatory and inhibitory synaptic inputs [71, 72, 81]. In awake animals, a well-documented brain state is the *high-conductance state*. Similar to the up-state in sleep, cortical neurons in the high-conductance state receive a bombardment of inputs from approximately 5,000 to 60,000 presynaptic neurons that fire at high rates. In *in vivo* conditions this leads to a persistent depolarization, highly fluctuating membrane potentials and tonic, irregular firing in the postsynaptic neurons as shown in different cortical regions [83, 238–241]. The high-conductance state during spontaneous activity is consistent with loosely balanced networks, as shown in the association cortex of awake cats [242]. In other brain states of awake animals, such as the quiet wakefulness condition, and during transitions between up- and down-states signatures of a tighter E/I balance in spontaneous activity have been found [69, 74].

Here, spontaneous activity is modeled in accordance with the balanced, high conductance state and the fluctuation-driven, irregular firing observed in the cortex of awake animals [83, 243, 244]. The dense presynaptic bombardment results in approximately continuous and smooth postsynaptic currents. The underlying *diffusion approximation* assumes a high number of independent presynaptic spikes that each only cause a very small postsynaptic potential change [56, 245, 246]. Formally, the input current during spontaneous activity can then be written as

$$I_0(t) = \mu + \xi_{\sigma_n,\tau_n}(t). \tag{1.31}$$

Here, μ denotes a constant input current and $\xi_{\sigma_n,\tau_n}(t) = \xi(t)$ is a fluctuating noise current. The noise current is a Gaussian process with zero mean, hence $\langle I_0(t)\rangle = \mu$. The temporal autocorrelation of the noise is $\langle \xi(t)\xi(t+h)\rangle = \frac{\sigma_n^2}{2\tau_n} e^{-|h|/\tau_n}$ with correlation time τ_n and noise variance σ_n^2. For $\tau_n > 0$ this describes an Ornstein-Uhlenbeck process [32, 83, 246] and corresponds to the assumption of exponential synapses [56]. The limit $\tau_n \to 0$ represents instantaneous synapses and in this limit the noise current becomes Gaussian white noise with autocorrelation $\langle \xi(t)\xi(t+h)\rangle = \sigma_n^2 \delta(h)$ [56]. Whenever $\tau_n = 0$, it is referred to the white noise case.

1.4.2 Signal-dependent current mean and variance modulation

In order to transmit information about a signal, the input to and activity of the encoding neurons must be altered by that signal. In the following, I introduce two important ways how an external signal can modulate E/I balanced inputs in cortical neurons – called *mean modulation* and *variance modulation*. I present a mathematical formulation of these encoding schemes in chapter 3 as a result of this work.

Mean modulation (MM) denotes signal-induced changes of the mean input current and can arise from a transient imbalance between excitation and inhibition. This may be the case if excitation and inhibition are not equally tuned to a sensory stimulus as for example observed in orientation tuning visual cortex of mice [247] or odor coding in the olfactory cortex [76, 248] (see also [68, 69]), or if they are balanced but inhibition is delayed [70, 74, 249]. In these cases, the mean somatic current in the respective neurons is modulated by the stimulus, and the net current change averaged across neurons (or trials, respectively) is representative of the stimulus. For example, it has been shown for orientation selective neurons in cat visual cortex that different orientations correspond to different levels of the average subthreshold voltages, and this orientation tuning is generated by an imbalance between excitatory and inhibitory conductances [250].

Variance modulation (VM) arises when excitation and inhibition are equally stimulus-tuned [73] and remain balanced under stimulation [71, 72, 75]. Then, in a loose balance setting, excitatory and inhibitory inputs both follow the signal but additionally possess independent fluctuations at faster time scales. As a result, external signals modulate and are encoded in the current and voltage fluctuations, whereas their mean remains unchanged. This has been observed in cat V1 under stimulation *in vivo* with broadband stimuli where the standard deviation of the membrane voltage is an increasing function of the stimulus strength [251]. Other studies on cat and primate V1 obtained similar results [252, 253]. Also, theoretical work on loosely balanced networks has demonstrated that membrane voltage fluctuations are modulated by the strength of external signals [254]. Similarly, other studies on cat visual cortex have measured the power of γ-range (25–70 Hz) membrane fluctuations in response to the presentation of moving gratings in the receptive field of the considered neurons [255, 256]. In this case, sensory information is encoded in the envelope of γ-band voltage fluctuations, suggesting that

the stimulus modifies the variability of the somatic input. A similar stimulus-tuned fluctuation strength in the θ-band (6–10 Hz) has been found in hippocampal place cells [257]. Moreover, remarkable evidence for coding through variance modulations is given by experimental studies on cortical pyramidal neurons that found a higher response sensitivity to changes in input fluctuation amplitudes than to mean changes [258, 259].

1.4.2.1 Mean and variance modulation in signal transmission.

Changes in both the mean and the variance (fluctuation) of input currents modify the postsynaptic activity. Both modulation schemes are thus capable of encoding and transmitting signals that induce the respective input changes.

The computational capabilities of mean and variance modulations have commonly been assessed in terms of linear response functions, which approximately describe the frequency-dependent response to changes in either modulation channel (section 1.3.1.3). Theoretical studies examining the leaky integrate-and-fire (LIF) and exponential integrate-and-fire (EIF) neuron models could obtain the linear response functions for mean and variance modulations analytically either over the full frequency range [107, 260] or for the limit of infinitely high frequencies [36, 56]. Furthermore, more general semi-analytic algorithms to obtain response functions were introduced [108]. The response functions generally have different functional shapes for variance and mean modulations and further depend on the spike initiation time (zero for LIF neurons and finite for EIF neurons, as explained in section 1.1.4) and steady-state characteristics [40]. In particular, for variance modulations, linear response functions in the LIF model can have finite response amplitudes even in the limit of infinitely high frequencies when white background noise is assumed [107]. This finding is supported by a whole-cell *in vitro* study on cortical neurons showing that encoding signals in the variance of uncorrelated synaptic inputs to a cortical ensemble enables reliable transmission of rapidly changing signals [85]. This can be explained analytically for LIF neurons where the instantaneous firing rate is directly proportional to the input variance, and consequently the firing rate response to changes in variance has an instantaneous component [36, 85, 107, 261]. In contrast, the encoding of signals in the mean current is subject to low-pass filtering as shown in experiments as well as analytically for LIF and EIF neurons [85, 107]. Table 1.2 lists the high-frequency limits of the linear response functions for LIF and EIF models.

Together, these findings suggest that the mean and variance coding strategies may use different mechanisms to convey information with mean modulations generally evoking larger response amplitudes and variance coding facilitating high frequency encoding [3].

On the other hand, there are some similarities between both coding channels. For example, correlated noise background results in a finite amplitude of the linear response functions for both mean and variance modulations at high frequencies in the LIF model. In EIF neurons, the responses decay to zero in all cases [36]. Overall, theoretical considerations show that neural responses can have small or high-frequency cut-offs, depending on electrophysiological parameters, modulation type, and noise parameters. Experimental studies of neocortical neurons have shown that the linear response functions for both MM and VM generally have a low-pass form, but their frequency cut-offs are located at high frequencies of a few hundred hertz [17, 112–115, 262, 263].

	$\tau_n = 0$		$\tau_n > 0$	
	MM	VM	MM	VM
LIF	$f^{-1/2}$	f^0	f^0	f^0
EIF	f^{-1}	f^{-1}	f^{-1}	f^{-1}

TABLE 1.2: Scaling of the high-frequency limits of the linear response functions with respect to stimulation frequency f for mean and variance modulation, LIF and EIF models, and for white ($\tau_n = 0$) and colored noise ($\tau_n > 0$), based on [36].

1.4.2.2 Combined mean and variance encoding.

Some of the experimental evidence for signal encoding via mean and variance modulations has remained inconclusive or ambiguous [68, 69]. Generally, it seems plausible that in *in vivo* situations, both the input mean and fluctuation power (variance) of input currents are modulated by relevant signals [251]. An earlier mentioned study on cat V1 (simple and complex orientation selective cells) shows that a slow changing visual stimulus modulates not only the mean voltage but at the same time serves as an envelope for the power in γ-band voltage fluctuations [255]. A computational follow-up study suggests that this variance modulation crucially enhances the stimulus information transmission compared to encoding in slow mean and rate modulations only. This points towards a synergetic effect, or "information gain", through interactions of MM and VM. However, the mentioned study is based on lower bound approximations for the

information [256], and a thorough information-theoretic investigation of *combined mean and variance encoding* – going beyond γ-band fluctuations only – has been lacking.

Open questions. As explained, experimental evidence suggests that mean and variance modulations are both present in the cortex and capable of signal-encoding by shaping neural activity. To understand how the coding strategies of current mean modulation and variance modulation operate and to quantify their capabilities of encoding incoming signals in spikes, it is crucial to determine their information transmission efficiency. This is an important effort not only because it will help clarify intra- and cross-areal communication in the brain but because the answer to it will shed light on the basics of analog-to-binary information transmission. While linear response theories can anticipate the amplitude and phase of the rate response approximately, it is not *a priori* clear how these findings translate into specific predictions about coding efficiency [3]. In particular, it remains open whether the putative benefit of encoding signals in input fluctuations withstands an information-theoretic analysis. Also, it is unclear how much of the total information is captured by the linear estimates used in numerous previous studies [218]. Moreover, investigations of mean and variance coding often relied on the assumption of white background noise which can be considered non-biological because postsynaptic potentials always possess finite decay times, and thus noise currents cannot be perfectly white [260]. Following these gaps, I address these specific questions in my work:

What is the complete information content of spike trains about mean and variance modulating signals in LIF and EIF neurons in the fluctuation driven regime? To what extent can this information transmission be understood in terms of linear input-output relations and what influences the encoding linearity?

Regarding combined mean and variance modulations, it has been suggested that changes in the input variance can have beneficial effects on the overall encoding performance that may otherwise be driven by mean modulations only [251, 256]. However, this topic has been analyzed only in a narrow context and not by means of exact information measures. I hence investigate the following question:

Can signal encoding via a combination of mean and variance modulations give rise to synergetic effects that facilitate information transmission?

The according results are presented in chapter 3.

1.5 Outline of the thesis

In the preceding sections, I have introduced the biological and mathematical foundations underlying the research program of this dissertation. Along that way, I have derived and formulated the guiding scientific questions that I answer in the forthcoming chapters, and have introduced the relevant methods. Here, I give a general overview of the structure and outline of the remainder of this dissertation. A respective visualization of the different parts of the results chapters 2 and 3 and their interrelations is shown in the diagram of figure 1.17. More detailed outlines are given at each chapter's beginning. Generally, and following the work process behind it, in each chapter, I first summarize the considered research questions, I then recapitulate the applied methods and subsequently the present the results. Additional concise summaries of the results are given in the last sections of chapters 2 and 3 and at the beginning of chapter 4.

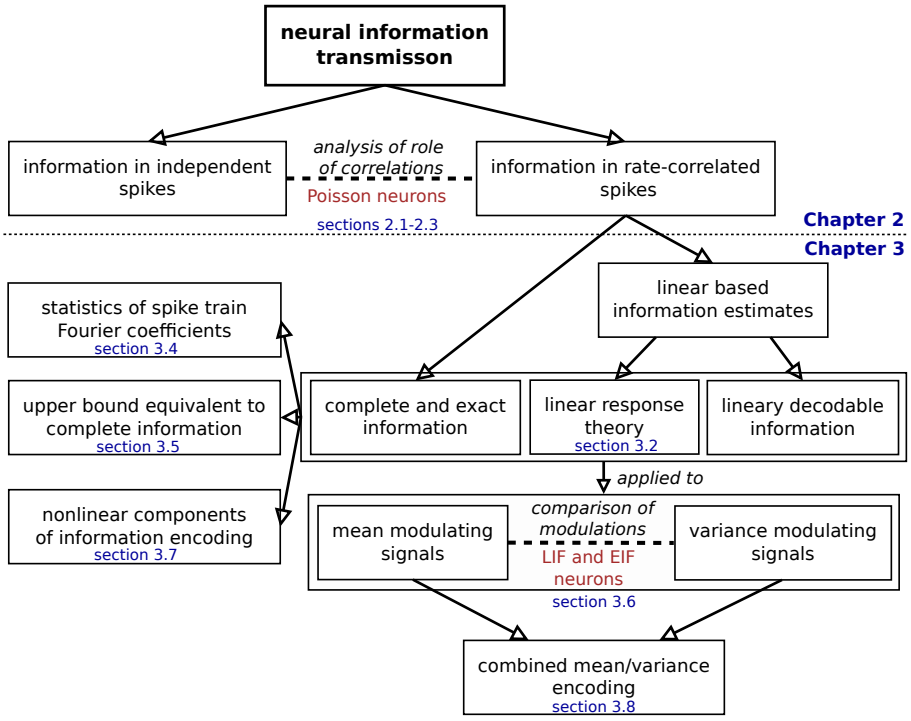


FIGURE 1.17: **Structure and outline of result chapters 2 and 3.**

Chapter 2 deals specifically with the information transmission about signals that are encoded in dynamic, correlated rate modulations. This information is contrasted with and systematically compared to the information transmission under neglect of rate correlations (sections 2.1–2.2). Thereby, I analyze the resulting redundancy, investigate the role of temporal spike correlations, and discuss different limiting cases (section 2.3). The

chapter ends with a summary and discussion, including a dedicated discussion of related work (section 2.4).

Chapter 3, on one hand, addresses a specific comparison of the information transmission capabilities of mean and variance modulation encoding. On the other hand, it contains more general analyses of the linear and nonlinear components of the mutual information and elaborates on the methodical foundations of the correlation method. The model framework and the MM and VM input currents are formally introduced in section 3.1. Subsequently, I present a linear response theory for the mutual information (section 3.2) and provide details on the simulation routines and parameter choices (section 3.3). In sections 3.4 and 3.5, I investigate the coding-relevant statistics of the spike train Fourier coefficients and based on that link different information measures. The explicit comparison of MM and VM is detailed in section 3.6, information encoding linearity is examined in section 3.7. Moreover, I consider the combined application of mean and variance modulations (section 3.8). The chapter is completed by a discussion of the results (section 3.9).

The final chapter 4 contains a general discussion of this dissertation in a broader context and addresses limitations and caveats of the work (sections 4.1–4.6). Moreover, I give an outlook on potential research projects extending the present work (section 4.7).

Chapter 2

The role of spike correlations in neural information transmission

Remark: *Some methods, results, figures and elements of the text in this chapter are also contained in this research article that Dr. Tchumatchenko and I published in 2019:*

- Herfurth, T. & Tchumatchenko, T. (2019). Quantifying encoding redundancy induced by rate correlations in Poisson neurons. Physical Review E, 99(4), 042402. [2]

As described in introduction section 1.16, a method to compute the information contained in spike trains about a stimulus capturing all types of contributing spike correlations is given by the correlation method – denoted by $\mathcal{I}_{\text{tot}}$. The information contained in spikes that are assumed independent, $\mathcal{I}_{\text{ind}}$, by design does not account for rate correlations or any other spike correlations. Now, systematically comparing $\mathcal{I}_{\text{ind}}$ and $\mathcal{I}_{\text{tot}}$ for a given scenario enables an analytical investigation of the role of spike correlations – both intrinsic and signal-induced – for encoding redundancy. This allows one to address open questions regarding the role of spike correlations for neural coding. I analyze how different signal properties and temporal spike correlations shape the information contained in correlated and independent spikes (the terms "independent spikes" and "correlated spikes" are used in this short form for brevity; they refer to cases where spikes are *treated as* being independent or correlated, respectively). I focus on the effects of rate correlations, in particular, temporal spike correlations that are signal-induced, e.g. by sensory

inputs, and do not stem from intrinsic dynamics or spontaneous activity. I hereby establish mathematical links between information-theoretic methods – including lower and upper bound Eq. (1.24) and (1.25) – with and without incorporation of temporal spike correlations [25, 218, 227]. Methodically, I consider Poisson neurons whose rate follows and represents a signal. Poisson neurons do not exhibit intrinsic spike correlations and thus are eligible to analyze signal-induced correlations (see figure 2.1). Nonetheless, the information measure derived for Poisson neurons is also expedient for non-Poisson spiking neurons where it can be used to dissect the effects of intrinsic and signal-induced rate correlations.

In section 2.1, I review the expressions for the mutual information of signals and spikes with and without consideration of temporal spike correlations. Based on this, I derive a method to analytically compute the information carried by spike trains of rate modulated Poisson neurons in terms of the firing rate autocorrelation function and the mean firing rate.

By means of three example rate processes that are introduced in section 2.2, I demonstrate how overall rate distribution and correlation structure of these processes influence the amount of information which is transmitted by independent and temporally correlated spikes, respectively. I investigate the role of different features of temporal correlations for coding redundancy in detail in section 2.3. I thereby link expressions for the information transmission through independent and correlated spikes and show that both information measures are equivalent in the regime of small rate modulations (section 2.3.3) where only the average modulation depth determines information content in both cases. I further illustrate that vanishing rate correlation times maximize the information in correlated spikes in section 2.3.4.

The results are summarized and discussed in section 2.4.

2.1 A method to measure the effect of rate correlations on information transmission

The model framework of this chapter is first introduced in section 2.1.1, and a method to analytically compute the information carried by spike trains of rate modulated Poisson neurons is subsequently derived in section 2.1.2.

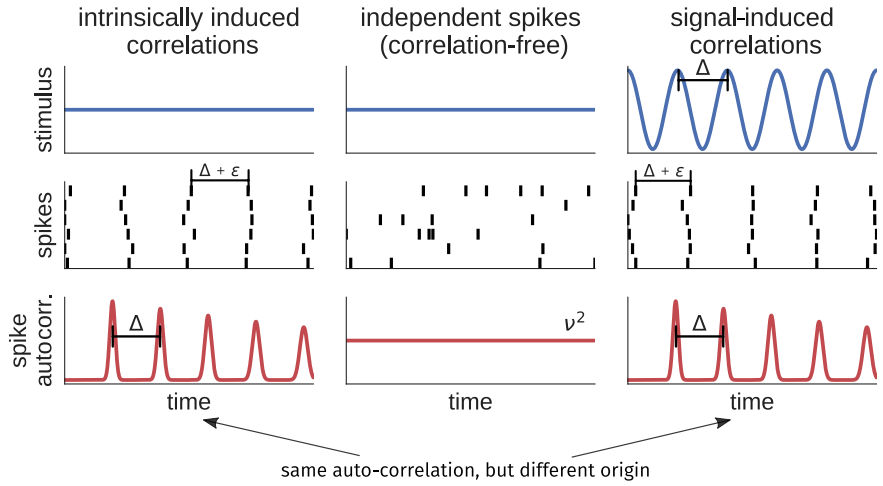


FIGURE 2.1: **Spike autocorrelations can emerge from intrinsic neural dynamics or can be signal-induced.** Here, I illustrate both cases. Intrinsic correlations (left column): a constant signal (top) causes regular spiking responses with only small variations ϵ in the interspike intervals (middle). This could represent an integrate-and-fire neuron that is stimulated by a constant suprathreshold current and weak noise. The resulting autocorrelation of the spike trains (see section 1.3.2.2) shows oscillations (bottom). Absence of temporal correlations (center column): neurons that respond to a constant signal in a Poisson manner have independent interspike intervals. Signal-induced correlations (right column): the signal is a sinusoidal function and neurons fire only around the local maxima of this signal. The resulting autocorrelation is shown on the bottom. Even though the spiking correlations look similar to those on the left, they are induced by signal correlations rather than internal dynamics. More generally, these correlations can be classified as rate correlations. In this chapter, I mostly examine the effect of rate correlations on information transmission.

2.1.1 Rate encoding Poisson neurons

To quantitatively analyze the impact of temporal spike correlations on information transmission I consider Poisson neurons[1]. As introduced in section 1.1.4.3, Poisson spiking is characterized by a spiking probability that is proportional to the instantaneous firing rate and independent of the occurrence of any other spikes. The firing rate $r(t)$ of these Poisson neurons follows a time-dependent signal $s(t)$ which implies a rate encoding of the signal. Poisson neurons do not exhibit intrinsic spike correlations, e.g. noise correlations, and therefore allow to specifically investigate the effect of signal-induced correlations only (see figure 2.1). Furthermore, the Poisson distribution is mathematically tractable and allows for analytic calculations [42, 264, 266]. The notation in the following implies $s(t) = r(t)$, which introduces an equivalence of the signal and rate trajectories. Temporal correlations in the firing rate are then equivalent to the temporal correlations in the signal. However, generally the results presented below are also valid

[1]Poisson firing can account well for the irregular firing pattern and noise characteristics of cortical neurons [43, 45, 47–49, 264, 265] (see section 1.1.4.3).

for other deterministic rate encodings $r(s(t))$, where the rate trajectory can be considered to be the signal itself[2]. Rate correlations, in this case, may not be exclusively signal-induced. Importantly, as I explain below, the information measure derived for Poisson neurons can also be utilized for non-Poisson spiking neurons where it can be used to dissect the effects of intrinsic and signal-induced rate correlations.

2.1.2 Information carried by correlated and independent spikes of Poisson neurons

The information per spike contained in entire spike trains of *correlated* spikes – also referred to as "correlation method" – in its general form is given by $\mathcal{I}_{\text{tot}}$ [see Eq. (1.23)] and is determined through the spike cross- and autocorrelation functions $C_{\text{cross}}(f)$ and $C_{\text{auto}}(f)$, respectively. Notably, $\mathcal{I}_{\text{tot}}$ is exact under the assumption of stationary signals and finite memory in the neurons but does not require Gaussianity in the input currents (see sections 1.3.2.2 and 3.4.2). In Poisson neurons, spike correlations only emerge from signal-induced correlations in the rate trajectory (cf. figure 2.1) and are equivalent to rate correlations. In this case, the Fourier transformed correlation functions Eq. (1.20) and (1.22) can be expressed in terms of the mean firing rate ν and the rate power spectrum $S_{rr}(f) = |\tilde{r}(f)|^2$ only (a detailed derivation is given in Appendix A.1):

$$C_{\text{cross}}(f) = S_{rr}(f), \tag{2.1}$$

$$C_{\text{auto}}(f) = C_{\text{cross}}(f) + \nu. \tag{2.2}$$

This is consistent with the fact that the cross-correlation is equivalent to the autocorrelation of the PSTH (peri-stimulus time histogram) [98], which in turn is given by $r(t)$. It follows that under the assumption of only rate-induced spike correlations $\mathcal{I}_{\text{tot}}$ can be written as

$$\mathcal{I}_{\text{corr}} = -\frac{1}{\nu}\int_0^\infty \mathrm{d}f \log_2\left(1 - \frac{S_{rr}(f)}{\nu + S_{rr}(f)}\right). \tag{2.3}$$

[2]Technically, this holds if there exists a *homeomorphism*, i.e. a smooth and uniquely invertible function, that maps the signal value $s(t)$ to the firing rate $r(t)$ and vice versa. Then, the mutual information of signal and spike trains and the mutual information of firing rate and spike trains are identical [267] (under the given assumption that spike trains are generated as a function of the rate only).

The subscript "corr" is chosen in order to distinguish Eq. (2.3) from $\mathcal{I}_{\text{tot}}$ and to emphasize that $\mathcal{I}_{\text{corr}}$ captures the effect of rate correlations, as opposed to $\mathcal{I}_{\text{ind}}$.

It is interesting to note that, given a linear relation between signal and firing rate, $\mathcal{I}_{\text{corr}}$ coincides with the information lower bound $\mathcal{I}_{\text{ld}}$ [Eq. (1.24)]. This is due to the fact that, in this case, the response-response cross-correlation C_{cross} and the signal-response cross-correlation $C_{s\rho} = S_{s\rho}$ are equivalent and both given by S_{rr}[3].

As a reminder, the information carried by (assumedly) *independent* spikes that are generated via stationary point processes with time-dependent firing rate $r(t)$ and overall rate distribution $p(r)$ is given by [see Eq. (1.28) and (1.29)]

$$\mathcal{I}_{\text{ind}} = \frac{1}{T}\int_0^T \mathrm{d}t\ \frac{r(t)}{\nu}\log_2\left(\frac{r(t)}{\nu}\right) = \left\langle \frac{r(t)}{\nu}\log_2\left(\frac{r(t)}{\nu}\right)\right\rangle_s = \int_0^\infty \mathrm{d}r\ p(r)\frac{r}{\nu}\log_2\left(\frac{r}{\nu}\right). \tag{2.4}$$

[Here, ν is the mean of $r(t)$.] It is important to note that Eq. (2.3) can be computed knowing only $r(t)$ and therefore does not require more knowledge than Eq. (2.4). Both expressions can be determined in terms of the PSTH, which corresponds to $r(t)$ and is experimentally accessible. Importantly, for non-Poisson neurons Eq. (2.3) can be considered an approximation of the information transmission when only rate correlations are taken into account and all other correlations (e.g. intrinsic or noise correlations) are neglected.

The obtained expression for $\mathcal{I}_{\text{corr}}$ allows for an analytic and exact investigation of the role of spike correlations for neural information transmission. The difference $\mathcal{I}_{\text{ind}} - \mathcal{I}_{\text{corr}}$ is the redundancy among spikes that arises from signal correlations in Poisson neurons, or generally from rate correlations as obtained from the PSTH. Moreover, the difference $\mathcal{I}_{\text{ind}} - \mathcal{I}_{\text{tot}}$ is the redundancy introduced by all present spike correlations and therefore offers a way to compute the difference in information if all but rate correlations are ignored.

[3] Setting $C_{s\rho} = S_{rr}$ and $C_{\text{auto}} = S_{rr} + \nu$ in Eq. (1.24), yields $\mathcal{I}_{\text{corr}}$; see Appendix A.1. A more general discussion of the relation of $\mathcal{I}_{\text{tot}}$ and $\mathcal{I}_{\text{ld}}$ is given in section 3.7.

2.2 Example processes for the rate trajectories

Most of the results of this chapter are illustrated, supplemented or supported by means of specific stochastic ensembles describing the rate (signal) trajectories $r(t)$. Thereby, using adequately parametrized dynamic processes allows to specify $r(t)$ in Eqs. (2.3) and (2.4), and to quantitatively analyze and visualize the influence of different signal properties, as is done in section 2.3. I introduce the example rate-generating processes in this section.

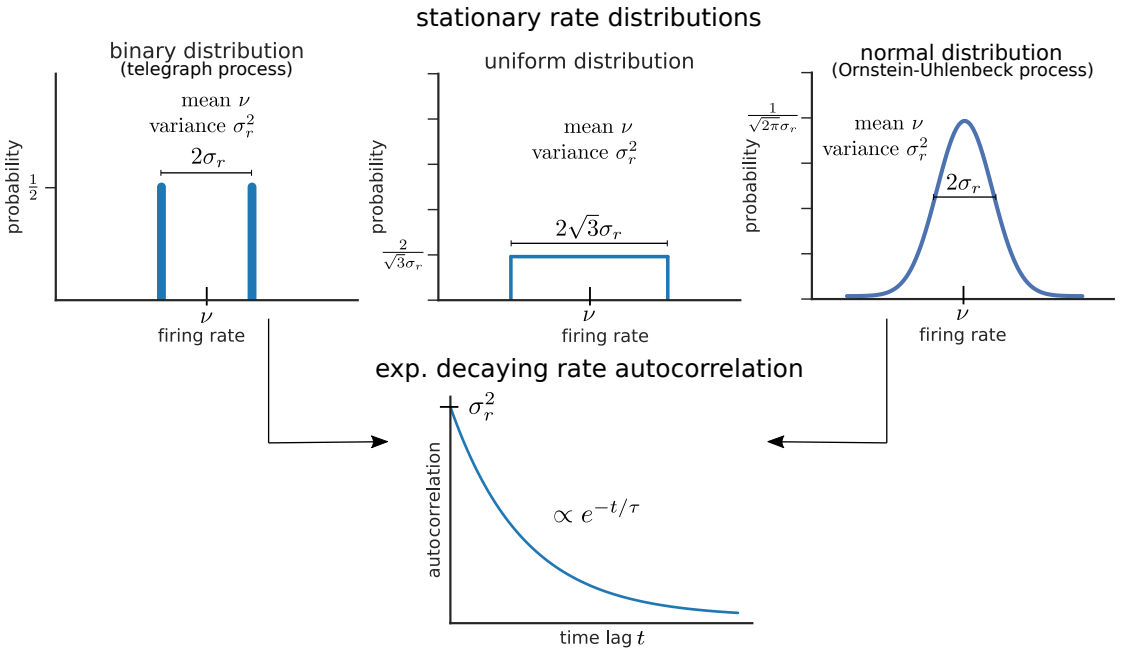


FIGURE 2.2: **Processes that describe firing rate dynamics are characterized by their stationary probability distribution and autocorrelation function.** The telegraph process is a process that jumps between two discrete values and, therefore, the stationary rate distribution is binary (top left). The next rate process possesses a uniform rate distribution (top center). The third process is an Ornstein-Uhlenbeck process with a normal distribution of the rates (top right). Here, all processes are assumed to have an exponentially decaying autocorrelation function (in the time domain; bottom). This dynamic property is independent of the stationary rate distributions. The processes are specified by the mean rate ν, rate variance σ_r^2, and correlation time τ. More details are given in the main text.

All processes for the rate trajectory $r(t)$ are assumed to be stationary. Their stationary probability distribution, i.e. the distribution of $r(t)$ across long time windows $T \to \infty$, has mean (firing rate) ν and standard deviation σ_r:

$$\langle r(t) \rangle = \frac{1}{T} \int_0^T \mathrm{d}t \; r(t) = \nu, \tag{2.5}$$

$$\langle (r(t) - \nu)^2 \rangle = \frac{1}{T} \int_0^T \mathrm{d}t \; (r(t) - \nu)^2 = \sigma_r^2. \tag{2.6}$$

Moreover, as discussed below, the dynamic properties of the processes are characterized by their temporal correlation function that has decay time τ.

2.2.1 Telegraph process

The first stochastic process that I consider is the *telegraph process*, also called random telegraph noise. This process is a binary process that here assumes the values $\nu+\sigma_r$ and $\nu-\sigma_r$. Jumping between these two processes occurs randomly with a specified rate that determines the correlation time of the process (see below) [268]. The telegraph process resembles the switching between up- and down-states of cortical activity [72, 269, 270]. Here, I assume that both values occur with equal probability (see figure 2.2). Hence,

$$p(r = \nu + \sigma_r) = p(r = \nu - \sigma_r) = 1/2. \tag{2.7}$$

Obviously, it is required that $\sigma_r \leq \nu$.

2.2.2 Process with uniform distribution

In the second rate process, the firing rate is assumed to follow a uniform distribution with a given mean ν and variance σ_r^2. Then, the stationary probability for a rate r is

$$p(r) = \begin{cases} \frac{1}{2\sigma_r\sqrt{3}} & \text{if } -\sigma_r\sqrt{3} \leq r - \nu \leq \sigma_r\sqrt{3}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

This imposes a constraint on the maximum variance at a given mean rate to prevent overmodulation and mathematical inconsistencies: I require $\sigma_r \leq \nu/\sqrt{3}$.

2.2.3 Ornstein-Uhlenbeck process

The third considered process is the Ornstein-Uhlenbeck process (OUP), also known as colored noise. OUP are band-pass filtered versions of Gaussian white noise and have been found to be a good description of neuronal activity [32, 83, 271] (see also section 1.4.1). For rates following an OUP with mean ν and variance σ_r^2 the stationary probability

distribution is given by a normal distribution,

$$p(r) = \mathcal{N}(r|\nu, \sigma_r^2). \tag{2.9}$$

Importantly, this normal distribution does not bound r and technically involves negative rates. In the results below, integrations over $p(r)$ are restricted to non-negative rates, resulting in approximations to the mutual information.

At the same mean and variance, the uniform, telegraph, and Ornstein-Uhlenbeck processes correspond to fundamentally different rate distributions and hence allow for comparison of the effects of differences in the higher moments of the rate distribution.

2.2.4 Exponentially decaying autocorrelation

Generally, the probability distribution of a stochastic process is – apart from mean and variance – independent of its autocorrelation function. In fact, stochastic processes with an arbitrary probability distribution and autocorrelation can be generated [272, 273]. An important class of stochastic processes are those with exponentially decaying autocorrelations. Here, in contrast to section 1.4.1, the rate rather than the synaptic current is modeled through such an autocorrelation. The autocorrelation of a (signal-dependent) rate process $r(t)$ then reads

$$\frac{1}{T}\int_0^T \mathrm{d}t\; r(t)r(t-h) - \nu^2 = \sigma_r^2\; e^{-|h|/\tau}, \tag{2.10}$$

where σ_r^2 and τ are the variance and correlation time, respectively. The corresponding power spectrum of a firing rate trajectory with exponentially decaying autocorrelation is given by

$$S_{rr}^{\exp}(f) = \frac{\sigma_r^2\tau}{\pi(1+f^2\tau^2)}. \tag{2.11}$$

Telegraph and Ornstein-Uhlenbeck processes naturally have exponentially decaying autocorrelations [268]. This assumption is also made for the process with uniform distribution in order to completely specify its dynamical properties.

2.3 Analyzing factors of encoding redundancy in rate codes

As shown above, the information in independent and correlated spikes can both be expressed as a function of the rate trajectory, but it is different features of that trajectory that determine the information in either case. In this section, I elaborate on these differences between $\mathcal{I}_{\text{ind}}$ and $\mathcal{I}_{\text{corr}}$ in more detail. This is partly done using the three previously introduced examples processes for the rate/signal.

In section 2.3.1, I discuss the upper bound property of $\mathcal{I}_{\text{ind}}$. In section 2.3.2, the role of the baseline firing rate is investigated, and in sections 2.3.3 and 2.3.4 I consider the limits of very small modulation depths and correlation times.

The `Mathematica` and `Python` code for the result figures and analytic calculations of this section can be found on this[4] github repository.

2.3.1 Assuming independent spikes yields an upper bound approximation for information transmission

Considering independent spikes means neglecting all temporal spike correlations, both intrinsic and signal-induced. Accordingly, $\mathcal{I}_{\text{ind}}$ is entirely determined by the distribution of $r(t)$, and its computation does not depend on temporal correlations. On the other hand – considering Poisson neurons with rate correlations – $\mathcal{I}_{\text{corr}}$ depends on the power spectrum of $r(t)$ [see Eq. (2.3)]. I discuss the implications of this difference for information transmission by means of the introduced rate processes in the following.

The information in spikes that are assumed to transmit information independently can be computed for the three processes using Eq. (2.4) and Eq. (2.8)–(2.9). It is

$$\mathcal{I}_{\text{ind}}^{\text{tele}} = \frac{1}{2} \sum_{k=1,-1} \frac{\nu + k\sigma_r}{\nu} \log_2 \left(\frac{\nu + k\sigma_r}{\nu} \right), \tag{2.12}$$

[4] https://github.com/t8ch/dissertation-code/tree/master/Ch2

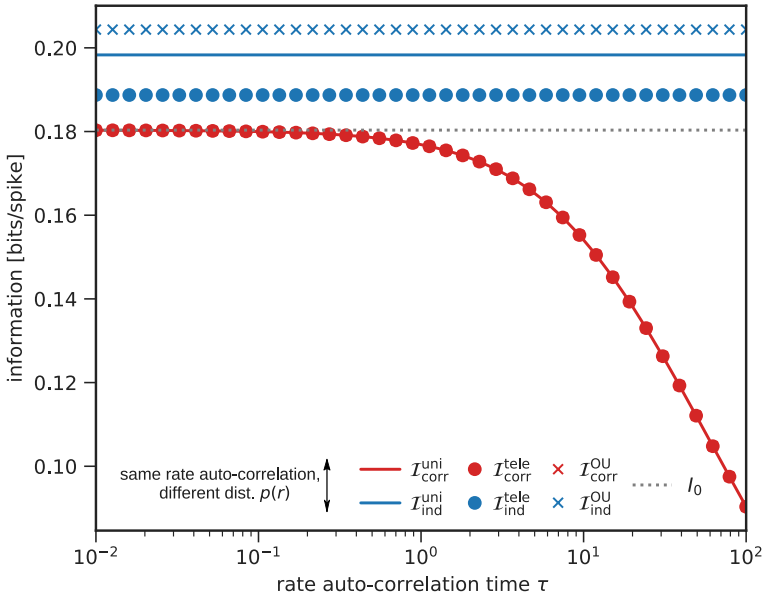


FIGURE 2.3: **Signal power spectrum determines the information contained in correlated spikes, whereas the information in independent spikes is given through the signal distribution.** The red line and symbols show the information in correlated spikes for a uniform, a telegraph, and an Ornstein-Uhlenbeck process, respectively. It is the same in all cases and varies with τ. On the other hand, the information contained in independent spikes (blue line and symbols) does not vary with τ but is different for all processes despite their identical spectra. Here, the differences in the probability distributions of $r(t)$ in the three processes lead to different encoding redundancies. The dashed line is the analytic limit $\lim_{\tau\to 0}\mathcal{I}_{\text{corr}} = \mathcal{I}_0 = \frac{\sigma_r^2}{2\log(2)\nu^2}$ [Eq. (2.24)]. For all curves it is $\nu = 1$ and $\sigma_r = 0.5\nu$.

for the telegraph process and

$$\begin{aligned}\mathcal{I}_{\text{ind}}^{\text{uni}} &= \frac{1}{2\sigma_r\sqrt{3}}\int_{\nu-\sqrt{3}\sigma_r}^{\nu+\sqrt{3}\sigma_r} \mathrm{d}r \frac{r}{\nu}\log_2\left(\frac{r}{\nu}\right) \\ &= \frac{-\left(\nu^2+3\sigma_r^2-2\sqrt{3}\nu\sigma_r\right)\log\left(1-\frac{\sqrt{3}\sigma_r}{\nu}\right)}{4\sqrt{3}\nu\sigma_r\log(2)} \\ &+ \frac{\left(\nu^2+3\sigma_r^2+2\sqrt{3}\nu\sigma_r\right)\log\left(\frac{\sqrt{3}\sigma_r}{\nu}+1\right)-2\sqrt{3}\nu\sigma_r}{4\sqrt{3}\nu\sigma_r\log(2)},\end{aligned} \tag{2.13}$$

assuming a uniform rate distribution. For OUP, the information contained in independent spikes is

$$\mathcal{I}_{\text{ind}}^{\text{OU}} \approx \frac{1}{\sqrt{2\pi\sigma_r^2\nu^2}}\int_0^\infty \mathrm{d}r\ \exp\left(\frac{-(r-\nu)^2}{2\sigma_r^2}\right) r\log_2\left(\frac{r}{\nu}\right). \tag{2.14}$$

Here, the integration must be truncated to avoid negative firing rates. As a consequence, $\mathcal{I}_{\text{ind}}^{\text{OU}}$ is an approximation which becomes more accurate for smaller σ_r/ν. However, this

does not affect the generality of the following results since all results are confirmed by and consistent with the analytic and exact results for the other processes.

According to Eq. (2.3), calculating the information in rate-correlated spikes requires knowledge of the rate power spectrum. All processes are assumed to possess an exponentially decaying autocorrelation function and the corresponding power spectrum Eq. (2.11). Using Eq. (2.3), it follows that the information in correlated spikes at the given rate power spectrum is the same for the three processes,

$$\mathcal{I}_{\text{corr}}^{\text{exp}} = \mathcal{I}_{\text{corr}}^{\text{uni}} = \mathcal{I}_{\text{corr}}^{\text{tele}} = \mathcal{I}_{\text{corr}}^{\text{OU}}, \tag{2.15}$$

and reads

$$\begin{aligned} \mathcal{I}_{\text{corr}}^{\text{exp}} &= -\frac{1}{2\nu}\int_{-\infty}^{\infty} \mathrm{d}f \log_2\left(1 - \frac{S_{rr}^{\text{exp}}(f)}{\nu + S_{rr}^{\text{exp}}(f)}\right) \\ &= \frac{\pi^{3/2}\nu - \pi\sqrt{\nu\left(\pi\nu + \sigma_r^2\tau\right)} + \sqrt{\pi}\sigma_r^2\tau}{\tau\log(2)\sqrt{\nu^3\left(\pi\nu + \sigma_r^2\tau\right)}}. \end{aligned} \tag{2.16}$$

With the results Eq. (2.12)–(2.14) and (2.16) at hand, I make some interesting observations. First, even though all processes have the same spectral decomposition and variance, the information in spikes that are assumed to be independent is different because the distributions of $r(t)$ are different. Accordingly, if the overall distribution of $r(t)$ is fixed but the correlation time is varied, $\mathcal{I}_{\text{ind}}$ remains unchanged for the three processes. On the other hand, the information contained in correlated spikes at a given signal power spectrum does not depend on the distribution of $r(t)$ (see also figure 2.3).

As a general finding, in rate modulated Poisson neurons the inclusion of signal-induced spike correlations always diminishes the information contained in spike trains compared to when spikes are treated as independent (cf. figures 2.3, 2.5 and 2.6). This redundancy among spikes has been proposed before [227, 236]. Qualitatively speaking, the redundancy arises as follows: a spike following another spike that carries information about a correlated signal does not provide entirely new, or independent, information. This is because knowledge about the signal at the time of one spike also provides information about the signal at later times that are not much larger than the signal correlation time. Since, by design, $\mathcal{I}_{\text{ind}}$ treats spikes as independent information carriers, irrespective of their temporal proximity, it overestimates the actual transmitted information

$\mathcal{I}_{\text{corr}}$. From this picture also follows that the redundancy decreases for smaller signal correlation times as is demonstrated in figure 2.3 and section 2.3.4.

2.3.2 $\mathcal{I}_{\text{ind}}$ only depends on relative modulation depth

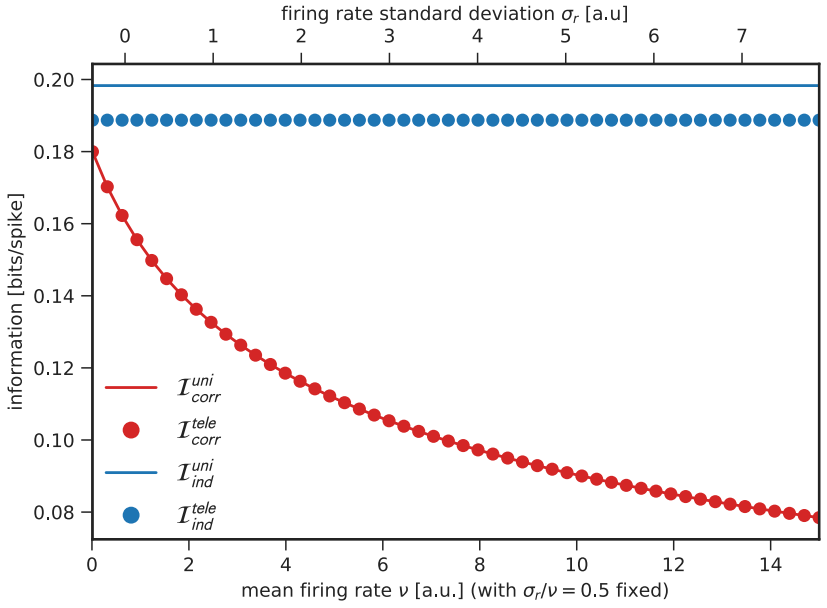


FIGURE 2.4: **For a given type of rate distribution, the information in independent spikes is determined by σ_r/ν.** Lines correspond to uniform rate processes and circles correspond to telegraph processes (OUP not shown here). The information in independent spikes (blue line and circles) does not depend on the mean firing rate ν if σ_r is chosen such that the ratio σ_r/ν is constant (in this figure it is $\sigma_r = 0.5\nu$ for all data points). Incorporating correlations in the information yields a strong decrease in information transmission per spike for increasing ν even at constant σ_r/ν (red line and circles). Analogously to figure 2.3, $\mathcal{I}_{\text{ind}}$ is different for all processes whereas $\mathcal{I}_{\text{corr}}$ is only a function of the rate autocorrelation and not sensitive to the distribution of $r(t)$. In this figure it is τ=10 [in inverse units of ν].

I make another noteworthy observation for the information in independent spikes. Rewriting Eq. (2.4) in terms of a rescaled firing rate yields

$$\begin{aligned}\mathcal{I}_{\text{ind}} &= \frac{1}{T}\int_0^T \mathrm{d}t\ \frac{r(t)}{\nu}\log_2\left(\frac{r(t)}{\nu}\right)\\ &= \frac{1}{T}\int_0^T \mathrm{d}t\ \frac{r_0(t)\sigma_r+\nu}{\nu}\log_2\left(\frac{r_0(t)\sigma_r+\nu}{\nu}\right)\\ &= \frac{1}{T}\int_0^T \mathrm{d}t\ (\alpha r_0(t)+1)\log_2\left(\alpha r_0(t)+1\right). \end{aligned} \tag{2.17}$$

Here, a linear rescaling $r(t) \to r_0(t)\sigma_r + \nu$ was applied such that $\langle r_0 \rangle = 0$ and $\mathrm{Var}(r_0) = 1$ (z-score normalization). Moreover, in the last step the relative modulation depth $\alpha \equiv \sigma_r/\nu$ was defined. From Eq. (2.17) follows that for any simultaneous change of ν and σ_r that leaves the ratio α unchanged the information in independent spikes remains the same because $r_0(t)$ by definition does not change under rescaling. Therefore, $\mathcal{I}_{\mathrm{ind}}$ only depends on the ratio σ_r/ν for a given type of stationary rate distribution. The information in rate-correlated spikes, on the other hand, strongly depends on the mean firing rate even if α is kept constant. This is demonstrated in figure 2.4.

The finding for independent spikes implies that the rate of information transmission, i.e. $\nu \cdot \mathcal{I}_{\mathrm{ind}}$ (bits per second), can become arbitrarily high by increasing the mean firing rate even at a small, finite modulation depth. More precisely, $\nu \cdot \mathcal{I}_{\mathrm{ind}}$ scales linearly with the firing rate at a given modulation depth. This does not apply to $\mathcal{I}_{\mathrm{corr}}$ where the decrease in the information per spike implies a sublinear, saturating increase in the overall information transmission $\nu \cdot \mathcal{I}_{\mathrm{corr}}$ for increasing ν.

2.3.3 Vanishing rate modulations render spikes independent irrespective of temporal correlation structure

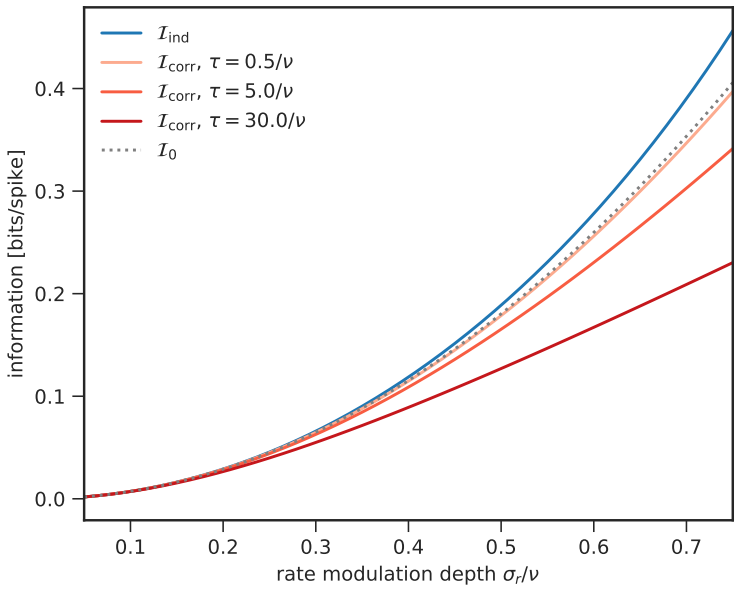


FIGURE 2.5: **Temporal correlations introduce redundancy (example of a telegraph process).** The information in independent spikes is always larger than that in correlated spikes. Shown here is the information per spike for telegraph processes with σ_r/ν as given on the x-axis. In the limit of very small rate modulations [$\sigma_r \to 0$], both cases converge to $\mathcal{I}_0$ [Eq. (2.20), shown as dashed line]. For all curves it is $\nu = 2$.

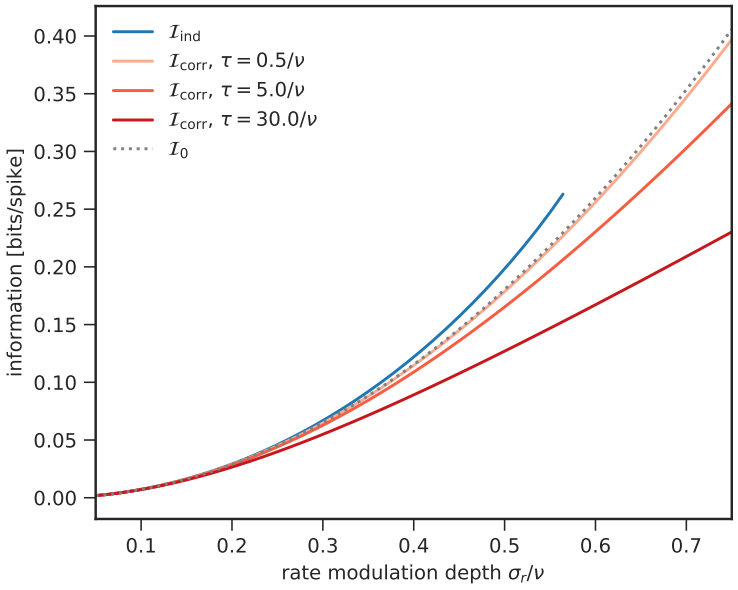


FIGURE 2.6: **Temporal correlations introduce redundancy (example of a process with uniform rate distribution).** Same as figure 2.6 but for a uniform rate distribution. The curve of $\mathcal{I}_{\text{ind}}$ is truncated at high σ_r/ν due to overmodulations that would correspond to negative firing rates (see section 2.2).

As discussed above, correlated rate modulations diminish the transmitted information. The impact of correlations on the mutual information can be expected to vanish in the limit $\sigma_r \to 0$ because in this limit correlations trivially vanish, too. In the following, I show how this limit indeed leads to an equivalence of $\mathcal{I}_{\text{ind}}$ and $\mathcal{I}_{\text{corr}}$.

Because $[r(t) - \nu] \to 0$ when $\sigma_r \to 0$, one can expand the integrand of Eq. (2.4) in $r(t)$ around ν. Expanding up to second order yields

$$\mathcal{I}_{\text{ind}} \approx \frac{1}{T} \int_0^T \mathrm{d}t \; \frac{r(t) - \nu}{\nu \log(2)} + \frac{(r(t) - \nu)^2}{2\nu^2 \log(2)} = \frac{\sigma_r^2}{2 \log(2) \nu^2}. \tag{2.18}$$

The first term of the integrand vanishes because the time average of $r(t)$ is ν, and the time average of the numerator of the second term is the rate variance. This result is consistent with second order expansions in σ_r of $\mathcal{I}_{\text{ind}}^{\text{tele}}$ and $\mathcal{I}_{\text{ind}}^{\text{uni}}$ [Eq. (2.12) and (2.13)] for the uniform and telegraph processes.

For $\mathcal{I}_{\text{corr}}$ the limit $\sigma_r \to 0$ implies that one can expand $\mathcal{I}_{\text{corr}}$ in $S_{rr}(f)$ at each frequency[5]. Expanding to first order in $S_{rr}(f)$ leads to

$$\mathcal{I}_{\text{corr}} \approx \frac{1}{2\log(2)\nu^2}\int_{-\infty}^{\infty} \mathrm{d}f\ S_{rr}(f) = \frac{\sigma_r^2}{2\log(2)\nu^2}. \tag{2.19}$$

Thus, in the limit of vanishing rate modulations, both information estimates are equivalent as they possess the same leading order expansion in σ_r which is given by

$$\mathcal{I}_0 \equiv \frac{\sigma_r^2}{2\log(2)\nu^2}. \tag{2.20}$$

This result emphasizes that both the information in independent and correlated spikes are in leading order determined by the rate processes' variance, which corresponds to the modulation depth around the baseline firing rate, or dynamic range, irrespective of other statistical properties of $r(t)$. The asymptotic convergence of the two information measures for vanishing modulations depths is also observable in figures 2.5 and 2.6.

2.3.4 Information in correlated spikes is maximized by vanishing correlation times

Intuitively, another way to reduce the impact of signal-induced rate correlations is to decrease the signal correlation time τ, because this reduces the width of the autocorrelation function. In what follows, I investigate the limit of vanishingly small correlation times of the information carried by correlated spikes and uncover a somewhat surprising discrepancy between $\mathcal{I}_{\text{ind}}$ and $\mathcal{I}_{\text{corr}}$ in this limit.

A vanishing correlation time is equivalent to a flat signal power spectrum (cf. the power spectrum Eq. (2.11) as an example). As a rule of thumb, it holds that "narrow functions have wide Fourier transforms, and vice versa"[6]. Therefore, I now consider a rate power spectrum $S_{rr}(f)$ that is constant for all frequencies $|f| \leq f_c$ and zero otherwise (e.g. band-limited white noise). Later I consider the limit $f_c \to \infty$ and confirm the result by explicitly taking the limit $\tau \to 0$ in the example rate processes. For a given rate variance

[5] Because the variance is equivalent to the integral of the power spectrum over all frequencies and the power spectrum is non-negative [30].

[6] For example, stated on this web page hosted by the University of California, Berkeley: https://casper.ssl.berkeley.edu/astrobaki/index.php/Fourier_Transform

σ_r^2 and flat power spectrum it is

$$\int_{-f_c}^{f_c} \mathrm{d}f \; S_{rr}(f) = \sigma_r^2 \quad \Rightarrow \quad S_{rr}(f) = \sigma_r^2/(2f_c). \tag{2.21}$$

Inserting in Eq. (2.3) and expansion in $1/(2f_c)$ yields

$$\mathcal{I}_{\mathrm{corr}} = -\frac{1}{2\nu} \int_{-f_c}^{f_c} \mathrm{d}f \log_2 \left(1 - \frac{\sigma_r^2/(2f_c)}{\nu + \sigma_r^2/(2f_c)}\right) \tag{2.22}$$

$$= \frac{f_c}{\nu} \left[\frac{\sigma_r^2/(2f_c)}{\nu \log(2)} + \mathcal{O}\left(\left(\frac{\sigma_r^2}{2f_c}\right)^2\right)\right]. \tag{2.23}$$

Now, the limit $\tau \to 0$ of vanishing correlation time corresponds to taking the limit $f_c \to \infty$ in the last equation. This limit precisely yields $\mathcal{I}_0$ as defined in Eq. (2.20):

$$\lim_{\tau \to 0} \mathcal{I}_{\mathrm{corr}} = \mathcal{I}_0 \leq \mathcal{I}_{\mathrm{ind}}. \tag{2.24}$$

This result is consistent with the limit $\tau \to 0$ of Eq. (2.16), representing the special case of an exponentially decaying autocorrelation function. In figures 2.3 and 2.5, the result of Eq. (2.24) is illustrated by the fact that the curves of $\mathcal{I}_{\mathrm{corr}}$ approach that of $\mathcal{I}_0$ for decreasing correlation time. At a given rate variance, $\mathcal{I}_0$ serves as an upper bound for $\mathcal{I}_{\mathrm{corr}}$.

Interestingly, however, $\mathcal{I}_0$ is equivalent to the information in independent spikes only in the limit of vanishing rate modulations. Therefore, vanishing rate correlation times maximize the information in correlated spikes for an otherwise fixed rate process but do not generally coincide with the case where spikes carry information independently (cf. Fig. 2.3). From one perspective, this can be explained by the cumulative redundant effect of the spike correlations across frequencies, here captured by σ_r^2, which does not have to be small even if $S_{rr}(f)$ is small everywhere. Moreover, the result of Eq. (2.24) expresses an essential difference between the notions of independent spikes in $\mathcal{I}_{\mathrm{ind}}$ and the zero correlation time limit of $\mathcal{I}_{\mathrm{corr}}$. The derivation of $\mathcal{I}_{\mathrm{ind}}$ not only relies on the assumption of statistically independent firing rates at different times but additionally assumes that the signal entropy is that of a homogeneous Poisson process [186, 227, 236] with rate ν and no rate modulations. The model thereby omits the rate variance as a general and not signal-specific property of the observed rate trajectories. The zero correlation time limit of $\mathcal{I}_{\mathrm{corr}}$, on the other hand, takes into account that rate modulations (with

variance σ_r^2) are present, albeit independent across time. The issue is elaborated on in discussion section 2.4.3.

2.4 Summary and discussion

A long-standing challenge in neuroscience has been to understand the code that is used in sequences of stereotyped action potentials (spike trains) to transmit information about relevant signals. Earlier studies have suggested that the efficiency of the neural code can be increased when temporal correlations in the spike trains are minimized, or equivalently, that temporal correlations in spike trains introduce redundancy to signal encoding [205, 227, 234, 235]. This coding redundancy among spikes can be quantified as the (negative) difference between the mutual information in spikes about a signal that is carried by the spike trains as a whole, including all correlations within, and the one carried by their individual spikes, if they are treated as being independent [25, 184, 227]. However, exact methods to simultaneously determine both quantities were lacking until recently and therefore quantifying the effect of temporal spike correlations on coding redundancy remained an open challenge. For example, fundamental open questions are (cf. section 1.3.2.2):
How can the impact of different temporal spike correlation structures on coding redundancy be quantified? And what are the properties of a rate code that determine these effects?

In this chapter, I provided a general method to compute the information carried by spikes with correlated rate modulations only, neglecting other spike correlations, and used it to investigate the effect of rate correlations on encoding redundancy. I derived it analytically by calculating the mutual information between a time-correlated, rate modulating signal and the resulting spikes of Poisson neurons. Based on this method and contrasting it with the information in (supposedly) independent spikes, I investigated how different properties of signal-induced firing rate trajectories shape information transmission and linked the results to other related findings. The general findings were complemented and illustrated by means of three example rate processes. Overall, the results shed light on the role of signal-induced temporal correlations for neural coding, by providing insight into how signal features shape redundancy and by establishing mathematical links between existing methods.

In sections 2.4.1 and 2.4.2, I summarize and discuss the general methods and the results that were derived by applying them. A specific discussion of related other studies is given in section 2.4.3.

2.4.1 A method to determine the impact of temporal correlations

The main results of section 2.1 and 2.3.1 are concerning a methodical framework for determining the effect of temporal spike correlations on information transmission. They can be summarized as follows:

- Using analytic calculations, I derived an expression that yields the complete information contained in rate modulations while neglecting the effects of intrinsic correlations ($\mathcal{I}_{\text{corr}}$, see section 2.1.2).
- I illustrated that the information in independent and rate-correlated spikes is completely determined by first-order (overall distribution) and second-order (autocorrelation) statistics of the rate trajectory, respectively. As a consequence, the information in rate modulated Poisson neurons only depends on the PSTH autocorrelation, regardless of other properties of the rate trajectory (see figure 2.3).
- Contrasting $\mathcal{I}_{\text{ind}}$ and $\mathcal{I}_{\text{corr}}$, I confirmed that temporal spike correlations introduce redundancy to information encoding, as stated previously [227]. Hence, $\mathcal{I}_{\text{ind}}$ can be considered an upper bound approximation of $\mathcal{I}_{\text{corr}}$ (section 2.3.1 and figure 2.5).

These results provide a general framework for computing the redundancy in the neural code, dissecting the effects of rate correlations and other temporal correlations. The introduced methodology, on the one hand, allows investigating the impact of signal properties on information encoding on a fundamental level. At the same time, $\mathcal{I}_{\text{corr}}$ [Eq. (2.3)] can be applied to experimentally recorded spike trains to obtain a better estimate of the mutual information that includes the effect of rate correlations and does not require more knowledge than $\mathcal{I}_{\text{ind}}$.

To derive $\mathcal{I}_{\text{corr}}$, I assumed Poisson firing, which has been proven a valid assumption for cortical neurons in many cases but may not always hold [43, 50]. Nonetheless, $\mathcal{I}_{\text{corr}}$ can be computed for any spiking data irrespective of their correlation structure as long as stationarity criteria are fulfilled [98, 227]; e.g. data obtained from simulations, recordings, or analytical calculations. The value of $\mathcal{I}_{\text{corr}}$ then corresponds to the information

carried by rate modulations only, ignoring other spike correlations and temporal codes with high temporal precision, and hence can also be used if the Poisson assumption does not hold (and is easy to compute). The correlation method captures the effect of all correlations. Thus, one can dissect the effects of rate correlations and other correlations, such as intrinsic noise correlations, by comparing $\mathcal{I}_{\text{corr}}$ to $\mathcal{I}_{\text{tot}}$ and $\mathcal{I}_{\text{ind}}$, if these can be computed.

In a linear rate code, i.e. if $r(t) \propto s(t)$, signal correlations and rate correlations are equivalent (cf. figure 2.1). Often, linear approximations of the input-output relations are appropriate [1] and can be obtained analytically for certain neuron models [108, 260, 274] (see sections 1.3.1.3 and 4.2). Then, the difference of $\mathcal{I}_{\text{corr}}$ and $\mathcal{I}_{\text{ind}}$ measures the redundancy effect of signal-induced correlations. However, in neural circuits, stimulus changes are not generally translated into proportional rate changes, and signal and rate correlations are not necessarily equivalent [29]. If the assumptions of rate encoding and linearity are not given, a comparison of $\mathcal{I}_{\text{ind}}$ and $\mathcal{I}_{\text{tot}}$ can be used to determine the combined effect of all present correlations, such as higher-order rate correlations [275] and intrinsic correlations [9]. Additionally, under the assumption of Poisson neurons, the difference of $\mathcal{I}_{\text{corr}}$ and $\mathcal{I}_{\text{tot}}$ reveals the impact of nonlinear, signal-dependent rate modulations on information transmission.

Here, I restricted my analysis to signal-induced rate correlations to introduce a general, analytic framework. However, incorporating more general signal-to-rate functions could certainly be an important focus for future experimental and theoretical studies. Applying my approach to *in vivo* data by calculating the respective first and second order spiking statistics could help shed light on the controversial role of spiking correlations for neural coding.

Neurons are embedded in networks in which inter-neural correlations can arise [188], even though these correlations are considered to vanish in balanced networks [66] (see sections 1.3.1.6 and 4.4 for more details). In particular, noise correlations can generally have additional synergetic or redundant effects on population encoding (see [190, 196] for reviews). Here, signal correlations are incorporated through the signal-carrying rate trajectories, and noise correlations are not present in the Poisson framework but have been found to be small [198]. Moreover, because my work represents the limit of disconnected neurons, it can promote network-level studies by providing a reference for redundancy

effects that are not mediated by inter-neural coupling but by temporal interactions only. Future extensions of the present work to population encoding could be guided by previous studies that include cross-neural noise and signal correlations [181–184, 276] (see also section 4.7.4). These studies are discussed in section 2.4.3.

2.4.2 Parameter dependence of encoding redundancy

In section 2.3, I investigated the roles of mean firing rate, rate modulation depth, and rate autocorrelation time for encoding redundancy, and I obtained these results:

- For a given rate distribution, $\mathcal{I}_{\text{ind}}$ only depends on the rate modulation depth and, therefore, the information rate in independent spikes scales with the firing rate (section 2.3.2 and figure 2.4).
- There is a mathematical limit at which the information for independent spikes and the correlation method converge to an expression that only depends on the dynamic range of the rate trajectory (section 2.3.3 and figures 2.5 and 2.6).
- Signals with a vanishing correlation time (a flat spectrum) maximize the information in correlated spikes but surprisingly are not equivalent to independent spikes in terms of information transmission, even in Poisson neurons (section 2.3.4 and figure 2.3).

The results of section 2.3.2 suggest that independent spikes can (putatively) transmit arbitrarily large amounts of information, in terms of the information rate, if the mean firing rate of a modulated rate trajectory is increased. However, for real biological systems, the amount of actually transmittable information in Poisson neurons is limited by several factors. First and foremost, normally correlations in the rate trajectory have to be expected. In this case, the information is correctly determined by $\mathcal{I}_{\text{corr}}$, and the corresponding information rate scales with ν such that it converges to a finite value (see section 2.3.2). Only in the limit of vanishing correlations where the information can be reduced to $\mathcal{I}_0$ [Eq. (2.20)] is the linear scaling of the information rate with ν recovered (a similar optimization of the information rate through temporal whitening is discussed below). Moreover, biological firing rates are naturally limited, thereby setting an upper bound to the information rate $\nu \cdot \mathcal{I}_{\text{ind}}$. One also needs to bear in mind that the metabolic cost of spiking is generally neglected when optimization is considered in terms of information rates.

I proved mathematically that the information in correlated and independent spikes becomes identical in the limit of vanishing rate modulations (section 2.3.3). Whereas this appears to be an intuitive result, it has not been formally obvious because $\mathcal{I}_{\text{tot}}$ and $\mathcal{I}_{\text{ind}}$ have been formulated in the time and frequency domain, respectively. The finding links previous work on the Poisson limit [227], lower and upper bound calculations [218], and the more recent analysis of the complete mutual information [98]. At the same time, it provides an intuitive notion of the information transmission in a rate code at the lowest order: information is proportional to the squared coefficient of variation of the rate process. This, in turn, agrees with the view that the capacity of a rate code grows with its dynamic range – as long as approximately linear signal-rate relations hold. Importantly, my results are not restricted to Gaussian rate processes, and I showed that a given signal power spectrum can be realized through different (non-Gaussian) processes [277] (see section 2.2), which yield different redundancy.

From section 2.3.4 follows that it may be beneficial for a neural system to temporally decorrelate either its input statistics or intrinsically decorrelate the response to correlated inputs. This could be realized through low signal-dependent rate modulation depths. However, this, at the same time, reduces the overall transmitted information. As follows from figures 2.3, 2.5 and 2.6, a better strategy would be to minimize the correlation time in order to maximize information transmission. This is known as temporal whitening and can be implemented, e.g., through fast synapses or adaptation mechanisms [25, 56, 223, 278–280]. One of my findings is that even complete whitening does not lift the information to the level of genuinely independent spikes as characterized by $\mathcal{I}_{\text{ind}}$ (see also section 2.4.3 below). Whereas previous theoretical and experimental studies have shown that neurons can transmit information independently [29, 167, 201], it seems to be unclear whether whitening of the neural responses also implies independent spikes as is often assumed [25, 29]. However, remaining levels of redundancy after whitening may be beneficial and increase the robustness to noise by providing longer windows for reliable signal integration. This is particularly true if external noise sources are relevant. In my model framework, all noise effects are summarized in the Poisson spiking variability (Fano factor of 1), and noise sources cannot be separated.

In summary, my work provides analytically and experimentally accessible methods to compute the influence of rate correlations on redundancy in neural information encoding. These influences are reducible to simple first- and second-order statistics of the rate

process. Comparing my results to existing results reconciles different methodologies that have been present in the field.

2.4.3 Discussion of related work

Some studies had previously addressed the issue of information encoding redundancy in spikes both within and across cells (see e.g. [181, 182, 276, 281, 282] or [184, 185] for reviews). Here, I review these studies and relate them to the present work. I thereby focus on spike correlations in single neurons across time, which have been the subject here, rather than across neurons.

Other authors have devised approximate expansions of the mutual information around $\mathcal{I}_{\text{ind}}$ [Eq. (2.4)], incorporating the effect of signal and noise correlations in higher order corrections to this expression. Such expansions of the mutual information were done with respect to the recording length for spike count and rate codes [281] as well as temporal codes [182, 276]. These result in a systematic decomposition of the mutual information in parts containing different types of signal and noise cross- and autocorrelations. Investigating these components allows to determine whether the different correlation types add synergy or redundancy to the neural code. Interestingly, if Poisson spiking is assumed, the only remaining correction to the information is given in terms of signal correlations and always reduces the overall information, rendering Eq. (1.29) an upper bound, in accordance with my results (see e.g. section 2.3.4). The second-order decomposition of the mutual information is a good approximation for short windows of stimulus presentation and low firing rates [182]. Effectively, correlations between more than two spikes are assumed to be negligible for information transmission in this approximation[7] [184]. The correlation method and its Poisson neuron version, on the other hand, assume recording windows that are much longer than the signal correlation time and that contain many spikes. In this setting, the presented results are exact.

A similar expansion to [276] has been put forward by DeWeese [186, 236] for single cells only and comprises a cluster expansion in the signal correlation time around Eq. (1.29). This yields similar results as [276] and [182] but omits terms that contain spike correlations at equal times [182, 236]. Taking the limit of zero correlation time in this expansion

[7] The correlation method also relies on two-point correlation functions, but this is a result for very long recording lengths for which the Fourier statistics become multivariate normal (see section 3.4).

only leaves the leading order term $\mathcal{I}_{\text{ind}}$. This, however, is only an upper bound for the same limit of the correlation method Eq. (2.3), as I have demonstrate. Again, I argue that this discrepancy reflects the different assumptions about the signal-averaged spiking statistics in the two methods (Eq. (2.3) vs Eq. (2.4)). In one case, the zeroth order expansion of the signal entropy is that of a homogeneous Poisson process (with the signal-averaged firing rate) and hence oblivious of rate modulations. The correlation method takes these modulations into account and depends on the variance of the rate process (cf. Eq. (2.24)). Rate modulations, in this case, thus contribute to the information even if they are independent across time. Therefore, both methods provide different reference values for the information in independent spikes, depending on the moments of the stationary rate distribution that are considered relevant. Often, the effect of correlations is measured using a reference that is computed on shuffled spiking data [184, 185]. Shuffling a modulated rate process in time leaves both mean and variance unchanged. Therefore, I think the limit $\mathcal{I}_0$ serves as a good alternative reference for the information in independent spikes and could be used as such in future studies.

An exact decomposition of the mutual information has been put forward by Pola *et al.* [181] and enables the investigation of synergy/redundancy among spikes, including all effects of correlations within and across cells. Similar to [182, 276], the information is decomposed into a linear term, a signal-similarity term, a stimulus-independent correlational component, and a stimulus-dependent correlational component [181] (an analogous decomposition of the population information is presented in [184]). This method, thereby, generally allows for an exact and rigorous investigation of the effects of different kinds of correlations on the mutual information. In particular, the role of cross-neural correlations, which are not investigated here, can be addressed with this method. However, the decomposition method does not explicitly provide a means to determine the effect of temporal spike correlations on information transmission. In fact, considering single neurons, the only remaining term in the method of [181] is the so-called linear term and merely describes the information in single neurons in most general terms as mutual information of signal and spiking response (for which the correlation method provides an expression).

Because the method of [181] is based on computing probability distributions over all possible spike trains it is in practice limited by the analytical tractability of these probabilities and encounters problems with sampling biases for experimental data of limited

size [181, 184, 282] (cf. section 1.3.2.2). Other studies addressed this problem and introduced shuffling methods to reduce the bias in lower and upper bounds to the information [282, 283]. Their approach comprises an estimation of the role of correlations for information transmission in terms of general response distributions that had previously been proposed [183, 201] (and whose validity has been debated [184, 185]). However, due to the "curse of dimensionality", these methods cannot overcome the sampling problem when long sequences with high temporal resolution are considered.

The correlation method determines the mutual information in terms of pairwise spike correlations and hence requires only knowledge of the second order statistics of the spiking responses. These properties are well accessible, both analytically and experimentally. This renders Eqs. (1.23), (2.3) and (2.4) convenient tools for the investigation of the role of spike correlations in cases where long, stationary stimuli with finite correlation time can be assumed. In situations where short recording lengths, low spike counts or very slow signals are present, or when cross-neural correlations are considered, the methods discussed in this section are more appropriate (see e.g. [158, 159, 201] for applications on experimental data).

Chapter 3

Information transmission through mean- and variance-coding in integrate-and-fire neurons

Remark: *Some methods, results, figures and elements of the text in this chapter are also contained in this research article that Dr. Tchumatchenko and I published in 2019*[1]*:*

- Herfurth, T. & Tchumatchenko, T. (2019). Information transmission of mean and variance coding in integrate-and-fire neurons. Physical Review E, 99(3), 032420. [3]

Summarizing the introduction chapter, the fundamental units of computation in the brain are spikes, which are binary all-or-none events generated by neurons in response to input currents. Sensory or internal signals influence the dynamics of these input currents. For instance, signals can modulate the average input current (mean modulation, MM) or vary its fluctuation amplitude (variance modulation, VM). Experimental evidence suggests that both mean and variance modulation may be present in neurons. However, it is an open question how mean and variance coding compare in terms of the information that is transmitted about the underlying signals by spike trains. Regarding encoding linearity, it is unclear – generally as well as for mean and variance coding – how the

[1]Note that in this chapter, values of mutual information per frequency and mutual information per spike are indicated by subscript and superscript labels, respectively, for better distinction. This is in line with the notation of [3].

fraction of linearly encoded information depends on neuron, signal and noise parameters. These issues are considered in this chapter.

In section 3.1, I introduce the model under consideration formally and develop a linear response theory-based approximation of the mutual information for mean and variance modulation in section 3.2. As many results are obtained by means of model simulations, the respective parameters and simulation details are introduced in section 3.3. In section 3.4, I analyze statistic (coding) properties of the spike train Fourier coefficients that result from mean and variance modulated input currents, and based on that I derive the identity of the complete information and its upper bound in section 3.5. The main results regarding the comparative information-theoretic analysis of mean and variance modulation as well as leaky and exponential integrate-and-fire neurons are presented in section 3.6. The results concerning the linearity of information encoding are subject of section 3.7. An investigation of signal encoding in both mean and variance modulated currents simultaneously is given in section 3.8. Finally, I summarize and discuss all results of the present chapter in section 3.9.

3.1 Model framework

3.1.1 Formalizing mean and variance modulation

The mean and variance modulated input currents are generated as modulations of the spontaneous activity Eq. (1.31), which can be considered to stem from background activity. Now, the signal recruits additional presynaptic excitatory and inhibitory neurons and modulates their firing rate around baseline firing ν_0 such that their firing rate is $\nu_0(1 + s(t))$. As is explained in figure 3.1, dominant excitatory or balanced input from this presynaptic activity can lead to mean and variance modulated input currents, respectively. Using the diffusion approximation, the mean and variance modulated input currents can be written as (this is similar to previously used expressions [107, 114]; a detailed derivation is given in Appendix A.2):

$$I_{\mathrm{MM}}(t) = \mu(1 + s(t)) + \xi_{\sigma_n,\tau_n}(t), \tag{3.1}$$

$$I_{\mathrm{VM}}(t) = \mu + \sqrt{1 + s(t)}\xi_{\sigma_n,\tau_n}(t). \tag{3.2}$$

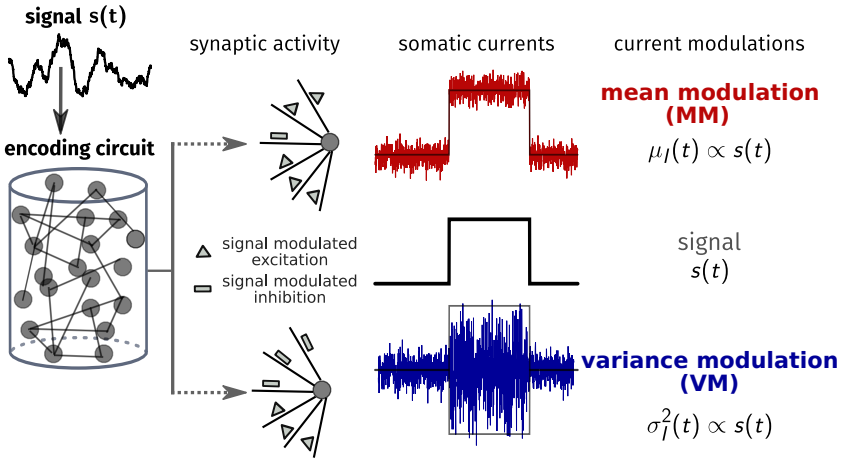


FIGURE 3.1: **Encoding signals by modulating either the mean or the variance of the somatic input currents.** An external signal $s(t)$ (step-like for illustration) arrives at the excitatory (exc) and inhibitory (inh) neurons of an encoding neural circuit. These respond with a signal-dependent firing rate $\nu_0[1+s(t)]$. I consider a neuron that receives inputs from this circuit. Depending on the number and type (exc vs inh) of its synaptic inputs, two cases can occur. In the first case (top, mean modulation), signal modulated exc currents are larger than their inh counterparts, thereby generating a net somatic current in individual neurons whose mean follows the signal $s(t)$ (red trace). In the second case (bottom, variance modulation), the exc and inh signal-modulated currents are balanced. The resulting somatic current has a signal-dependent variance (blue trace). Other signal-independent exc and inh inputs contribute to a background (not shown). Appendix A.2 provides additional details.

Here, μ is a constant input current and ξ_{σ_n,τ_n} is the noise part of the current that is inherent to spontaneous activity (see section 1.4.1). As explained in section 3.3.1, the noise variance σ_n^2 is incorporated in the model such that it also represents firing rates. Examples for the currents in both modulations are shown in figure 3.2.

In the unperturbed, stationary/spontaneous state, when no signal is present $[s(t) = 0]$, both modulation currents are equivalent with $I(t) = I_0 = \mu + \xi_{\sigma_n,\tau_n}(t)$ [Eq. (1.31)]. In this state, the current mean is μ and its variance is given by the variance of ξ_{σ_n,τ_n}. In MM, the signal modulates the current mean relative to the steady-state mean μ, which gives rise to the term $\mu\,(1+s(t))$. Equivalently, the modulation of the variance relative to its steady-state value is represented by $\sqrt{1+s(t)}\xi_{\sigma_n,\tau_n}$, whereby the square root accounts for the fact that the noise variance is modulated rather than the standard deviation[2]. Thus, the signal's standard deviation σ_s reflects how much the signal contributes to the current mean (for MM) or variance (for VM) relative to the steady state, i.e. state of spontaneous activity. It is hence called (relative) signal strength. Further it

[2]Regarding the role of noise in Eq. (3.1) and (3.2), MM corresponds to the more conventional case of *additive noise*; VM is a less investigated instance of *multiplicative noise*.

is $I_{\mathrm{VM}}(t) = \mu$ whenever $s(t) < -1$, which is equivalent to the assumption of a threshold for activation in the presynaptic encoding population.

All neurons in a population receive currents modulated by the signal $s(t)$, but have individual noise sources $\xi_{\sigma_n,\tau_n}(t)$ which are independent across trials and across neurons. I assume independent noise sources because very weak cross-correlations have been found in cortical circuits [66, 243, 284].

3.1.2 Signal properties

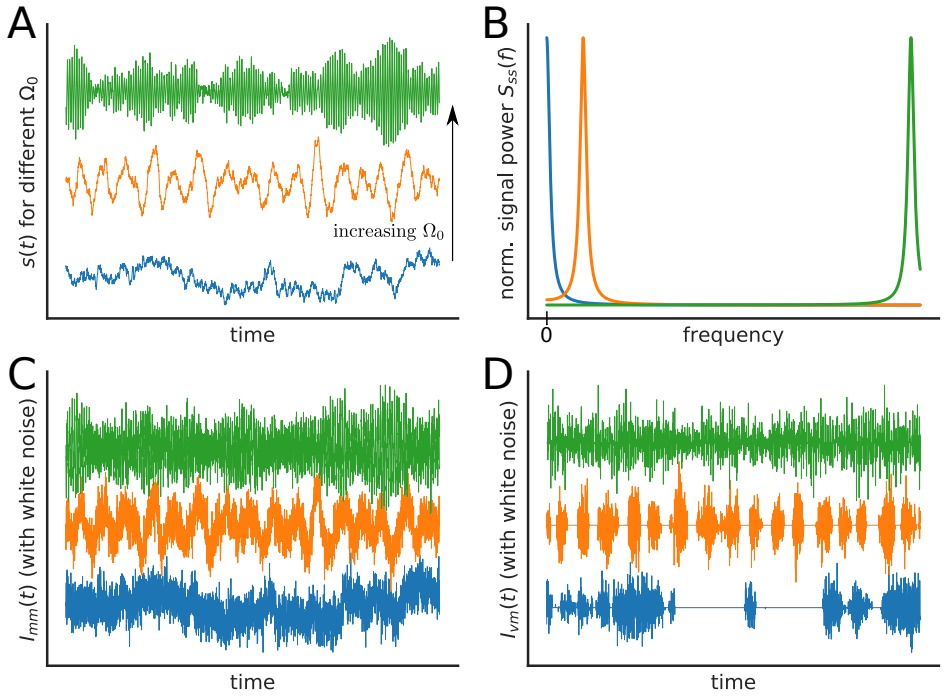


FIGURE 3.2: **Illustration of signals and respective mean- and variance modulated currents for different central frequencies Ω_0.** In all subplots the central signal frequency is ascending in the order blue, orange, green. **(A)** Example signal trajectories for different Ω_0 reveal stochastic oscillations with increasing frequency as Ω_0 increases. **(B)** The corresponding signal power spectra, given in Eq. (3.4), have a maximum around Ω_0, their width is the inverse of the signal correlation time τ_s. For $\Omega_0 = 0$, the process is an Ornstein-Uhlenbeck process. **(C),(D)** The mean and variance modulated currents obtained from the signals in (A) with white noise as "carrier" [cf. Eq. (3.1) and (3.2)].

The signals $s(t)$ are stationary, Gaussian processes that represent *stochastic oscillations*. These signal processes are characterized through their temporal autocorrelation function,

$$\langle s(t)s(t+h)\rangle = \sigma_s^2 e^{-|h|/\tau_s}\cos(\Omega_0 h)\ . \tag{3.3}$$

The parameter σ_s determines the signals standard deviation and strength (cf. previous section), and $\tau_s > 0$ is the signal correlation time and defines how quickly the autocorrelation decays, i.e. represents the "memory time" of the process. Moreover, the dominant oscillation (angular[3]) frequency in $s(t)$ is given by Ω_0, called *central signal frequency*.

The power spectrum of $s(t)$ is given by

$$S_{ss}(\omega) = \left(\frac{\sigma_s^2 \tau_s}{1 + [\tau_s(\omega + \Omega_0)]^2} + \frac{\sigma_s^2 \tau_s}{1 + [\tau_s(\omega - \Omega_0)]^2} \right) . \tag{3.4}$$

[Examples of $s(t)$ and their power spectra are shown in figure 3.2A,B.] The argmax of $S_{ss}(\omega)$ is well approximated by the central frequency Ω_0 within the range of parameters used in this work (see tables 3.2 and 3.3); the width at half-maximum corresponds to the inverse of τ_s. The described stochastic oscillations facilitate a systematic investigation of the influence of spectral signal properties on information transmission: by tuning τ_s and Ω_0 particular center values and bandwidths of the signal spectrum, i.e. signals with different "velocity", can be analyzed. Furthermore, stimuli with periodic properties and distinct frequency decomposition can be represented by the chosen signal process. Similar approaches have been used before, e.g. in [285].

It is noteworthy that the Gaussian property of $s(t)$ is not required for the exact information calculations in this work [98]. In fact, the VM current Eq. (3.2) does not generally follow a normal distribution as demonstrated in section 3.4.2.

3.2 A linear response theory for mutual information

The spike correlation functions $C_{\text{auto}}(f)$ and $C_{\text{cross}}(f)$ [Eq. (1.20) and (1.22)] that are the determinants of the mutual information are exactly computable only under simplifying assumptions. There are no analytic means to analytically calculate these functions for general signal and noise currents in integrate-and-fire neurons [9, 102, 124] (so-called "threshold neurons" are derived from LIF neurons and are analytically more tractable [109, 111]). A convenient and successful method to approximate the correlation functions is based on *linear response theory* (LRT) whose basic idea I introduced in section 1.3.1.3: the firing rate is estimated as a convolution of the stimulus and a linear kernel function [286, 287]. Here, I present a LRT approximation for the mutual

[3] In this chapter, frequencies are normally given in terms of angular frequencies $\omega = 2\pi f$.

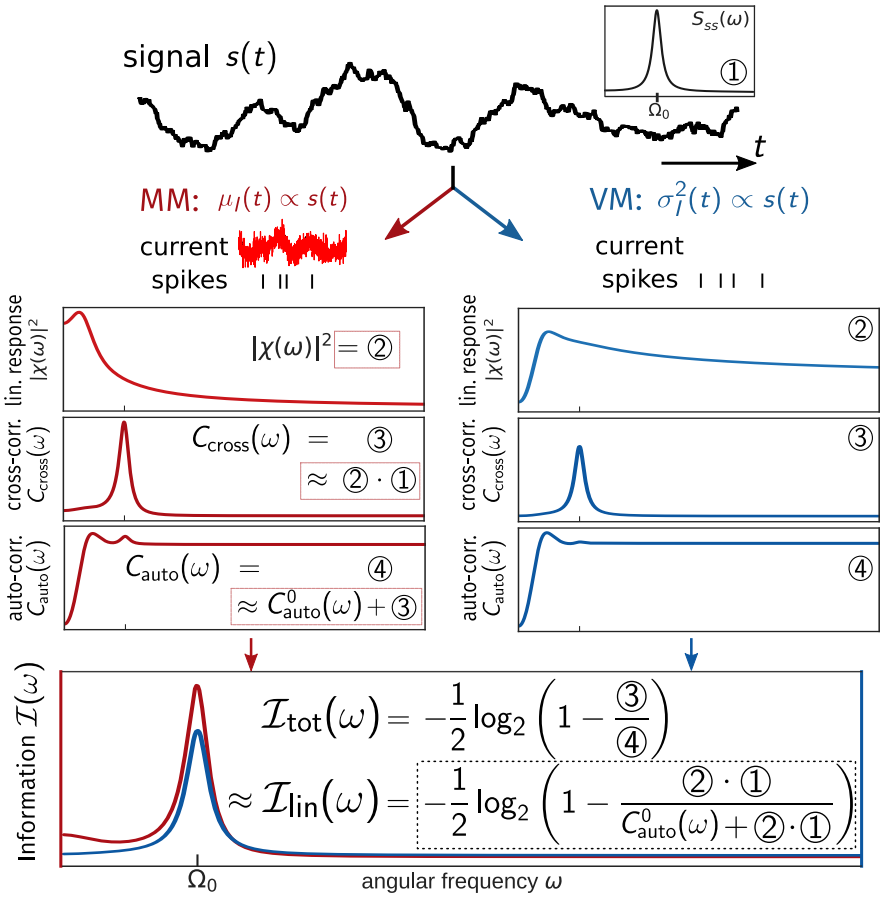


FIGURE 3.3: **Linear response theory reveals the role of response functions for information transmission.** Shown on top are the stimulus and its power spectrum (denoted by ①; cf. figure 3.2). The signal power spectrum has a central frequency Ω_0, its width is determined by τ_s (see Eq. 3.4). A given signal is encoded in either the mean (MM left, red) or the variance of the somatic current (VM right, blue). The functions ②-④ describe the responses and their correlations evoked by mean or variance coded signals, respectively. Displayed from top to bottom are linear response function ②, linear approx. to the trial averaged spike cross-correlation function ③, and the linear approx. to the spike autocorrelation function ④. The resulting information $\mathcal{I}_{\text{tot}}(\omega)$ and its linear approx. $\mathcal{I}_{\text{lin}}(\omega)$, as given by Eq. (1.23) and (3.9), are shown for MM and VM (bottom). For clarity, displayed quantities are boxed.

information $\mathcal{I}_{\text{tot}}(\omega)$ and thereby discuss how linear response functions can be related to the computation of this information. This allows to explore how signal features and neuronal dynamics that shape the linear response function determine the information content using closed-form analytical solutions. Figure 3.3 shows a corresponding illustration.

Assuming that the signal $s(t)$ is a weak perturbation to the input current, i.e. small σ_s, the dynamics of the firing rate in the Fourier domain within LRT follow from Eq. (1.11)

using the convolution theorem,

$$\tilde{r}(\omega) = \langle \tilde{\rho}(\omega) \rangle_{\mathrm{n}} = 2\pi\nu_0\delta(\omega) + \chi(\omega)\tilde{s}(\omega). \tag{3.5}$$

Here, $\chi(\omega)$ represents the linear response function, or susceptibility, and can be considered the Fourier transform of $L(h)$ in Eq. (1.11), and $\langle\rangle_{\mathrm{n}}$ stands for averaging over the noise. The response functions depend on the neuron model, the noise characteristics and, importantly, the modulation type. For LIF neurons and white noise the response functions have been derived for mean and variance modulation [56, 107] and are denoted by $\chi_{\mathrm{MM}}(\omega)$ and $\chi_{\mathrm{VM}}(\omega)$, respectively. The relevant functions for the LRT, in this case, are given in Appendix A.3. For noise with nonzero correlation time and EIF neurons approximations or semianalytic solutions of the linear response functions have been developed but are not part of this work [40, 108, 260].

Based on Eq. (3.5), the cross-correlation function of two neurons that share a weak signal $s(t)$ on top of a noise background (③ in Fig. 3.3) is given by the product of the amplitude squared linear response $|\chi(\omega)|^2$ (②) and the signal power spectrum $S_{ss}(\omega)$ (①) [220, 288]:

$$C^{\mathrm{lin}}_{\mathrm{cross}}(\omega) = |\chi(\omega)|^2 S_{ss}(\omega) \approx C_{\mathrm{cross}}(\omega). \tag{3.6}$$

The relevant regions of $\chi(\omega)$ that determine information transmission are given by the shape of the signal's power spectrum. This can be understood by considering a signal power spectrum which is constant everywhere ("white"). Then, according to Eq. (3.6), $|\chi(\omega)|^2$ will be relevant to the information content across all frequencies. On the other hand, if the power spectrum is peaked around frequency Ω_0 then only the values $|\chi(\omega \approx \Omega_0)|^2$ will be relevant for the information content (cf. figure 3.3).

A linear approximation of the spike autocorrelation function (④) – the denominator in Eq. (1.23) – is [288]

$$C^{\mathrm{lin}}_{\mathrm{auto}}(\omega) = C^{0}_{\mathrm{auto}}(\omega) + |\chi(\omega)|^2 S_{ss}(\omega) \approx C_{\mathrm{auto}}(\omega). \tag{3.7}$$

Here, $C^{0}_{\mathrm{auto}}(\omega)$ denotes the spike autocorrelation function in the absence of a signal. Note that $C^{0}_{\mathrm{auto}}(\omega)$ converges to the stationary firing rate ν_0 for high frequencies $\omega \to \infty$ (cf. Appendix A.3). Furthermore, in the given framework, the power spectrum $S_{ss}(\omega)$ decays with $1/\omega^2$ for large ω such that $C^{\mathrm{lin}}_{\mathrm{auto}}(\omega)$ also converges to ν_0 in the high frequency limit.

The influence of the signal on $C^{\text{lin}}_{\text{auto}}(\omega)$ is restricted to frequencies that are present in the signal. It is noteworthy that the approximation $C^{\text{lin}}_{\text{auto}}(\omega)$ does not strictly derive from the linear approximation Eq. (3.5) for the rate modulation and does not correspond to the analytic limit $\sigma_s \to 0$ [288].

Inserting Eq. (3.6) and (3.7) in Eq. (1.23) yields the fully linearized estimate [4] $\mathcal{I}_{\text{lin}}(\omega)$ of the mutual information:

$$\mathcal{I}_{\text{lin}}(\omega) = -\frac{1}{2}\log_2\left(1 - \frac{|\chi(\omega)|^2 S_{ss}(\omega)}{C^0_{\text{auto}}(\omega) + |\chi(\omega)|^2 S_{ss}(\omega)}\right) \approx \mathcal{I}_{\text{tot}}(\omega), \tag{3.8}$$

$$\mathcal{I}^{\text{lin}} = \frac{1}{2\pi\nu_0}\int_0^\infty \mathrm{d}\omega\, \mathcal{I}_{\text{lin}}(\omega). \tag{3.9}$$

[A factor of $1/(2\pi)$ arises because integration occurs over angular frequencies [25].] In cases where the signal strength is small one can assume $C_{\text{cross}}(\omega) \ll C_{\text{auto}}(\omega)$, which supports an expansion of Eq. (3.8) in $\frac{C^{\text{lin}}_{\text{cross}}(\omega)}{C^0_{\text{auto}}(\omega)}$ to get a better intuition for the role of $\chi(\omega)$. Then, it holds in linear order of this ratio:

$$\mathcal{I}_{\text{lin}}(\omega) \approx \frac{C^{\text{lin}}_{\text{cross}}(\omega)}{2C^0_{\text{auto}}(\omega)\log(2)}. \tag{3.10}$$

This expression highlights the importance of the linear response and cross-correlation function as a proxy for information transmission. In the linear regime ($\sigma_s \ll 1$), one can assume that $C^{\text{lin}}_{\text{cross}}(\omega)$ is a good approximation to the cross-correlation function and basic properties of the mutual information can be derived from the properties of known linear response functions [36, 56, 107, 108, 260]. However, outside the linear regime, nonlinear coding phenomena emerge and cannot be explained using linear responses alone (cf. section 3.7).

3.2.1 Difference to linearly decodable information

Both information estimates $\mathcal{I}^{\text{lin}}$ and $\mathcal{I}^{\text{ld}}$ [Eqs. (1.24) and (3.9)] rely on linearity assumptions. However, these assumptions are different and the information measures have distinct properties that need to be appreciated. First, $\mathcal{I}^{\text{lin}}$ relies on linear approximations of both cross- and autocorrelations whereas $\mathcal{I}^{\text{ld}}$ only contains another approximation of the cross-correlation. Generally, both approximations of the cross-correlation become

[4] "Linear" refers to the linear approximation Eq. (3.5) that underlies the information estimate. The approximations of the correlation functions are quadratic in χ and s.

equivalent in leading order when $\sigma_s \to 0$. In this case, both information approximations converge because the correction to the autocorrelation in $\mathcal{I}^{\text{lin}}$ becomes negligible (see section 3.7 for more details). Second, the linearly decodable information $\mathcal{I}^{\text{ld}}$ constitutes an exact lower bound for the information. $\mathcal{I}^{\text{lin}}$, on the other hand, may generally either under- or overestimate the exact mutual information. Moreover, $\mathcal{I}^{\text{lin}}$ also contains an estimate of the firing rate, which is incorporated as a normalization factor for the total information and hence is critical to be approximated correctly. Whereas deviations of $\mathcal{I}^{\text{lin}}$ from the exact information $\mathcal{I}^{\text{tot}}$ may be hard to interpret, deviations of $\mathcal{I}^{\text{ld}}$ and $\mathcal{I}^{\text{tot}}$ are unequivocal manifestations of nonlinearly encoded information (see section 3.7). The LRT-based information approximation $\mathcal{I}^{\text{lin}}$, however, can be used to analytically estimate the parameter dependencies of the neural information content (see figures 3.3-3.18) as schematically demonstrated in figure 3.3.

3.3 Model parameters and numerical simulations

The results for $\mathcal{I}^{\text{tot}}$ and $\mathcal{I}^{\text{ld}}$ and their corresponding correlation functions are obtained by numerical simulations because no closed-form expressions for these quantities exist in the given model framework. The simulations were executed using the programming language `Python`. As the simulations require to assign numerical values to all existing parameters, I here elaborate on the choice of parameters in this work. Moreover, I provide details on how I generated the signal and noise traces, produced the spike trains, and obtained the correlation functions in the simulations.

3.3.1 Choice of parameters

3.3.1.1 Neuron parameters

The neuron model parameters were chosen based on experimental findings for cortical pyramidal neurons. Generally, even within this class of neurons the electrophysiological properties vary considerably (see e.g. in the neurophysiology database *NeuroElectro* [289]). Here, the following parameters are used in accordance with *in vivo* and *in vitro* studies [38, 83, 290]:

$$\begin{aligned} \tau_m &= 10 \text{ ms}, \\ R &= 40 \text{ M}\Omega, \\ V_r &= 0 \text{ mV}, \\ \Theta &= 15 \text{ mV}. \end{aligned}$$

For the LIF model [Eq. (1.4)], refractory periods are considered to be negligible, which is commonly supported by the rather low firing rates (see next paragraph) [124, 260, 261]. For the EIF model [Eq. (1.6)], the refractory period and spike initiation time are based on fits of experimental recordings [38, 102],

$$\begin{aligned} \tau_r &= 0 \text{ ms (LIF)}, \\ \tau_r &= 5 \text{ ms (EIF)}, \\ \Delta_T &= 1.5 \text{ ms}. \end{aligned}$$

3.3.1.2 Noise and signal parameters

The constant input μ was restricted to lie in the *subthreshold* regime where $\mu \cdot R < \Theta$. The fluctuation-driven subthreshold regime has been linked to irregular cortical activity and balanced states [43, 44, 67, 291–293] (see also section 1.2.2). The focus on the fluctuation-driven regime is also in agreement with the settings and stimulation protocols in experimental studies [115, 293–295].

The noise parameters, i.e. τ_n and σ_n, and constant current μ (cf. sections 1.4.1 and 3.1.1) were chosen to produce cortical firing statistics: I aimed at firing rates between 5 and 30 Hz and at values of the coefficient of variation around 0.7 [44, 50, 83, 296, 297]. This was done by setting three different target firing rates and a target range of coefficients of variation and then tuning the parameters accordingly. In particular, I adjusted the noise strengths σ_n and μ for each noise correlation time τ_n and each neuron model such that the three different firing rates were realized by three different values of σ_n at fixed μ. The reference for the parameter tuning was the stationary state [$s(t) = 0$], and confounding through parameter adjustments to different signal parameters was avoided.

Following this procedure, I obtained for each τ_n three values of σ_n that correspond to the firing rates ν_1, ν_2, ν_3 and that I denote by $\hat{\sigma}_n^{(1)}, \hat{\sigma}_n^{(2)}, \hat{\sigma}_n^{(3)}$, respectively. These firing rates measured in the steady state are $\nu_1 = (11 \pm 1.1)$ Hz, $\nu_2 = (16.8 \pm 0.6)$ Hz, $\nu_3 = (21 \pm 0.4)$ Hz for the LIF model and $\nu_1 = (10 \pm 1.2)$ Hz, $\nu_2 = (16.3 \pm 0.8)$ Hz, $\nu_3 = (20.7 \pm 0.8)$ Hz

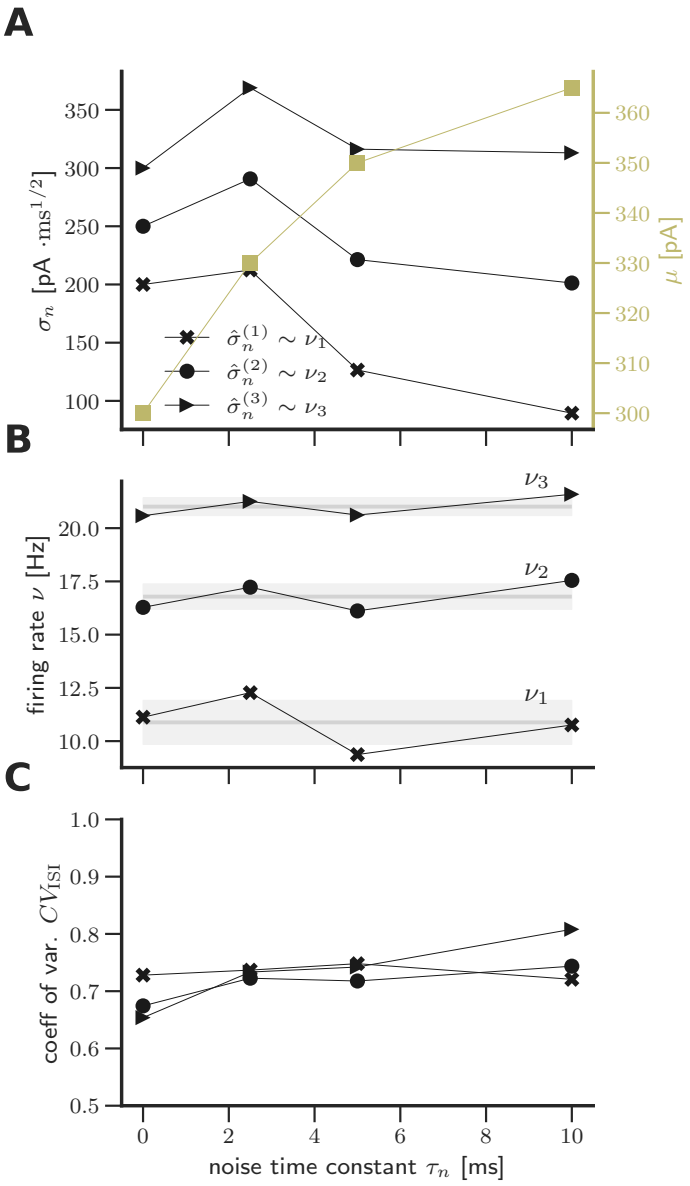


FIGURE 3.4: **Noise strength and constant input current for the LIF model were chosen to yield spiking statistics that agree with experimental findings.** The shown spiking statistics were obtained from simulations in the signal-free case [$\sigma_s = 0$]. **(A)** For each noise correlation time τ_n one value of μ (olive squares, right axis) and three equidistant values of σ_n (different symbols, left axis) were chosen and resulted in the spiking statistics shown in (B) and (C). As follows from **(B)**, for each τ_n one value of σ_n can be assigned to one of three firing levels. These are denoted by $\sigma_n^{(1)}$, $\sigma_n^{(2)}$, $\sigma_n^{(3)}$ and correspond to the firing rates ν_1, ν_2, ν_3 as explained in the text. **(C)** confirms that all chosen parameters yield values of CV_{ISI} in between 0.6 and 0.8.

for the EIF model. The values of the coefficient of variation are between 0.6 and 0.9. A graphical representation of these relations and the obtained firing statistics is shown in figures 3.4 and 3.5 for LIF and EIF neurons, respectively. Moreover, the chosen values of $\hat{\sigma}_n$ for each neuron model and noise correlation time are summarized in table 3.1.

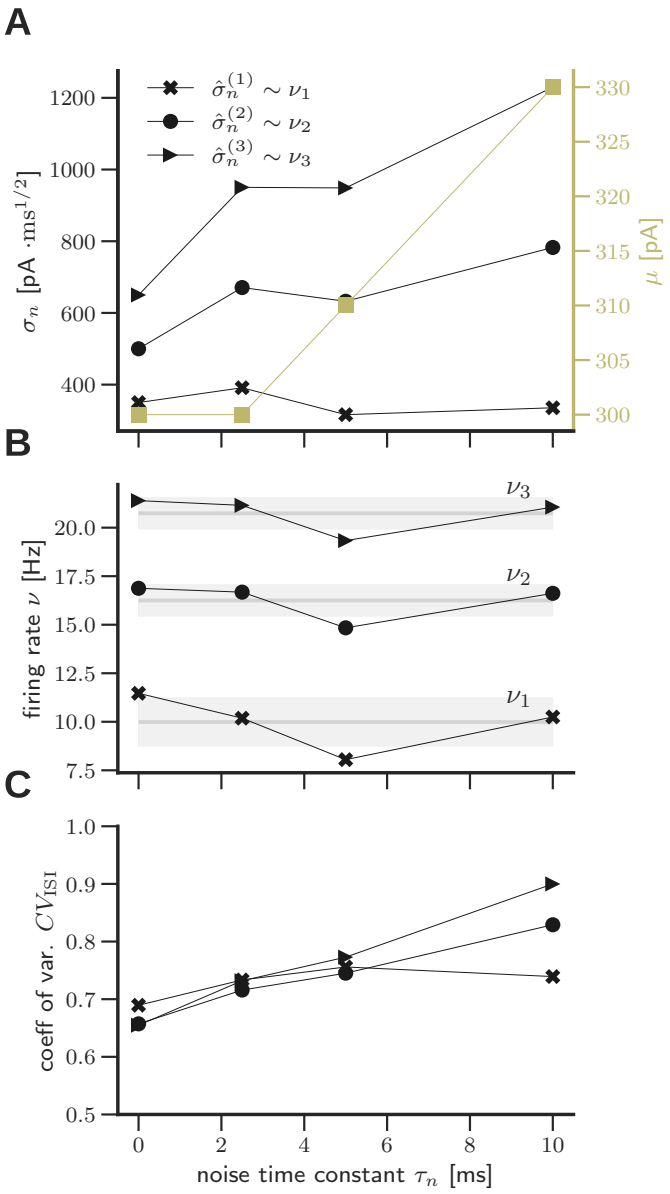


FIGURE 3.5: **Noise strength and constant input current for the EIF model were chosen to yield spiking statistics that agree with experimental findings.** Equivalent to figure 3.4 but for exponential integrate-and-fire neurons.

The values for the firing rates are well aligned along horizontal lines and show only mild deviations from the target firing rates (figures 3.4B and 3.5B). Importantly, these deviations are considerably smaller than the intervals between the different firing levels, such that findings regarding different levels of firing (e.g. when σ_n is the first axis) are well founded and refer to significant differences in the firing rate between neighboring points. This enables an evaluation of the role of different firing rates/noise strengths on information transmission. Similar arguments apply to the coefficient of variation CV_{ISI} (figures 3.4C and 3.5C). Again, only small variations exist, and all values lie well within

the previously identified biological, cortical regime [50, 297]. Nonetheless, the tuning of CV_{ISI} for different τ_n exhibits stronger deviations and a weak trend for EIF neurons because the fine tuning of the target firing statistics is highly non-trivial: Generally, for nonwhite noise the parameters μ and σ_n have to be determined through simulations and depend on the target rate and CV in a complex, non-monotonic way [220]. The parameter tuning is additionally constrained to fixed values of μ for a given τ_n at all firing rates to reduce the number of independent parameters and confounding factors. Moreover, the values of σ_n are restricted to be equidistant at a given τ_n for a better interpretability of different noise values.

The signal parameters Ω_0, σ_s and τ_s introduced in section 3.1.2 are independent variables and were chosen to yield a broad overview of their role for information transmission. The central signal frequency Ω_0 was sampled in a range from 0 to $\sim 2\pi\times$2.5kHz, covering the spectrum of relevant frequencies in neural filtering. A value of $\Omega_0 = 0$ corresponds to the popular choice of Ornstein-Uhlenbeck process signals, and the high-frequency end of the range lies above the observed cut-off for neural filtering [114, 115, 298]. The signal and noise time constants, τ_s and τ_n, are chosen according to experimental reports of AMPA and GABA currents [280, 299], and are in the range of 10–30 ms and 2.5–10 ms, respectively (plus the white noise case). The signal strength was chosen to span from very weak signals [$\sigma_s = 0.05$] to overmodulation [$\sigma_s = 2$], thereby covering (allegedly) linear and very nonlinear regimes. This range can be considered to represent the spectrum from inconspicuous to salient stimuli. I summarize all parameters sampled in the simulations in table 3.2.

	LIF		EIF	
τ_n (ms)	$\left[\hat{\sigma}_n^{(1)}, \hat{\sigma}_n^{(2)}, \hat{\sigma}_n^{(3)}\right]$ (pA$\sqrt{\text{ms}}$)	μ (pA)	$\left[\hat{\sigma}_n^{(1)}, \hat{\sigma}_n^{(2)}, \hat{\sigma}_n^{(3)}\right]$ (pA$\sqrt{\text{ms}}$)	μ (pA)
0	[200, 250, 300]	300	[350, 500, 650]	300
2.5	[95, 130, 165]$\cdot\sqrt{5}$	330	[175, 300, 425]$\cdot\sqrt{5}$	300
5	[40, 70, 100]$\cdot\sqrt{10}$	350	[100, 200, 300]$\cdot\sqrt{10}$	310
10	[20, 45, 70]$\cdot\sqrt{20}$	365	[75, 175, 275]$\cdot\sqrt{20}$	330

TABLE 3.1: **Values of $\boldsymbol{\sigma_n}$ corresponding to $\hat{\boldsymbol{\sigma}}_{\boldsymbol{n}} = \left[\hat{\boldsymbol{\sigma}}_{\boldsymbol{n}}^{(\mathbf{1})}, \hat{\boldsymbol{\sigma}}_{\boldsymbol{n}}^{(\mathbf{2})}, \hat{\boldsymbol{\sigma}}_{\boldsymbol{n}}^{(\mathbf{3})}\right]$ and μ for the different $\boldsymbol{\tau_n}$.** Parameters $\hat{\sigma}_n$ were chosen equidistantly and in order to produce three different, fixed firing rates at given CV_{ISI} and constant input current μ. The values can also be seen in figures 3.4, 3.5.

Parameter	(unit)	Simulated values
σ_s	[1]	[0.05, 0.15, 0.25, 0.5, 0.75, 1, 2]
Ω_0	(2πkHz)	[0, 0.25, 0.51, 1 , 2.54, 7.11*]
$\hat{\sigma}_n$	[1]	$[\hat{\sigma}_n^{(1)}, \hat{\sigma}_n^{(2)}, \hat{\sigma}_n^{(3)}]$
τ_n	(ms)	[0, 2.5, 5, 10]
τ_s	(ms)	[10, 20, 30]

TABLE 3.2: **Signal and noise parameters used in simulations.** The column *Simulated values* lists all values that were sampled in the simulations for the respective parameter. The parameters that were used to produce the figures are annotated in their respective captions. Varied parameters used in inset figures are annotated in the plot themselves. Each $\hat{\sigma}_n^{(m)}$ was chosen to realize a fixed firing rate ν_m as explained above and shown in table 3.1. *This value is not included for the EIF model.

3.3.2 Simulations

Here, I summarize the procedure that I used to simulate the spike trains and obtain the correlation functions and mutual information. The `Python` code for the simulations in this chapter can be found on this[5] github repository.

In each simulation, the signal and noise currents were generated first. For the current generation, time was discretized in bins of width $\Delta t = 0.02$ ms. White noise was then obtained from drawing independent numbers from the normal distribution $\mathcal{N}(0, 1/\sqrt{\Delta t})$. To generate the traces of signal and noise with $\tau_n > 0$, I used the algorithm from Ref. [300]. After obtaining the signal modulated currents through Eq. (3.1) and (3.2), I next iteratively solved Eq. (1.4) and (1.6) via the *Euler method* [301] with time discretization width Δt. The resulting spike times were discretized at 0.1 ms (sampling frequency 10 kHz).

In the following I describe how the correlation functions and mutual information values for sections 3.6, 3.7 and 3.8 were obtained (the procedure underlying the simulation results in section 3.4 is explained in the section itself.) To obtain the spike autocorrelation function, I generated 256,000 independent spike trains of 4 seconds length. The autocorrelation function was calculated using Eq. (1.20) by using *Fast Fourier Transform* [301] and averaging over all trials. To calculate the cross-correlation function, I generated 4,000 trials for each of 64 different stimuli $s(t)$ with varying, independent noise trajectories (256,000 trials in total; each trial 4 seconds). I evaluated all pairs $|\tilde{\rho}_i^k(\omega) \cdot \tilde{\rho}_j^k(\omega)|^2$ for the same stimulus trajectory $s_k(t)$ and averaged across trials to obtain $C_{\text{cross}}^k(\omega)$ (cf. figure 1.16). I repeated this procedure for all 64 independent stimuli and obtained the

[5] https://github.com/t8ch/dissertation-code/tree/master/Ch3

cross-correlation function as $\frac{1}{64}\sum_{k=1}^{64} C^k_{\text{cross}}(\omega)$. All correlation functions were smoothed with a Hanning window function with a width of 5 Hz [301]. To calculate the linearly decodable information $\mathcal{I}^{\text{ld}}$, the same procedure was used to determine the signal-response cross-spectrum $S_{sr}(\omega)$ by first evaluating the Fourier transforms of $s(t)$ and $\rho(t)$. I used cut-offs of $\omega_{max} = 3\Omega_0 + 3 \times 2\pi\text{kHz}$ and $\omega_{max} = 8 \times 2\pi\text{kHz}$ in the integration over frequencies to obtain $\mathcal{I}^{\text{tot}}$ and $\mathcal{I}^{\text{ld}}$ [Eq. (1.24) and (1.23)], respectively. Note that for long spike trains ($T \gg \tau_s, \tau_n, \tau_m$), averaging the information content over stimuli and to obtain the stimulus averaged cross-correlation function first and to compute the mutual information from that subsequently are equivalent. Here, I chose the latter.

3.4 Analysis of spike train Fourier coefficients

The correlation method for computing the full mutual information [Eq. (1.23)] relies on a theorem according to which stationary signals with finite correlation times lead to spike trains whose Fourier coefficients are multivariate normal distributed (Gaussian) and independent across trials [25, 98, 302]. Numerically, the validity of this theorem has only partly been demonstrated for mean modulating signals [98], and not at all for variance modulation.

In this section, I analyze the statistical properties of the spike train Fourier coefficients for both MM and VM in LIF neurons. In section 3.4.1, I demonstrate that the spike train Fourier coefficients are independent and normal distributed. This is the case irrespective whether the input current Fourier modes have this property (section 3.4.2). Moreover, in section 3.4.3 I show that the mean values of the Fourier coefficients are the information encoding features and thereby verify a fundamental assumption that I use in section 3.5 to derive the (putative) upper bound for information transmission.

3.4.1 Fourier coefficients are Gaussian and independent

The correlation method is based on the assumption that the distribution of the Fourier coefficients of the spike trains are Gaussian and independent if the eliciting currents are stationary and have finite-time correlations. Here, I explicitly confirm this assumption for both MM and VM by first showing in section 3.4.1.1 that the spike train Fourier coefficients are uncorrelated and each follow a normal distribution (a similar analysis

was done in [98]). Then, in section 3.4.1.2, I demonstrate that the spike train Fourier coefficients (FC) stemming from mean and variance modulated input currents are multivariate normal. Together this implies that the FC are independent and normal [303] as demanded for the correlation method [98]. Again, the corresponding `Python` code can be found on this github repository.

3.4.1.1 Fourier coefficients are uncorrelated and Gaussian

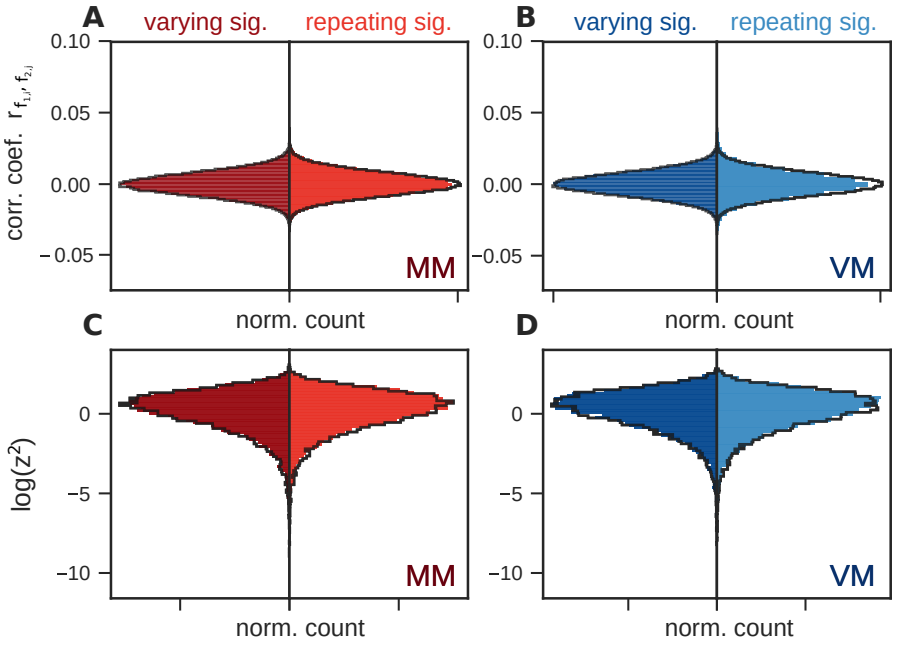


FIGURE 3.6: **Fourier coefficients of spike trains resulting from mean and variance modulated input currents in leaky integrate-and-fire neurons follow Gaussian distributions and are uncorrelated within and across frequencies.** **(A)** For mean modulated input currents, the distributions of cross-trial correlation coefficients $r_{f_{1,i},f_{2,j}}$ between spike train Fourier coefficients at frequencies f_1 and f_2 obtained at trials i and j are displayed. The left side of each panel denotes the distribution for trials with varying signals. The right side shows the distribution for trials with repeated signals. The black envelope line represents the corresponding histogram for surrogate data generated from independent Gaussian distributions with the same mean and variance for each frequency. The values of $r_{f_{1,i},f_{2,j}}$ are small for both varying and repeating signals, and their distribution matches with that resulting from independent Gaussian variables (black line). It is known that the width of this distribution is determined by the finite recording size and decays with increasing recording time [98]. **(B)** Same statistics as in (A) for variance modulated input currents. **(C),(D)** The logarithms of the squared Gaussianity test-scores z (see text for details) are displayed for varying/repeated signals (left/right) for mean ((C), red) and variance ((D), blue) modulated input currents. Again, the histograms obtained from the sampled and the surrogate normal distributions (black lines) coincide, indicating that the Fourier coefficients of the spike trains are Gaussian. The parameters I used are as follows $\sigma_s = 0.25$, $\Omega_0 = 0.5 \times 2\pi\text{kHz}$, $\sigma_n = 250\ \text{pA}/\sqrt{\text{ms}}$, $\tau_n = 0$, $\tau_s = 20$ ms.

The spike train Fourier coefficients are given by the real and imaginary parts of $\tilde{\rho}(f)$ of the Fourier transformed spike trains (see section 1.3.2.2). The conducted test consists of evaluating the correlation between the Fourier coefficients and a standard test score for normality. Therein, the correlation between two Fourier coefficients X and Y is measured by the Pearson correlation coefficient $\mathrm{r}_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$ [30]. Figure 3.6A,B visualizes an example of this analysis.

To further assess the normality of the Fourier coefficients, I used a standard Gaussianity test [304][6], yielding a test score z for each Fourier coefficient. This score is obtained as a combined test of skewness and kurtosis[7] of the distribution of each Fourier coefficient across trials in comparison to the skewness and kurtosis of a normal distribution. For each Fourier coefficient the z-scores z_s and z_k for skewness and kurtosis are calculated from the skewness/kurtosis x_s/x_k of the samples as $z_{s/k} = (x_{s/k} - \mu_{s/k})/\sigma_{s/k}$. The values $\mu_{s/k}$ and $\sigma_{s/k}$ are the known mean and standard deviation of the skewness/kurtosis distribution under the null hypotheses that the sample is drawn from a normal distribution with the sample mean and variance. This means if $|z_s| > 1.96$ the sample's skewness is significantly different from zero (significance level $\alpha = 0.05$). The total test score is $z^2 = z_s^2 + z_k^2$. Accordingly, samples with lower values of z^2 have a higher probability of originating from a normal distribution. Test statistics are graphically represented in figure 3.6C,D.

For both fixed and varying signal, I computed the distribution of pairwise correlations between Fourier coefficients using 300 randomly chosen frequencies ($\omega \in (0, 2.5] \times 2\pi\mathrm{kHz}$) and the resulting 90,000 frequency pairs, whereby for each frequency real and imaginary parts of the Fourier coefficients were considered separately. I used the same data for the normality test. As a reference, for each Fourier coefficient, samples from a normal distribution with the mean and variance as obtained from the original data and the same size were generated (independently across frequencies). The correlation coefficients and normality score were computed equivalently on this surrogate data and are shown as black envelope curves in figure 3.6. All test scores of the simulated spikes trains have the same distribution as those obtained by sampling from "true" normal distributions. The analysis thus demonstrates that uncorrelated Gaussianity of the Fourier coefficients holds for trials with varying and repeated signals.

[6] Implemented in `scipy.stats.normtest` in the Python library scipy.
[7] Third and fourth moment statistics representing symmetry and "tailedness" of a distribution [305].

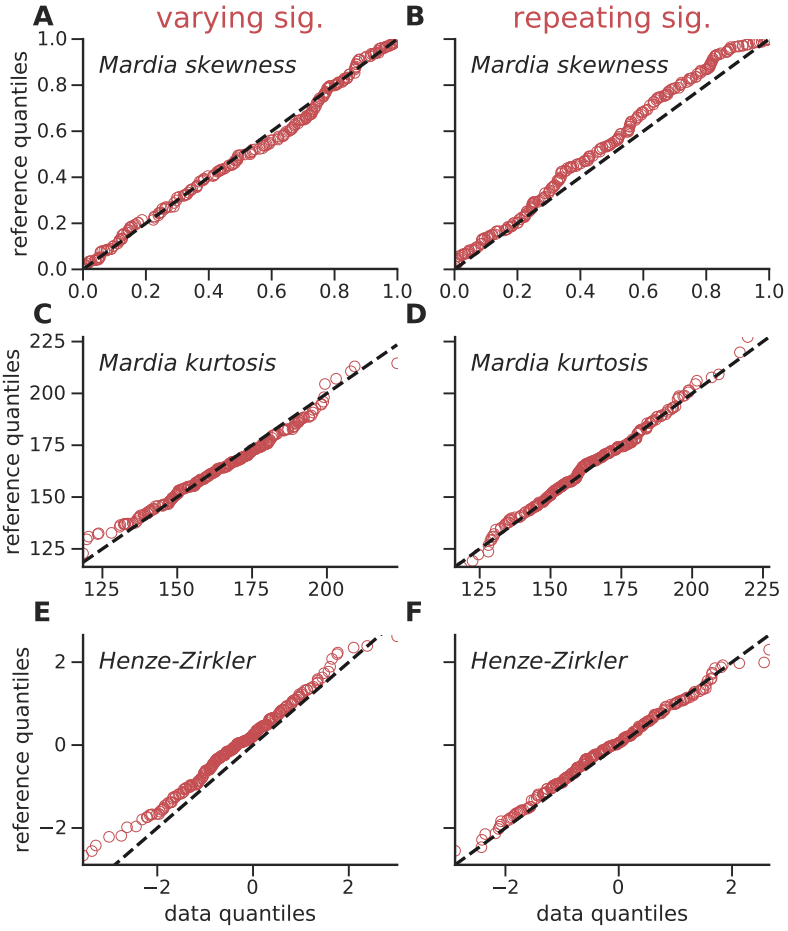


FIGURE 3.7: **Multivariate normality tests of the Fourier coefficients at nine frequencies around Ω_0 (mean modulation).** The plots show the QQ-plots of the tests statistics of Mardia's skewness test (top row plots **(A)** and **(B)**), Mardia's kurtosis test (center, **(C)** and **(D)**) and Henze-Zirkler test (bottom, **(E)** and **(F)**). Thereby, the x-axes represent quantiles from the distribution that was obtained from simulated spike trains, the y-axes correspond to independent Gaussian variables with the same mean and variance. **(A)**, **(C)**, **(E)**: tests for varying signals; **(B)**, **(D)**, **(F)**: tests for repeating (fixed) signal. The quantiles are closely aligned along the identity line which indicates multivariate normality and independence of the simulated Fourier coefficients. All tests were carried out using LIF neurons and only the statistics of the real parts of the Fourier coefficients are shown here. Used parameters: $\sigma_s = 0.25$, $\Omega_0 = 0.5 \times 2\pi$kHz, $\sigma_n = 250$ pA$/\sqrt{\text{ms}}$, $\tau_n = 0$, $\tau_s = 20$ ms, equivalent to figure 3.6.

3.4.1.2 Spike train Fourier coefficients are multivariate normal

I use three well-known test statistics for multivariate normality (MVN). The first two are Mardia's tests for multivariate skewness and kurtosis [306, 307] (multivariate extension of the standard normality test from the previous section), which belong to the most referenced tests for multivariate normality [308–310]. In the limit of an infinite number of samples the two scores of Mardia's test are asymptotically distributed as χ^2 with the

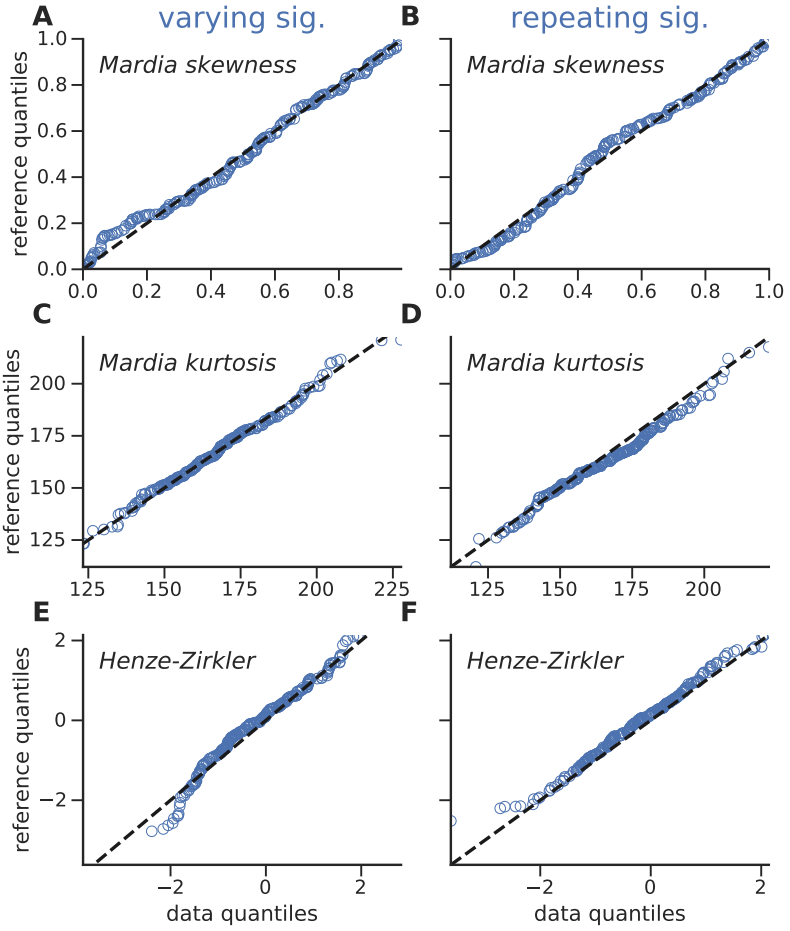


FIGURE 3.8: **Multivariate normality tests of the Fourier coefficients at nine frequencies around Ω_0 (variance modulation).** Same as figure 3.7 but for variance modulation. The quantiles are closely aligned along the identity line which indicates multivariate normality and independence of the simulated Fourier coefficients.

degree of freedom dependent on the dimensionality of the test data and standard normal, respectively [308, 311]. The third test is the Henze-Zirkler test whose score is based on the characteristic function of a multivariate distribution [308, 312]. The Henze-Zirkler (HZ) test score is approximately lognormal distributed in the aforementioned limit of very large samples [312]. The HZ test has been described as very powerful in the literature [308, 311].

In principle, it is possible to calculate the p-values that correspond to each of the three test scores obtained from a given number of data samples. These could then be used to evaluate the null hypothesis "the data has a multivariate normal density". However, I decided to follow a more informative approach for two reasons: First, the p-values given in the literature are based on the limiting distributions that arise for infinitely

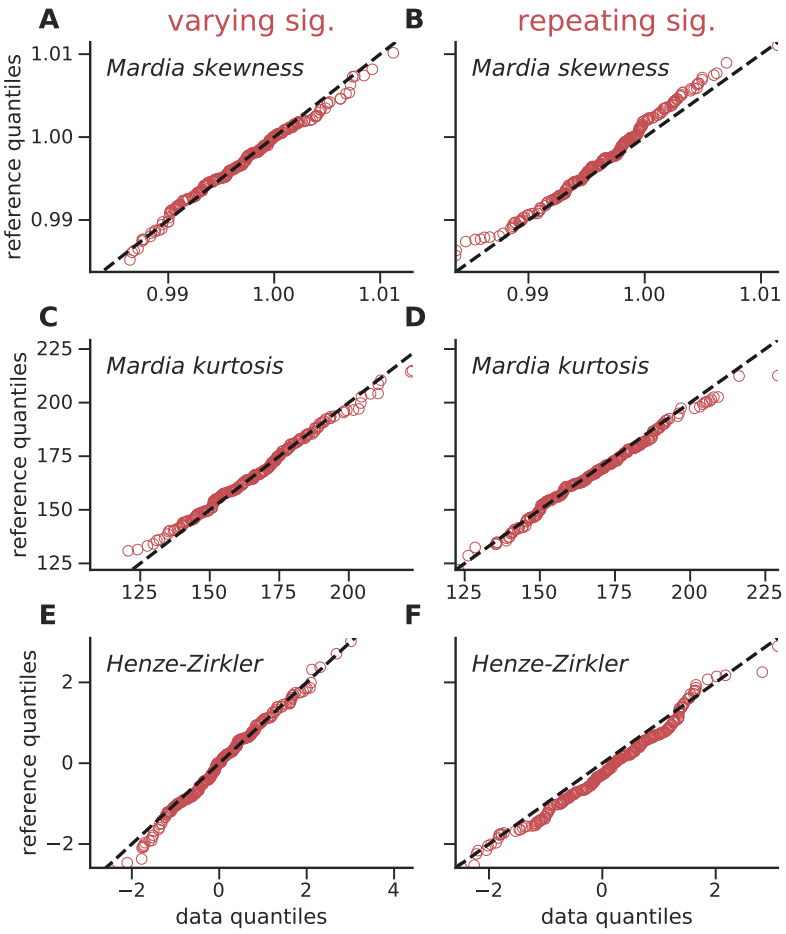


FIGURE 3.9: **Multivariate normality tests of the Fourier coefficients at nine random frequencies (mean modulation).** Same as figure 3.7 but for nine random frequencies. The quantiles are closely aligned along the identity line which indicates multivariate normality and independence of the simulated Fourier coefficients.

large sample sizes. In practice, however, this asymptotic convergence only occurs for increasingly large sample numbers when the dimensionality of the data (i.e. the number of considered Fourier coefficients in my case) is increased. I found that already for 9-dimensional data, I cannot expect convergence for a practicable number of samples. Second, as a consequence, an obtained p-value merely would describe the probability by which I would have found the respective test score under the null hypothesis. This means that even if the null hypothesis is true one may discard it by chance with a probability equal to the chosen confidence (e.g. 5%).

Therefore, I assessed multivariate normality with a procedure that rests on repeated, independent evaluations of the three test scores, as explained in the following. For each test case – i.e., a combination of MM or VM and repeated signals or varying signals –

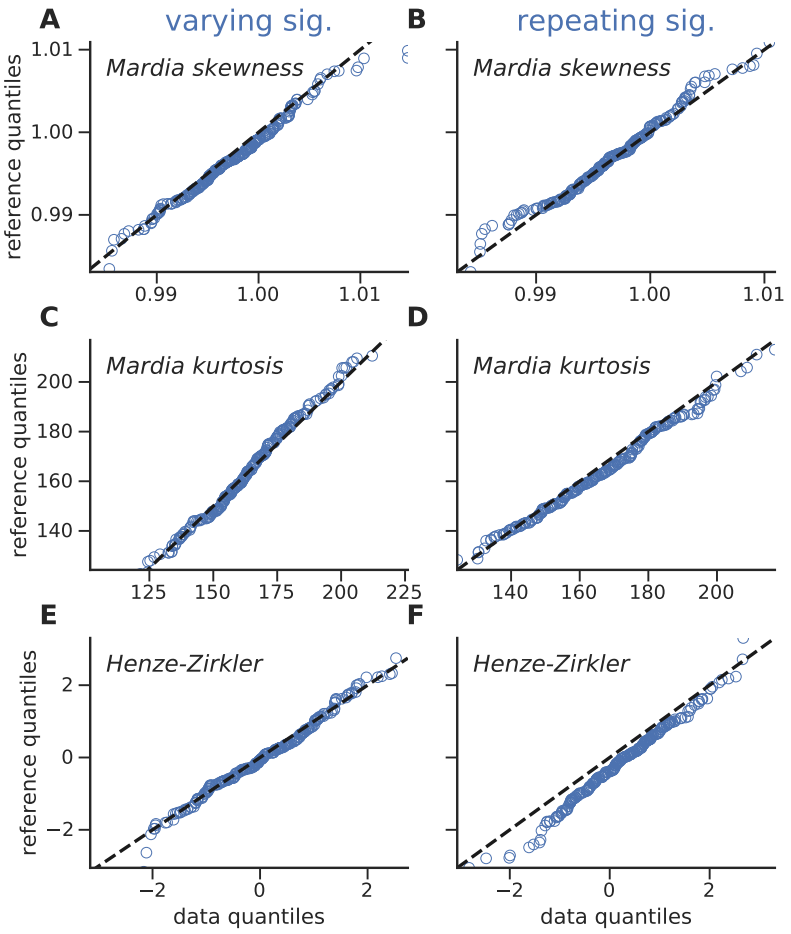


FIGURE 3.10: **Multivariate normality tests of the Fourier coefficients at nine random frequencies (variance modulation).** Same as figure 3.9 but for variance modulation. The quantiles are closely aligned along the identity line which indicates multivariate normality and independence of the simulated Fourier coefficients.

I chose nine frequencies (see below) whose corresponding Fourier coefficients I tested for multivariate normality. In each case, I simulated 16,000 spike trains, extracted the chosen Fourier coefficients (real or imaginary part depending on the case) and calculated the three test scores[8]. Additionally, I generated for each Fourier coefficient 16,000 values drawn from uncorrelated Gaussian distributions with the same mean and variance as in the simulated data. The three test scores were also computed for this reference data (equivalent to the surrogate data in section 3.4.1). This routine was repeated 250 times with independent simulations such that I obtained 250 values of each score for both simulated data and reference data for each test case. Under the assumption of multivariate normality, one would expect that the scores of the simulated data and those of the reference data follow the same distribution (cf. [308, 311, 314]). This is

[8]I implemented Mardia's test and adapted the Henze-Zirkler test from [313] (cf. online code resource).

strongly supported by my results in figures 3.7-3.10, which show quantile-quantile plots (QQ-plots)[9] of these distributions for the different tests and in different test cases.

The figures show the analysis of two different sets of frequencies. The first set of frequencies (figures 3.7 and 3.8) are aligned around $\Omega_0 = 0.5 \times 2\pi$kHz (nine frequencies; Ω_0 and the four adjacent smaller and four adjacent larger frequencies at spacing $\Delta_\omega = 2\pi/(20\text{ms})$ between two frequencies). The second set (figures 3.9 and 3.10) are nine randomly chosen frequencies within the range $[0, \sim 6\Omega_0]$. I restricted the analysis to nine frequencies because the amount of samples that would be required for more frequencies to obtain appropriate statistics grows exponentially. However, the fact that multivariate normality holds for both random frequencies and neighboring frequencies in the region of highest information transmission strongly supports the multivariate normality of the spike train Fourier coefficients.

Together with the results of section 3.4.1, I conclude that the spike train Fourier coefficients are Gaussian and independent as theoretically derived previously [302]. This is consistent with standard methods to compute the mutual information in spike trains where it is assumed that the FC at different frequencies carry information independently [25, 218].

3.4.2 Normal distribution of the input current is not required

Importantly, Gaussianity and independence of the Fourier coefficients emerge under the stationarity assumptions even if the input current does not itself follow a Gaussian distribution. This can be exemplified by considering the stationary statistics of variance modulated currents. Expanding $I_{\text{VM}}(t)$ [Eq. (3.2)] in $s(t)$, one finds

$$I_{\text{VM}}(t) = \mu + \left(1 + \frac{s(t)}{2} - \frac{s(t)^2}{8} + \mathcal{O}(s^3)\right) \xi_{\sigma_n,\tau_n}(t) \ . \tag{3.11}$$

This expression reflects that $I_{\text{VM}}(t)$ is generated by products of Gaussians as s and ξ both are Gaussian processes. Generally, the resulting time-series does not have Gaussian statistics [315]. Even if all summands in the expansion are taken into account, the central limit theorem should not apply because all terms are highly dependent [204].

[9] Here, I generated the QQ-plots by sorting the test scores in either case by size and plotting the pairs of values at the same position in the ordered list as points.

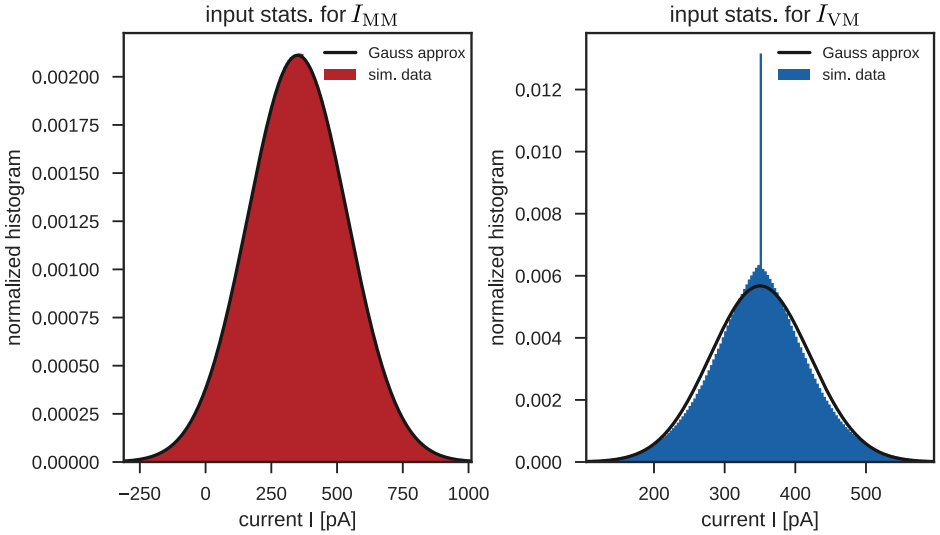


FIGURE 3.11: **Variance modulated currents are generally not normal distributed.** The figure shows the distributions of the currents $I_{\mathrm{MM}}(t)$ (red, left) and $I_{\mathrm{VM}}(t)$ (blue, right) across time as histograms for example signal/noise traces. The black curves show the Gaussian analogues with the same mean and variance. Whereas MM currents are always normal distributed, VM input currents typically have non-Gaussian histograms that approach normal distributions in the limit of small signals (cf. expansion of Eq. (3.11)). Moreover, if σ_s increases the amount of overmodulation $[s(t) > 1]$ increases. In these cases, it is $I_{\mathrm{VM}}(t) = \mu$ (see section 3.1.1) which results in the peaked maximum in the histogram as visible in the right plot. Parameters in this figure are $\sigma_s = 0.5$, $\mu = 350$ pA, $\Omega_0 = 0.5 \times 2\pi$kHz, $\tau_n = 5$ ms, $\sigma_n = 220$ pA$/\sqrt{\text{ms}}$.

Furthermore, overmodulation at high signal strengths produces non-Gaussian statistics in the VM current (see figure 3.11). $I_{\mathrm{MM}}(t)$, on the other hand, is always Gaussian.

Moreover, Eq. (3.11) reveals that the cross-correlation between $I_{\mathrm{VM},i}(t)$ and $I_{\mathrm{VM},j}(t)$ in trials i and j with the same signal vanishes,

$$\langle I_{\mathrm{VM},i}(t) I_{\mathrm{VM},j}(t+h) \rangle_{\mathrm{tr}_{i \neq j}} - \mu^2 = 0 \quad \forall\ h \neq 0, \tag{3.12}$$

because according to Isserlis' theorem, all terms resulting from expansion Eq. (3.11) can be expressed in terms of products of two-point cross-correlations between $s(t)$ and $\xi(t)$ at different trials [316]. However, all these signal-noise and noise-noise cross-correlations vanish due to independence. Nonetheless, for the spiking responses resulting from $I_{\mathrm{VM}}(t)$ one finds that the cross-correlation $C_{\mathrm{cross}}^{\mathrm{VM}}(\omega)$ is generally nonzero and that the spiking ensemble is fully described by the two-point correlation function, determining the variance of the Gaussian distributions for each Fourier coefficient. As an important finding, it can be concluded that the neurons' nonlinear spiking process maps all signal correlations onto two-point correlation functions in the spiking output, even if two-point

correlation functions are not present in the variance modulated current input. This mechanism forms a solid foundation for the applicability of the correlation method under stationary, finite-memory inputs.

3.4.3 Information is encoded in the mean of spike train Fourier coefficients

In this work, the variances of the spike train Fourier coefficients across trials are the decisive quantities because they determine the correlation functions entering the correlation method Eq. (1.23). Regarding neural encoding mechanisms, it is interesting to also inspect the characteristics of the mean values of the Fourier coefficients.

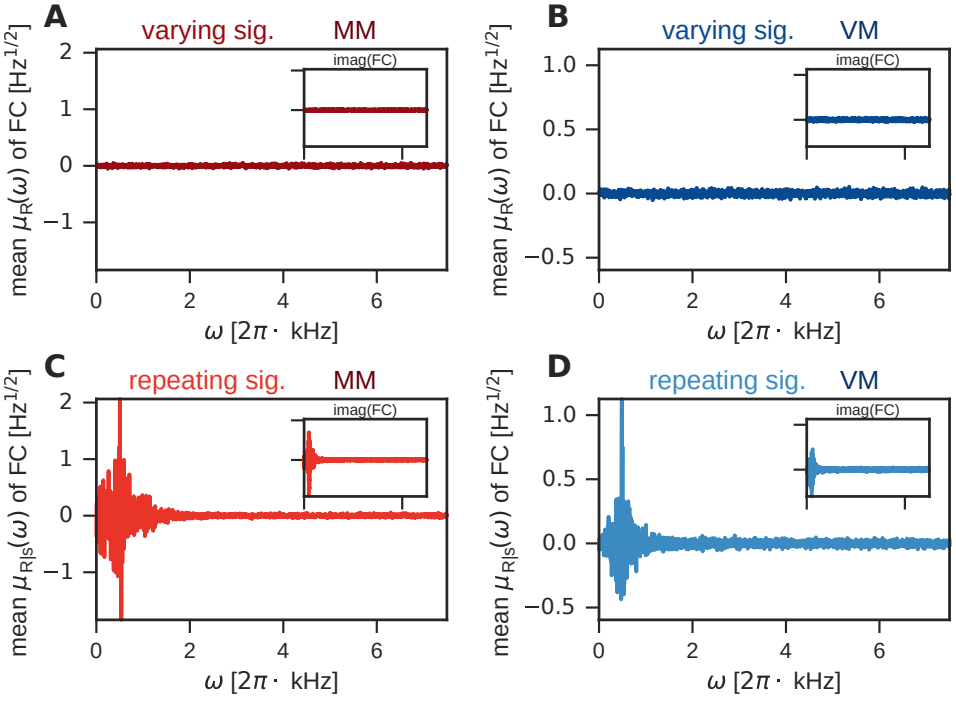


FIGURE 3.12: **Signal information is encoded in the mean values of the spike train Fourier coefficients.** All panels show the trial averaged mean values of the real (main plots) and imaginary (insets) Fourier coefficients (FC) as a function of angular frequency ω. No differences between the real and imaginary FC are observable. The Fourier coefficients are analyzed for trials with varying signals (**(A)**,**(B)**) and for repeating signals (**(C)**,**(D)**) for both mean and variance modulation (**(A)**,**(C)** and **(B)**,**(D)**, respectively). For varying signals, the mean values $\mu_R(\omega)$ average out while one finds finite values for the signal-conditioned FC means $\mu_{R|s}(\omega)$ at frequencies where the signal has major contributions (here $\Omega_0 = 0.5 \times 2\pi$kHz). This indicates that the mean responses encode the signal. Higher absolute values of $\mu_{R|s}(\omega)$ for MM than for VM are consistent with the observed higher information transmission through MM. The spikes were generated using the same parameters as in figure 3.6.

In fact, I find that the (complex) mean Fourier coefficients, $\mu_{R|s}(\omega)$, that result from repeated presentation of the same signal are the signal encoding quantities (see figure 3.12C,D). These signal-conditioned mean values are finite for frequencies $\omega \sim \Omega_0$ that are dominant in the signal. Within linear response theory, these quantities can be computed as $\mu_{R|s}(\omega) = \chi(\omega)\tilde{s}(\omega)$. Therefore, in the linear regime, a perfect decoder would need to extract the values $\mu_{R|s}(\omega)$ from the spike trains in order to reconstruct the Fourier transform of the signal. Outside the linear regime, at higher σ_s, signal decoding may require nonlinear transformations of $\mu_{R|s}(\omega)$, possibly including frequency coupling (see also sections 3.7 and 4.6). Nonetheless, the values $\mu_{R|s}(\omega)$ in principle carry all available signal information. Comparing the values of $\mu_{R|s}(\omega)$ for MM and VM in figures 3.12C,D, they show similar shape, but the amplitude of the deflections is higher in MM. This suggests a better decodability of MM and agrees with the higher information transmission through MM in the shown case.

If the signal is varied from trial to trial, the mean Fourier coefficients, μ_R, vanish as observable in figure 3.12A,B. This means that no specific signal can be decoded from spike train statistics that result from different signals, which is an intuitive finding.

Anticipating the discussion of the upper bound in section 3.5, the values $\mu_{R|s}(\omega)$ can be identified with the deterministic, signal-dependent – and signal-encoding – response feature $r_{\mathrm{det}}(s)$. According to section 3.4.1, the Fourier coefficients are, moreover, normal distributed around their means across trials for both varying and fixed signals. These two properties of the spike train Fourier coefficients – signal-dependent mean and normal distribution – imply that assumptions (1) and (2) in section 3.5 for deriving the upper bound $\mathcal{I}^{\mathrm{ub}}$ are naturally fulfilled.

Given that the values $\mu_{R|s}(\omega)$ encode the signal, one can develop a better intuition for the information-theoretic meaning of the Fourier coefficient variances that implicitly enter $\mathcal{I}^{\mathrm{tot}}$. An alternative representation of the mutual information and correlation functions is [98],

$$\mathcal{I}(\omega) = \frac{1}{2}\log_2\left(\frac{\sigma_R^2(\omega)}{\sigma_{R|s}^2(\omega)}\right), \tag{3.13}$$

$$\sigma_{R|s}^2(\omega) = \frac{1}{T}\left(C_{\mathrm{auto}}(\omega) - C_{\mathrm{cross}}(\omega)\right), \tag{3.14}$$

$$\sigma_R^2(\omega) = \frac{1}{T}C_{\mathrm{auto}}(\omega), \tag{3.15}$$

where $\sigma^2_{R|s}(\omega)$ and $\sigma^2_R(\omega)$ are the variances of the Fourier coefficients $\tilde{\rho}(\omega)$ at fixed and varying signal, respectively. According to Eq. (3.13), the signal-conditioned variances $\sigma^2_{R|s}(\omega)$ are inversely related to the mutual information which can be understood considering that the means $\mu_{R|s}(\omega)$ carry the signal information: the larger the variances $\sigma^2_{R|s}(\omega)$ the worse is the discriminability of different signal-encoding mean values $\mu_{R|s}(\omega)$. Or, in other words, a signal-encoding $\mu_{R|s}(\omega)$ can be decoded most reliably in a given trial when the trial-to-trial variability $\sigma^2_{R|s}(\omega)$ is small. For varying signals, on the other hand, the variances $\sigma^2_R(\omega)$ represent the overall range of the values $\mu_{R|s}(\omega)$ across different signals and, therefore, reflect the bandwidth in which signals can be encoded. Consequently, higher $\sigma^2_R(\omega)$ are associated with higher information transmission. A discussion of decoding from the Fourier coefficients is also part of section 4.6.

3.5 Putative upper bound is equivalent to exact information

The Fourier analysis of the spike trains of the previous section allows to derive an important result regarding the upper bound $\mathcal{I}^{\text{ub}}$ introduced as Eq. (1.25) in section 1.16. Here, I show that under the assumption of stationary, stochastic signals the upper bound actually corresponds to the exact information $\mathcal{I}^{\text{tot}}$.

The upper bound derived in previous studies relies on two key assumptions [218, 224, 226]:

(1) The information about a signal s is encoded exclusively in a deterministic response feature $r_{\text{det}}(s)$ (typically the PSTH, recorded from repeated presentations of s).

(2) At any given trial i, the response $r_i(s)$ has a Gaussian noise component such that $r_i(s) = r_{\text{det}}(s) + N_i$ with $N_i \sim \mathcal{N}$.

The mutual information of the responses $\{r_i\} = R$ and the signal $s \in S$ is then given by (see the definitions of entropy and mutual information in section 1.3.2.1):

$$\mathcal{I}(R,S) = H(R) - \langle H(R|s)\rangle_s = H(R) - \langle H(N|s)\rangle_s . \tag{3.16}$$

According to assumption (2), the noise entropy is that of a normal distribution with variance $\text{Var}[R|S] = \text{Var}[N|S]$. To obtain the upper bound, one assumes a Gaussian

response distribution[10] with a variance $\mathrm{Var}[R]$ to determine $H(R)$ in Eq. (3.16). This leads to (an alternative representation is given by Eq. (1.25); cf. also Eq. (3.13))

$$\mathcal{I}^{\mathrm{ub}}(R,S) = \frac{1}{2}\log_2\left(\frac{\mathrm{Var}[R]}{\mathrm{Var}[R|S]}\right) \geq \mathcal{I}^{\mathrm{tot}}(R,S). \tag{3.17}$$

Now, this calculation is applied to the Fourier coefficients of the spike trains for each frequency. The ensemble $R = R(f)$ then corresponds to the set of real and imaginary parts of the Fourier coefficients $\tilde{\rho}(f)$ of the response spike trains at each frequency f (see section 1.3.2.2).

In section 3.4.3, I showed that assumptions (1) and (2) are naturally fulfilled on the level of the spike train Fourier coefficients for stationary signals with finite correlation time. Moreover, in section 3.4.1, I demonstrated that Gaussianity and independence of the Fourier coefficients $R(f) = \{\mathrm{Re}(\tilde{\rho}(f)),\ \mathrm{Im}(\tilde{\rho}(f))\}$ emerge from stationarity, too, and are not just a limiting case. Overall, this leads to the conclusion that equality holds in Eq. (3.17), such that

$$\mathcal{I}^{\mathrm{ub}} = \mathcal{I}^{\mathrm{tot}}. \tag{3.18}$$

Hence, the putative upper bound is in fact equivalent to the full mutual information.

Alternatively, the equivalence of both information measures can be proven formally by rewriting Eq. (3.17) for the upper bound. For R_1 and R_2 being two responses to the same stimulus it follows from Eq. (B.2) of Roddey et al. [225]:

$$\langle \mathrm{Var}[R_1|S]\rangle = (1-\gamma_{R_1R_2})\,\mathrm{Var}[R_1], \tag{3.19}$$

with the response-response coherence $\gamma^2_{R_1R_2} = \frac{\langle R_1^* R_2\rangle\langle R_1 R_2^*\rangle}{\langle R_1^* R_1\rangle\langle R_2^* R_2\rangle}$. Insertion in Eq. (3.17) yields

$$\mathcal{I}^{\mathrm{ub}}(R,S) = \frac{1}{2}\log_2\left(\frac{\mathrm{Var}[R_1]}{(1-\gamma_{R_1R_2})\,\mathrm{Var}[R_1]}\right) = -\frac{1}{2}\log_2\left(1-\gamma_{R_1R_2}\right). \tag{3.20}$$

Again, considering the responses to be the spike train Fourier modes, i.e. $R_1 = \tilde{\rho}_1(f)$ and $R_2 = \tilde{\rho}_2(f)$, it is $\gamma^2_{R_1R_2} = \gamma^2_{\rho_1\rho_2}(f) = \lim_{T\to\infty}\frac{|\langle\tilde{\rho}_1(f)\tilde{\rho}_2^*(f)\rangle|^2}{|\langle\tilde{\rho}_1(f)\rangle|^2|\langle\tilde{\rho}_2(f)\rangle|^2}$. Now, for long recording sequences one can identify $|\langle\tilde{\rho}_1(f)\tilde{\rho}_2^*(f)\rangle| = C_{\mathrm{cross}}(f)$ and $|\langle\tilde{\rho}_1(f)\rangle|^2 = |\langle\tilde{\rho}_2(f)\rangle|^2 = C_{\mathrm{auto}}(f)$

[10] The normal distribution is the distribution of maximum entropy at a given mean and variance [204].

using standard stationarity assumptions and Eqs. (1.20), (1.22), and (1.26). From inserting these identities in Eq. (3.20) again follows the equivalence of $\mathcal{I}^{\mathrm{ub}}$ and $\mathcal{I}^{\mathrm{tot}}$.

3.6 Comparing information transmission in mean and variance coding

The introduced methods at hand, I conduct a comparative analysis of the information that is transmitted by the spikes of leaky and exponential integrate-and-fire neurons about mean- and variance-modulating signals. The analysis rests on evaluating the mutual information $\mathcal{I}^{\mathrm{tot}}$ (and partly $\mathcal{I}^{\mathrm{lin}}$) for the two encoding schemes at different parameters.

Figures 3.13 and 3.14 report the core results of the information-theoretic analysis with focus on the role of signal strength, noise strength (firing rate) and noise correlation time – for leaky integrate-and-fire (LIF) and exponential integrate-and-fire (EIF) neurons, respectively. Generally, the mean channel (always shown in red colors) invariably outperforms the variance channel (blue) in terms of information transmission when small to moderate central signal frequencies are considered (for signals with power at larger frequencies the assumption of white noise background can change this picture in LIF neurons; see section 3.6.2). This finding holds across a broad set of signal strengths (figures 3.13A,D and 3.14A,D), noise strengths/firing rates (figures 3.13B,E and 3.14B,E) and different noise time constants (figures 3.13C,F and 3.14C,F). [Gray, solid curves represent the channel information ratio $\beta^{\mathrm{tot}} = \mathcal{I}^{\mathrm{tot}}_{\mathrm{MM}}/\mathcal{I}^{\mathrm{tot}}_{\mathrm{VM}}$.]

Mean coding outperforms variance coding by as much as two orders of magnitude for small signal strengths in LIF neurons (figure 3.13D). This difference is less pronounced but still significant in EIF neurons (Figure 3.14D) where information transmission tends to be lower overall. For large signal strengths, the information contents of the two channels become more similar. Interestingly, for VM information transmission does not saturate at high signal strengths, even though a considerable overmodulation $[s(t) < -1]$ can be expected to occur frequently. In these instances, the VM input currents are zero (see section 3.1.1) and do not themselves convey detailed stimulus information, and should thus decrease the transmittable information. In MM, in contrast, overmodulated currents are distinguishable from subthreshold inputs.

Figures 3.13B and 3.14B show that increasing noise variance (and thereby increasing firing rate) reduces the mutual information in both coding strategies and neuron models, whereby the effect seems stronger for MM. This is consistent with (but not exclusively explainable by) the decay of the linear response functions Eq. (A.18) and (A.19): for large noise strengths, $\chi_{\mathrm{MM}}(\omega)$ scales inversely with σ_n whereas $\chi_{\mathrm{VM}}(\omega)$ is not proportionally related to σ_n. Moreover, because of previously described effects of *stochastic resonance* the information is not expected to be a monotonic function of the noise strength but may have local maxima [107], which are not detected here due to the constraints of sampling finite parameter sets. Increasing noise and firing rate also generally reduce the gap between variance and mean coding (see figures 3.13E and 3.14E). Furthermore, increasing the noise time constant strongly diminishes the information in the variance channel while boosting it in the mean channel (figures 3.13C and 3.14C). As a consequence, the analysis reports a growing advantage of MM over VM coding for larger noise time constants (figures 3.13F and 3.14F). I did not find a marked dependence of the information on the signal correlation time within the given parameter range (see figures 3.15 and 3.16).

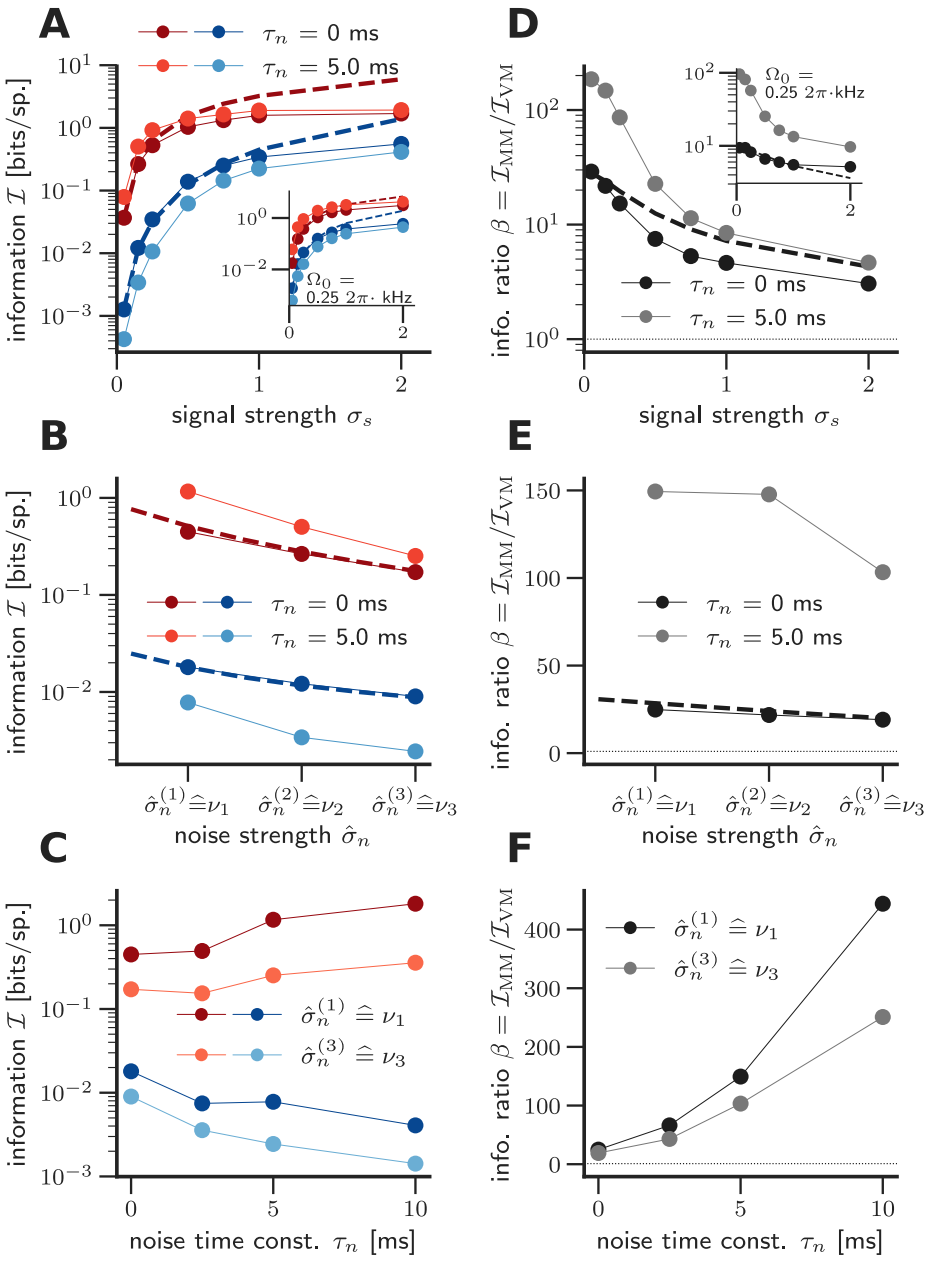


FIGURE 3.13: **Information capacity of mean and variance coding in leaky integrate-and-fire neurons.** **(A)** Information increases monotonically as signal strength grows for both mean (MM, red lines and symbols) and variance coding (VM, blue lines and symbols). The dashed lines (blue and red) denote the respective linear approximations $\mathcal{I}^{\mathrm{lin}}$, circles denote numerical simulations of $\mathcal{I}^{\mathrm{tot}}$. The information is an approx. linear function of the signal strength for smaller σ_s and sublinear for larger σ_s (cf. figure 3.24). This finding is conserved for alternative values of the central signal frequency Ω_0 (inset). **(B)** Increasing the noise amplitude while keeping the signal strength constant reduces the information content for both modulations. The normalized noise strengths $\hat{\sigma}_n = \hat{\sigma}_n^{(1)}, \hat{\sigma}_n^{(2)}, \hat{\sigma}_n^{(3)}$ are increasing with equal increments and chosen to produce the three different firing rates $\nu_1 \approx 12$ Hz, $\nu_2 \approx 17$ Hz, and $\nu_3 \approx 22$ Hz for each τ_n (see section 3.3). **(C)** Noise time constant τ_n has opposing effects: It reduces information for VM but increases it for MM. **(D)-(F)** The information channel ratio $\beta^{\mathrm{tot}} = \mathcal{I}^{\mathrm{tot}}_{\mathrm{MM}}/\mathcal{I}^{\mathrm{tot}}_{\mathrm{VM}}$ (solid lines and symbols) and $\beta^{\mathrm{lin}} = \mathcal{I}^{\mathrm{lin}}_{\mathrm{MM}}/\mathcal{I}^{\mathrm{lin}}_{\mathrm{VM}}$ (dotted lines) obtained from (A)-(C). The information content is higher for MM signals across a wide parameter regime: All curves lie above $\beta^{\mathrm{tot}} = 1$ (thin, dashed line). The information content for MM and VM becomes more similar for high σ_s and increasing noise $\hat{\sigma}_n$ ((D),(E)). Larger noise correlation times favor mean coding (F). Parameters here are: $\sigma_s = 0.15$ in (B), (C), (E), and (F); $\Omega_0 = 0$ except insets in (A) and (D); $\hat{\sigma}_n = \hat{\sigma}_n^{(2)}$ in (A), (D) and $\hat{\sigma}_n = [\hat{\sigma}_n^{(1)}, \hat{\sigma}_n^{(3)}]$ in (C), (F); $\tau_s = 20$ ms everywhere.

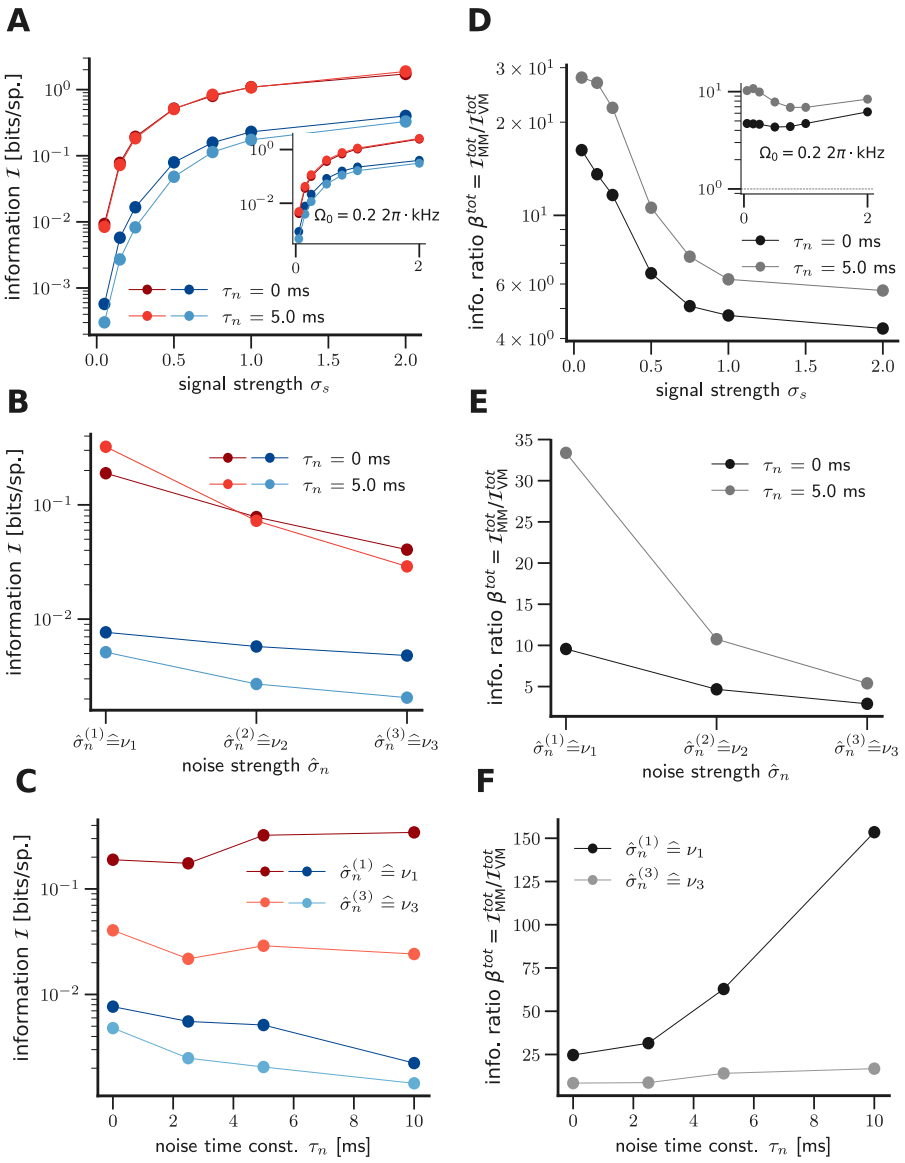


FIGURE 3.14: **Information capacity of mean and variance coding in exponential integrate-and-fire neurons.** In (A),(B),(C) information for mean coding (red) and variance coding (VM) is shown for EIF neurons as a function of signal and noise strength/firing rate and noise correlation time (see figure 3.13 for comparison to LIF neurons). **(A)** Information increases monotonically as a function of the signal strength. **(B)** Increasing the firing rate through increased noise has a reducing effect on the information content for both modulations. **(C)** The effects of increasing noise correlation time are opposite for mean and variance coding: $\mathcal{I}^{\text{tot}}_{\text{MM}}$ increases with τ_n, while $\mathcal{I}^{\text{tot}}_{\text{VM}}$ decreases. **(D)-(F)** The information ratio $\beta^{\text{tot}} = \mathcal{I}^{\text{tot}}_{\text{MM}}/\mathcal{I}^{\text{tot}}_{\text{VM}}$ corresponding to (A)-(C). All curves lie above the line $\beta^{\text{tot}} = 1$ and accordingly the information transmission for MM is consistently larger than for VM. From **(D)** follows that mean and variance coding become more similar for high σ_s. Panel **(E)** indicates that variance coding is less impacted by increasing noise strength/firing rate. **(F)** Larger noise correlation times favor mean coding. Parameters are $\sigma_s = 0.15$ in (B), (C), (E), and (F); $\Omega_0 = 0$ (except insets), $\hat{\sigma}_n = \hat{\sigma}_n^{(2)}$ in (A), (C), (D), and (F), and $\tau_s = 20$ ms everywhere.

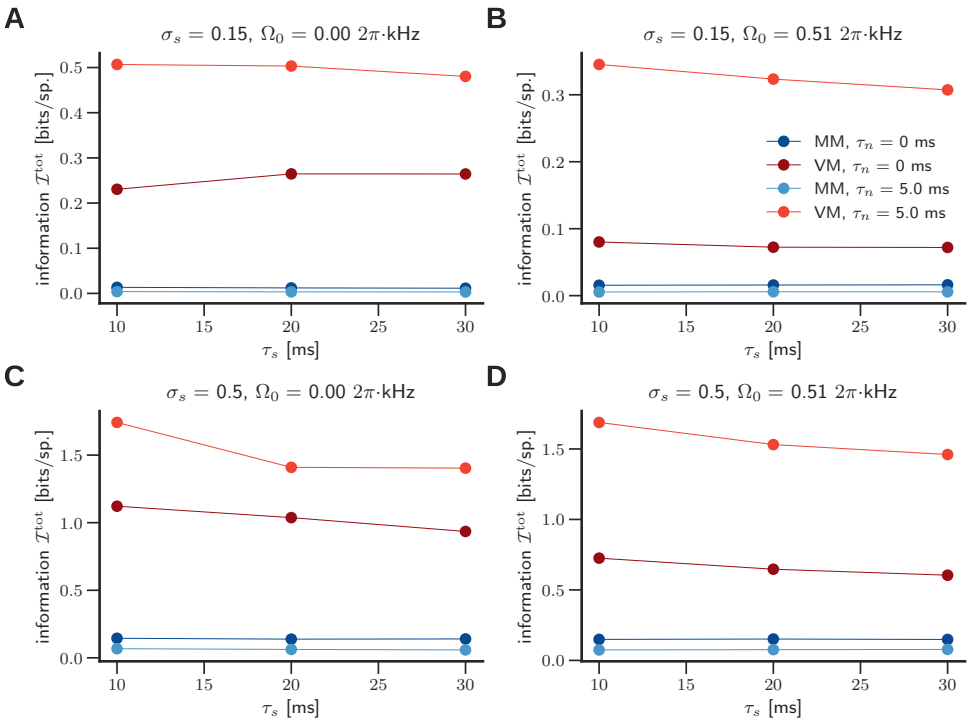


FIGURE 3.15: **Signal correlation time τ_s plays only a minor role for the information content in leaky integrate-and-fire neurons.** Mean and variance encoding are denoted by red and blue colors, respectively (dark colors: $\tau_n = 0$ ms, brighter colors: $\tau_n = 5$ ms). In all panels $\hat{\sigma}_n = \hat{\sigma}_n^{(2)} \mathrel{\widehat{=}} \nu_2$. Changes of the correlation time produce only relatively small changes in the mutual information.

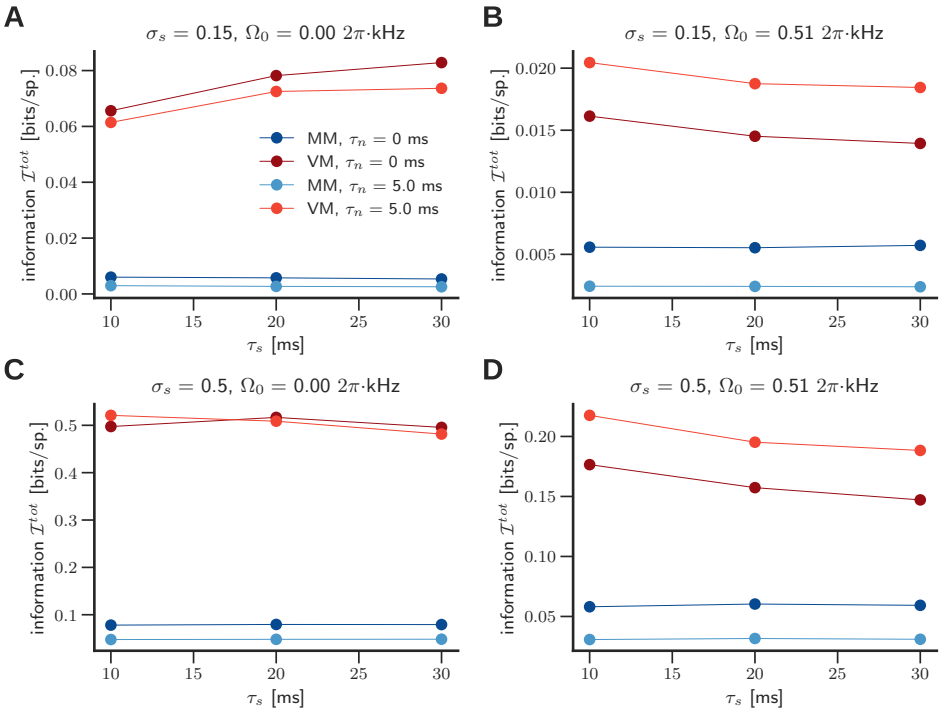


FIGURE 3.16: **Signal correlation time τ_s plays only a minor role for the information content in exponential integrate-and-fire neurons.** Same as figure 3.15 but for EIF neurons.

3.6.1 Quality of linear approximation $\mathcal{I}_{\text{lin}}$

In section 3.2, I introduced the linear response function to provide an intuition for the parameter dependence of the information transmission in mean and variance coding. But how well does $\mathcal{I}^{\text{lin}}$ approximate the true information $\mathcal{I}^{\text{tot}}$ quantitatively?[11]

As one would expect, in the linear regime of small signal strengths the linear information approximation closely resembles the full information content, but for large signal strengths, the linear approximation overestimates the full information content, as can be seen in figure 3.13A. This estimation error has two main origins. First, the assumption that the presence of a signal does not change the average firing rate compared to spontaneous activity is part of the linear approximation but does not hold at higher signal strengths. In these cases, the firing rate ν is underestimated by ν_0 [Eq. (A.16)] and consequently, according to Eq. (3.9), the overall information per spike is overestimated by $\mathcal{I}^{\text{lin}}$. This can be best seen in the autocorrelation functions in figures 3.17E,F where the effect is most pronounced for MM (the firing rate is represented by the high frequency limit of the autocorrelation). Moreover, the firing rate also influences the overall shape of the autocorrelation, and thus rate estimations errors also introduce more complex approximation errors to $\mathcal{I}_{\text{lin}}(\omega)$.

Second, when the dynamic range of the rate responses becomes large due to increased signal strength, this leads to overmodulation. Negative rate modulations are then "clipped" because the firing rate cannot become negative/imaginary, and similar effects occur for positive modulations when the firing rate saturates. However, linear response theory always assumes linear signal-induced rate modulations and thus predicts higher responsiveness, i.e. an unnaturally high dynamic range, to large rate modulations. As a result, linear response theory predicts too high mutual information. As an example, figure 3.17 shows the frequency-resolved correlation functions and mutual information (exact, linearly decodable and linear approximation) for a relatively high signal strength of $\sigma_s = 0.5$. It can be seen that the linear approximations overestimate the cross-correlation and information around Ω_0 because they cannot capture the damping of the response amplitudes due to saturation at large signals.

[11] Not to be confused with the linearly decodable information $\mathcal{I}^{\text{ld}}$, which is discussed in section 3.7.1.

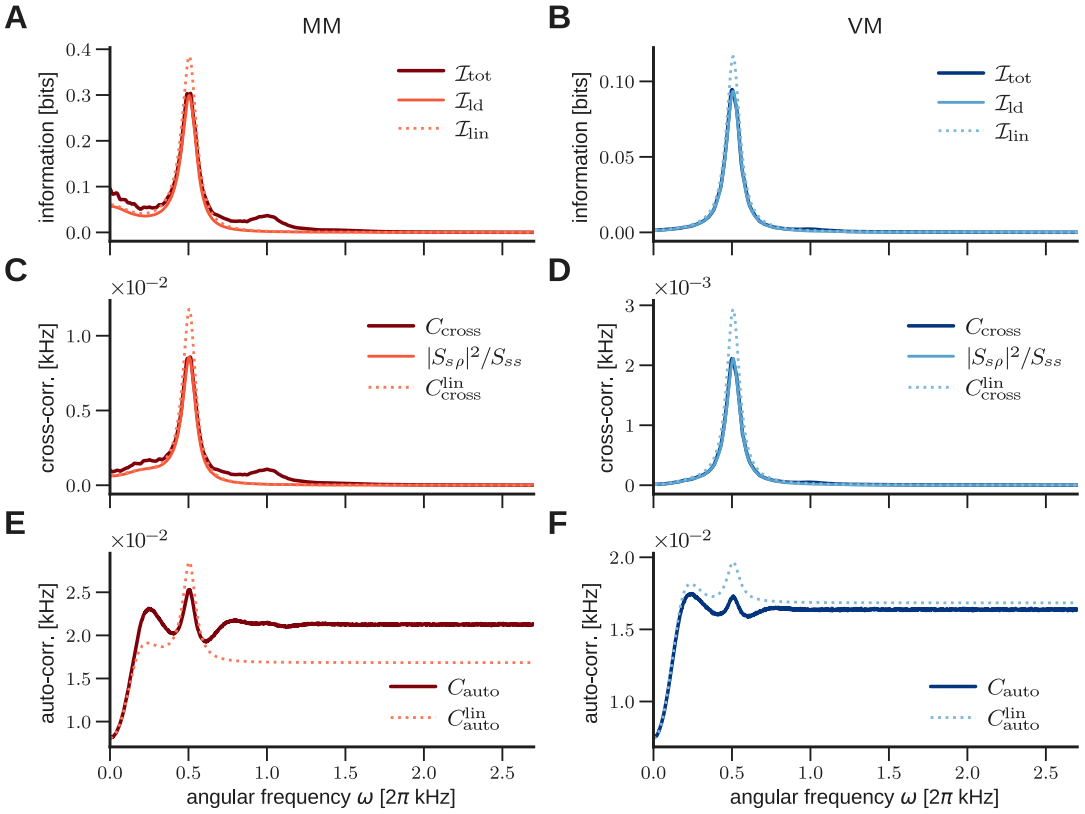


FIGURE 3.17: **Linear approximations of frequency-resolved information and spike correlation functions significantly deviate from their exact counterparts at large signal strengths.** I compare the frequency resolved information Eq. (1.23) and the corresponding correlation functions to the information approximations Eq. (3.9) and (1.24), and their corresponding correlation functions at a signal strength of $\sigma_s = 0.5$. MM is shown in the left column, VM in the right column. (**A**,**B**) Frequency resolved information $\mathcal{I}_{\text{tot}}$ (dark, solid line), linearly decodable information $\mathcal{I}_{\text{ld}}$ (light, solid) and linear approximation $\mathcal{I}_{\text{lin}}$ (dashed). The differences between the three information quantities are a direct consequence of the different estimates of the spike correlation functions. (**C**,**D**) Spike cross-correlation function C_{cross} and its respective approximations. Nonlinearities manifest themselves as second-order harmonics and in the low frequency range (much stronger for MM (C)). Furthermore, $C^{\text{lin}}_{\text{cross}}$ overestimates the correlation around $\omega = \Omega_0$. (**E**,**F**) The spike autocorrelation function C_{auto}. Significant deviations of the linear approximation $C^{\text{lin}}_{\text{auto}}$ and the complete autocorrelation can be observed around Ω_0 and a considerable underestimation of the firing rate is present in MM (E). I note that while the autocorrelations have similar amplitudes for mean and variance modulation, the value ranges of the cross-correlation functions differ in the two channels. Here, I used the following parameters $\Omega_0 = 0.51 \times 2\pi$kHz, $\hat{\sigma}_n = \hat{\sigma}_n^{(2)}$, $\tau_n = 0$, $\tau_s = 20$ms.

3.6.2 Influence of spike initiation time on information content

In this section, I analyze the effect of neuronal dynamics, more precisely the spike initiation time, on the transmission of signals with different "velocity". Here, velocity is represented by the dominant signal frequency Ω_0; high frequencies correspond to fast signals. As explained in section 1.1.4, the EIF model allows to incorporate a finite

spike initiation time, whereas the LIF model corresponds to the limit of instantaneous spiking. Figure 3.18 shows how the mutual information in LIF and EIF neurons scales as a function of Ω_0 and reveals crucial differences between the two neuron models as the central frequency increases.

In LIF neurons, information transmission converges to finite values for large signal frequencies if the noise is temporally correlated in both the mean and variance channel (figure 3.18A). This behavior is reproduced at different signal strengths (see inset to figure 3.18A), and is consistent with the fact that for colored noise the response functions of LIF neurons remain finite in the high frequency limit [107, 124]. If the noise is white, on the other hand, the information in mean coding declines with growing frequency. Figure 3.18C indicates that for white noise background in the limit of a large central signal frequency, variance modulation can be more beneficial for LIF neurons. This echoes the hypothesis by Silberberg *et al.* [85] and Tchumatchenko *et al.* [114] that variance modulation can evoke stronger responses to fast changing signals on a population level. In contrast, studying the information ratio of mean and variance coding (figure 3.18C) for colored noise (light gray), I find that mean coding is a more efficient coding strategy by approximately two orders of magnitude.

In figure 3.18B and its inset the information content in EIF neurons is shown to decline with the central signal frequency for both coding schemes – in accordance with the known $1/\omega$ decay of the linear response functions in both cases [36] (see table 1.2). Interestingly, mean coding nonetheless still outperforms variance coding by approximately one order of magnitude (see also figure 3.18D), but I notice that both coding schemes become more similar as the central signal frequency increases.

Overall, I find that mean coding outperforms variance coding, except for the case of LIF neurons with white background noise and high-frequency signals. Since postsynaptic potentials always possess finite decay times, noise currents cannot be perfectly white [260]. In cases where finite correlation times in the noise are present (even if they are as small as 5 ms) mean coding provides higher information transmission also in LIF neurons. Interestingly, in the EIF model, which incorporates the finite spike initiation times as generated by the sodium channel dynamics in cortical neurons [35], I find a consistent advantage of mean over variance coding for both white and nonwhite noise. I, therefore, conclude that mean coding has higher information transmission capabilities

in biologically plausible settings. Comparing the values for information content and responsiveness to fast signals in LIF and EIF neurons, I emphasize the importance of spike initiation dynamics for information transmission. In particular, a finite spike initiation time – which can be considered the biological relevant case – introduces cut-offs for information carrying frequencies.

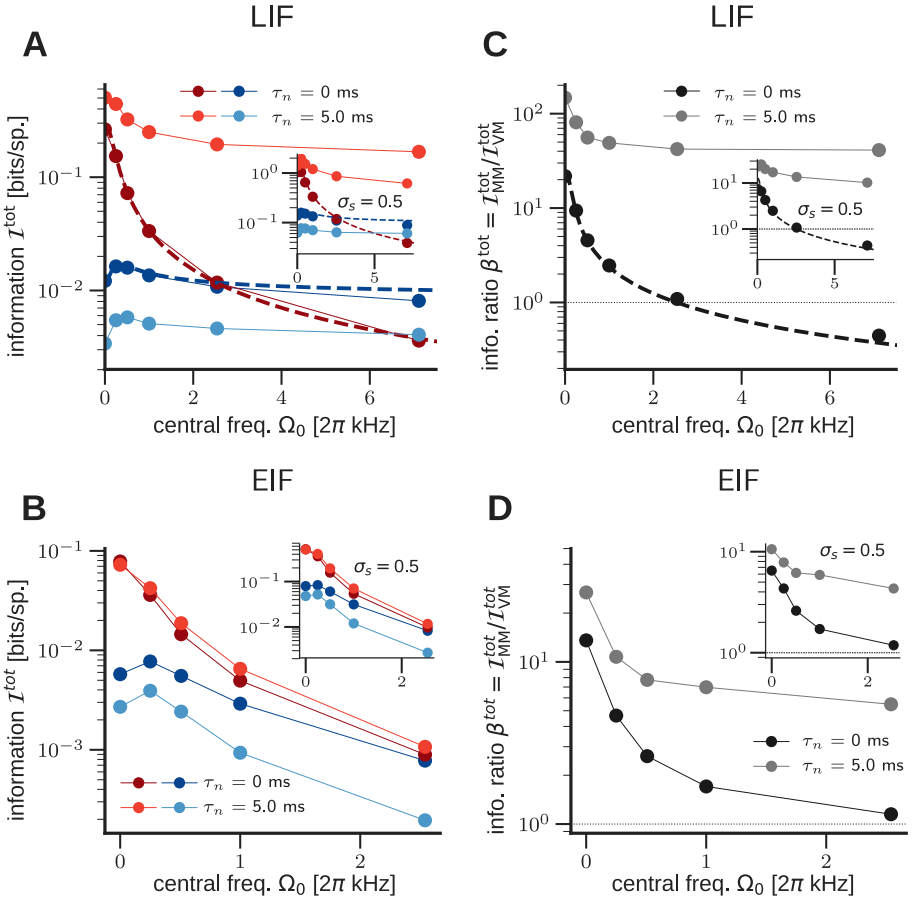


FIGURE 3.18: **Introducing a finite spike initiation time reduces overall information transmission but has less effect on variance coding. (A)** The mutual information for mean (red lines and symbols) and variance coding (blue lines and symbols) for LIF neurons with nonwhite noise saturates to a finite value as the central signal frequency grows. Mean encoded information content for large signal frequencies Ω_0 and white noise decays to zero (dashed red line). **(B)** In EIF neurons, I also observe that mean coding has higher information content. Information content in both channels decays as the central signal frequency grows regardless of noise correlation time. Local maxima exist for VM in both neuron models at small Ω_0. **(C)** Mean coding carries more signal information for finite noise correlation times (gray line). If the noise is white, then the ratio $\beta^{\mathrm{tot}} = \mathcal{I}^{\mathrm{tot}}_{\mathrm{MM}}/\mathcal{I}^{\mathrm{tot}}_{\mathrm{VM}}$ crosses 1, and variance outperforms mean coding for sufficiently large Ω_0 in LIF neurons. **(D)** In EIF neurons, mean coding has higher information transmission regardless of noise correlation times. For both neuron models, increasing Ω_0 reduces the ratio β^{tot} and favors VM coding. Parameters are $\sigma_s = 0.15$ (except insets), $\hat{\sigma}_n = \hat{\sigma}_n^{(2)}$ and $\tau_s = 20$ ms everywhere.

3.7 Linear and nonlinear components of information encoding

In introduction section 1.4, I explained that generally little is known about the amount of nonlinearly encoded, or decodable, information in spike trains. This also applies to encoding via mean and variance modulations and is due to a (previous) lack of a method to compute the mutual information capturing all linear and nonlinear effects. As written before, an according research question can be phrased as, "To what extent can information transmission be understood in terms of linear input-output relations and what influences the encoding linearity?". Having the correlation method for the full mutual information at hand and comparing it to the linearly decodable information (cf. section 1.3.2.2), it is possible to elaborate on this question. I conduct a respective investigation in this section based on mean and variance modulated input currents and LIF/EIF neurons. In subsection 3.7.1, I uncover general relations of linearly and nonlinearly encoded information, and in subsection 3.7.2 I analyze the major effects in more detail on the level of spike correlation functions.

3.7.1 Proportion of linearly encoded information is inversely related to total information

The information content that can be decoded linearly is given by $\mathcal{I}_{\mathrm{ld}}(\omega)$, defined in Eq. (1.24), which is a lower bound for the complete mutual information. Moreover – and this is the crucial property of $\mathcal{I}_{\mathrm{ld}}(\omega)$ here – the lower bound does not capture nonlinear signal-response correlations, and thus contrasting it with $\mathcal{I}_{\mathrm{tot}}(\omega)$ reveals linearly and nonlinearly decodable components of information transmission. As a reminder, $\mathcal{I}_{\mathrm{ld}}(\omega)$ is not equivalent to $\mathcal{I}_{\mathrm{lin}}(\omega)$; the latter is not a strict approximation or bound to the full information (see section 3.2). To assess the fraction of the mutual information that is encoded linearly, I define the information linearity index

$$\lambda^{\mathrm{ld}} = \frac{\mathcal{I}^{\mathrm{ld}}}{\mathcal{I}^{\mathrm{tot}}} \leq 1.$$

Figures 3.19 and 3.20 show the dependencies of this index on the complete information and different signal and noise parameters, as explained in the following.

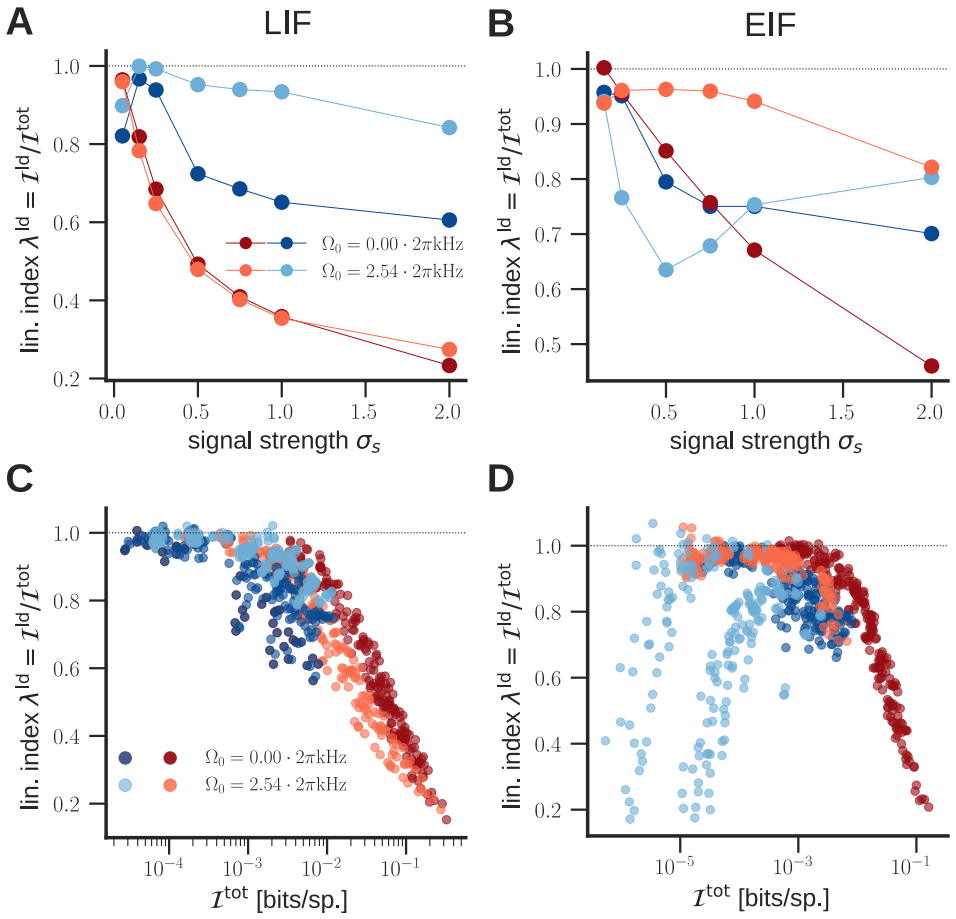


FIGURE 3.19: **Linearity of encoding is a function of the total information.** The fraction of linearly decodable information is measured by $\lambda^{\mathrm{ld}} = \mathcal{I}^{\mathrm{ld}}/\mathcal{I}^{\mathrm{tot}}$. **(A)** As a function of increasing signal strength the deviations from linearity increase faster for mean coded (MM, red circles) than for variance coded signals (VM, blue circles) in LIF neurons. **(B)** The situation is different for EIF neurons. As shown here, for VM the fraction of nonlinearly encoded information can be a non-monotonic function of the signal strength. For MM the curves are similar to those in LIF neurons. **(C),(D)** Many simulated data points are shown together (see below). The two different values for the central signal frequency Ω_0 are shown as different shades of red circles/blue crosses. For LIF neurons, the linearity λ^{ld} is predominantly a function of the total information $\mathcal{I}^{\mathrm{tot}}$ that appears to be the same for MM and VM (shades of red and blue). The effect of increased central signal frequencies tends to be opposite for VM and MM: higher frequencies (lighter colors) increase λ^{ld} in variance coding and decrease it in mean coding (C). For EIF neurons, high values of Ω_0 are related to clear reductions of the linearity λ^{ld} within variance coding (D). The effect is illustrated in Fig. 3.21 and yields to deviations in the overall dependence of total information and its linearly decodable fraction. In all plots the thin, dashed line denotes $\lambda^{\mathrm{ld}} = 1$ where exact information and lower bound coincide. Parameters are $\sigma_s = [0.15, 0.25, 0.5, .75, 1, 2]$ in (C) and (D); $\hat{\sigma}_n = \hat{\sigma}_n^{(2)}$ in (A) and (B), and $\hat{\sigma}_n = [\hat{\sigma}_n^{(1)}, \hat{\sigma}_n^{(2)}, \hat{\sigma}_n^{(3)}]$ in (C) and (D); $\tau_n = 10$ ms in (A) and (B), and $\tau_n = [0, 2.5, 5, 10]$ ms in (C) and (D); $\tau_s = 20$ ms everywhere.

Figure 3.19A,B shows that λ^{ld} drops with increasing signal strength in most cases and approaches 1 for vanishing σ_s. This is in line with the intuition that small signals are encoded linearly and that nonlinearities arise progressively as the signal grows. However, a crucial observation can be made. In EIF neurons, variance coding of high-frequency

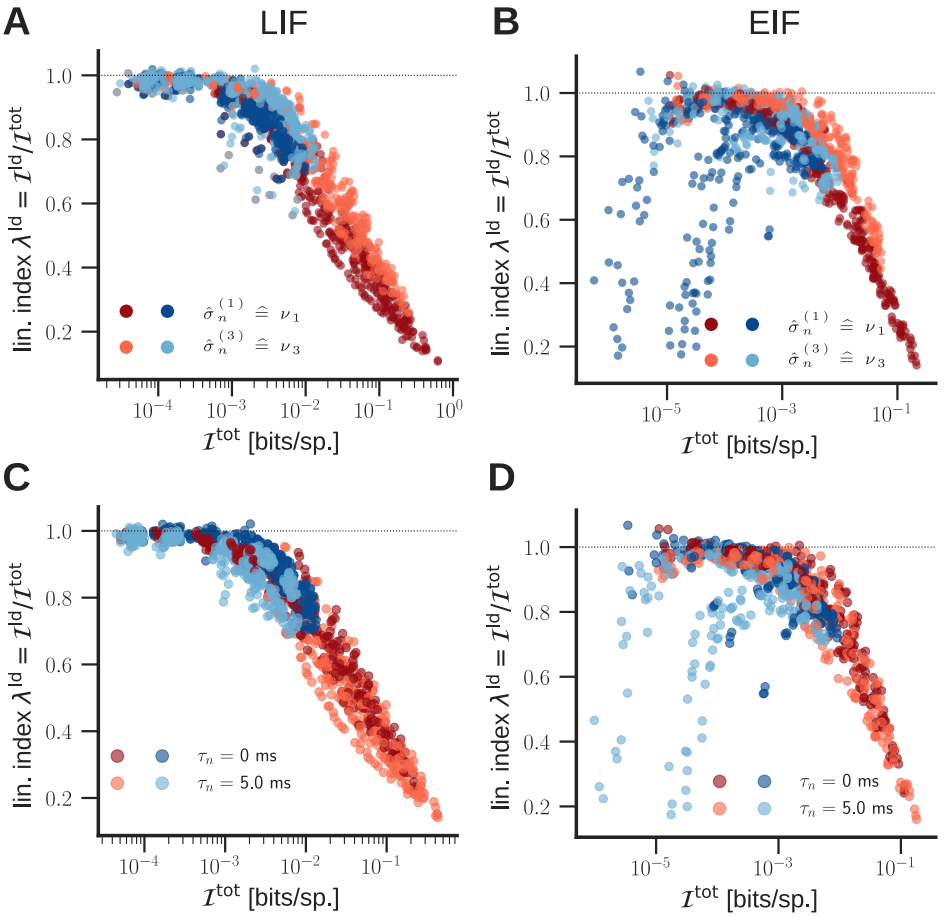


FIGURE 3.20: **The fraction of linearly decodable information is enhanced by larger noise strengths/firing rates.** As in figure 3.19, a clear overall dependence of λ^{ld} on the total information can be observed in both neuron models. The color-code highlights the role of firing rate determining noise strength ((A),(B)) and noise correlation time ((C),(D)) for both LIF and EIF neurons ((A),(C) and (B),(D), respectively). Lighter colors correspond to higher values. **(A)** Increased σ_n (ν) increases the linearity $\lambda^{\mathrm{ld}} = \mathcal{I}^{\mathrm{ld}}/\mathcal{I}^{\mathrm{tot}}$ in both modulation schemes. **(B)** The same effect is visible in EIF neurons but less clear. Moreover, the data points stemming from large Ω_0 introduce subbranches to the overall structure (cf. section 3.7.2). **(C)** The noise correlation time τ_n has a de-linearizing effect in mean and variance coding for LIF neurons, which is more pronounced for VM. **(D)** No clear tendency is observable in EIF neurons. In all plots the thin, dashed line denotes $\lambda^{\mathrm{ld}} = 1$ where exact information and lower bound coincide. The parameters used in this figures are the same as in Fig. 3.19C,D but including all Ω_0 and with the color-coded parameter as indicated in the legends.

signals leads to local minima of λ^{ld} as a function of the signal strength (figure 3.19B). For low Ω_0 (dark colors), the curves are similar in both neuron models and share a feature: the fraction of linearly encoded information remains considerably higher when variance modulations are used rather than mean modulations. This is not an obvious finding because the signal strength σ_s in both cases equivalently describes the signal-driven modulation depth (of the input's mean or variance; see Eqs. (3.1), (3.2)), and conceptually the linearly decodable information stems from a linearization in σ_s [128]

(an analytic derivation is given below). I find a distinct behavior for EIF neurons at higher Ω_0 for which a local minimum of λ^{ld} exists.

Taking another point of view, from figures 3.19C,D and 3.20 also follows that the main determinant for the linearity is given by the total information $\mathcal{I}^{\mathrm{tot}}$ itself. Here, a visualization of pooled data for both coding schemes and neuron models (with different color lightness for different parameter values) demonstrates that data points are distributed around a decaying curve in the $\mathcal{I}^{\mathrm{tot}}$-$\lambda^{\mathrm{ld}}$ plane. Again, an aberrant dependency is found for variance coding in EIF neurons at high signal frequencies (figures 3.19D and 3.20B,D). This particular behavior emphasizes the importance of the spike initiation time for the responsiveness of neurons to high frequencies and is analyzed in more detail in figure 3.21B,D. For LIF neurons, higher signal frequencies generally decrease the linearity at a given total information in mean coding and increase it in variance coding. In EIF neurons, the effect of larger signal frequencies is generally de-linearizing, in accordance with experimental work [225, 317].

Additionally, I find that larger noise amplitudes and firing rates lead to a more linear encoding for both neuron models and encoding schemes (figures 3.20A,B), which is in line with previous reports [275, 288, 318]. In contrast, larger noise correlation times increase the contribution of nonlinear encoding. This effect is most pronounced in mean coding (see figure 3.20C,D).

3.7.1.1 Mapping the complete mutual information and linearly decodable information for small signals

As shown, the linearly decodable information and complete information converge to the same values for small signal strengths, i.e. $\lambda^{\mathrm{ld}} \to 1$ as $\sigma_s \to 0$ (see figures 3.19A,B). Here, I derive this relation analytically.

As explained in section 3.2, the cross-correlation function can be analytically approximated through $C_{\mathrm{cross}}(f) \approx |\chi(f)|^2 S_{ss}(f)$ for small signals. Inserting in the correlation method Eq. (1.23) yields

$$\mathcal{I}_{\mathrm{tot}}(f) \approx -\frac{1}{2}\log_2\left(1 - \frac{|\chi(f)|^2 S_{ss}(f)}{C_{\mathrm{auto}}(f)}\right). \tag{3.21}$$

Similarly, the signal-response cross-correlation $C_{s\rho}(f)$ can be approximated using Eq. (3.5), and it follows

$$|C_{s\rho}(f)|^2 = |\langle\langle\tilde{\rho}(f)\rangle_n\tilde{s}(f)\rangle_s|^2 \approx |\langle\chi(f)\tilde{s}(\omega)\tilde{s}^*(f)\rangle_s|^2 = |\chi(f)|^2 S_{ss}^2(f). \tag{3.22}$$

Finally, inserting this in Eq. (1.24) for the linearly decodable information yields the small-signal approximation,

$$\mathcal{I}_{\mathrm{ld}}(f) \approx -\frac{1}{2}\log_2\left(1 - \frac{|\chi(f)|^2 S_{ss}^2(f)}{S_{ss}(f)C_{\mathrm{auto}}(f)}\right) = -\frac{1}{2}\log_2\left(1 - \frac{|\chi(f)|^2 S_{ss}(f)}{C_{\mathrm{auto}}(f)}\right). \tag{3.23}$$

Accordingly, in the linear regime at sufficiently small signals, $\mathcal{I}_{\mathrm{tot}}(f)$ and $\mathcal{I}_{\mathrm{ld}}(f)$ converge, as expected. I showed an equivalent relation for rate-modulated Poisson neurons in chapter 2.

3.7.2 Different types of nonlinearities are dominant in mean and variance coding

Here, I elaborate on the origins and differences of the previously described encoding nonlinearities in mean and variance coding and the two neuron models.Following the notion of λ^{ld}, the differences of the total information content $\mathcal{I}_{\mathrm{tot}}(\omega)$ and the linearly decodable information $\mathcal{I}_{\mathrm{ld}}(\omega)$ serve as a traceable measure of nonlinearity – the larger the differences, the more nonlinear the information transmission at a given frequency.

I identify two major types of nonlinearities that are most prominent in different signal modulations and neuron models. For LIF neurons, I find higher harmonics to be the most important nonlinearies, in particular, for mean modulations (figure 3.21A,C). Higher harmonics are contributions to the information at frequencies that are multiples of the central signal frequency and stem from the equivalent higher harmonics in the cross-correlation function (figure 3.21C; see also figure 3.17). At a given signal strength, higher harmonics are much stronger in MM. This matches the faster drop of λ^{ld} as a function of σ_s as visible in figure 3.19A. However, at a given information, higher harmonics can be expected to be equally present in VM, as figures 3.19C and 3.20A,C suggest.

The finite spike initiation time in EIF neurons gives rise to another type of nonlinearity at very low frequencies for signals with high central frequencies. Contributions at low

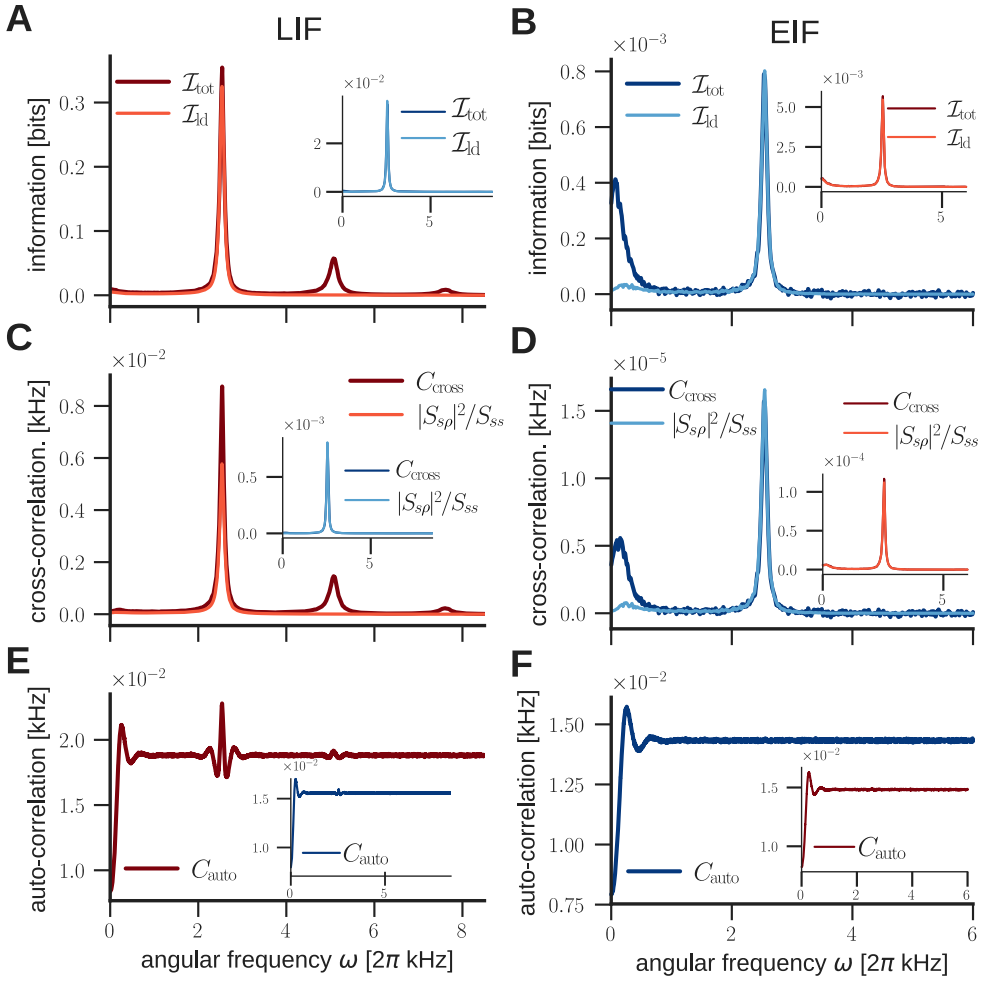


FIGURE 3.21: **Nonlinearities are observable as higher harmonics and as contributions at low frequencies.** Both columns show the information-relevant, frequency-resolved functions for a high central signal frequency $\Omega_0 = 2.54 \times 2\pi$kHz and $\sigma_s = 0.5$. The left column represents data from LIF neurons, the right column from EIF neurons. Main plots in the left column represent MM (red), and insets correspond to VM (blue); vice versa in the right column where insets refer to MM. **(A)** Nonlinearities manifest themselves in differences of the total information $\mathcal{I}_{\mathrm{tot}}(\omega)$ (dark colors) and its linear counterpart $\mathcal{I}_{\mathrm{ld}}(\omega)$ (light). Here, nonlinearities are higher harmonics of the signal frequency and are visible as local maxima at multiples of Ω_0 in mean coding. As shown in the inset, variance coding does not possess these nonlinearities at $\sigma_s = 0.5$. **(B)** In EIF neurons, instead of higher harmonics prominent nonlinearities are present at very low frequencies in the information for VM. Contributions at these frequencies are also present in the linearly decodable information but are much smaller. For mean modulations none such nonlinearities are observable (inset). (**C**,**D**) The shape of the information and its lower bound can well be traced back to the respective cross-correlation functions. Accordingly, nonlinear contributions are reflected in differences of $C_{\mathrm{cross}}(\omega)$ and $|S_{s\rho}(\omega)|^2/S_{ss}(\omega)$ [cf. Eq. (1.24)]. (**E**,**F**) The autocorrelation functions show clear signal-dependent deflections in (E) and are almost unchanged in comparison to the steady-state in (F). This is in agreement with the different sizes of the cross-correlation functions in (C) and (D) that reflect the impact of the signal. Autocorrelations do not have much influence on the information in the shown examples. Parameters are $\sigma_s = 0.5$, $\Omega_0 = 2.54 \times 2\pi$kHz, $\hat{\sigma}_n = \hat{\sigma}_n^{(2)}$, $\tau_n = 5$ ms and $\tau_s = 20$ ms.

frequencies can principally be understood in a linear picture. EIF neurons are low-pass filters whose response function is highest at low frequencies. Therefore, even if the signal power is relatively small at low frequencies, the response function can compensate for that and can produce contributions to the cross-correlation at frequencies close to 0. This gives an intuition for the origin of the small contributions to $\mathcal{I}_{\mathrm{ld}}$ and $|S_{s\rho}(\omega)|^2/S_{ss}(\omega)$ that occur in both modulations (figure 3.21B,D). However, for variance coding also contributions that are considerably higher than these linear ones are present (figure 3.21B). They are most likely due to higher order signal-response interactions, such as previously described couplings of similar frequencies that are each close to Ω_0 [275]. These nonlinear contributions at low frequencies to the cross-correlation function and, consequently, to the mutual information are barely observable in LIF neurons but explain the mentioned behavior of λ^{ld} for EIF neurons at high Ω_0 (see figures 3.19 and 3.20). Moreover, also higher harmonics can generally be observed in EIF neurons but are significant only when the total information is high.

For completeness, I also show the autocorrelation functions in all cases in figures 3.21E,F. The size of signal-dependent modulations of the autocorrelation depends on the size of the cross-correlation. The shape of the autocorrelation is not crucial for understanding the fundamental characteristics of information transmission.

3.8 Combined mean and variance modulation

I presented a comprehensive investigation of mean and variance modulations as information encoding strategies in the previous sections. Thereby, both of these strategies were considered separately. From introduction section 1.4.2, however, follows that a combined and simultaneous encoding via MM and VM could potentially enhance information transmission. This could happen in a nonlinear, synergetic manner, i.e. the combined effect of MM and VM may lead to higher information transmission than the sum of information transmission through MM and VM separately. I analyze the signal and noise parameter dependence of the information conveyed through combined encoding in this section. In particular, I compare combined encoding to the information in the separate channels under equivalent conditions.

3.8.1 Formalizing combined encoding

The signal encoding input current I_{com} in the combined MM-VM encoding is designed according to I_{MM} and I_{VM} [Eq. (3.1) and (3.2)], assuming that the signal $s(t)$ modulates both current mean and variance:

$$I_{\mathrm{com}}(t) = \mu(1 + s(t)) + \frac{\sigma_n}{\sqrt{2}}\sqrt{1 + s(t)}\tilde{\xi}_1(t) + \frac{\sigma_n}{\sqrt{2}}\tilde{\xi}_2(t). \tag{3.24}$$

Here, $\tilde{\xi}_1$ and $\tilde{\xi}_2$ denote two independent instances of the noise process (white or colored noise, as before) with unit variance[12]. This reproduces the MM and VM currents plus an additional independent noise term when in either case the other modulation in Eq. (3.24) is omitted. This is because in Eq. (3.24) the total noise is divided in a signal-independent part $\tilde{\xi}_1$, corresponding to MM, and a modulated part $\tilde{\xi}_2$, corresponding to VM. Considering the stationary state at $s(t) = 0$, it is

$$I_{\mathrm{com}}^{s=0}(t) = \mu + \frac{\sigma_n}{\sqrt{2}}\left(\tilde{\xi}_1(t) + \tilde{\xi}_2(t)\right) \widehat{=} I_0(t). \tag{3.25}$$

This is analog to the previously assumed state of spontaneous activity $I_0(t)$ (cf. section 3.1.1) because the sum of two independent Gaussian processes is another Gaussian process [204]. Thereby, the total variance of $I_{\mathrm{com}}^{s=0}(t)$ is σ_n^2. Complete equivalence would apply if $\tilde{\xi}_1 = \tilde{\xi}_2$. However, the choice here takes into account that the noise traces are not identical since the variance modulated and signal independent presynaptic populations are not identical. The mutual information $\mathcal{I}^{\mathrm{com}}$ for combined encoding is obtained through Eq. (1.23) using $I_{\mathrm{com}}(t)$ as input current.

3.8.2 A linear response theory for combined coding

I derive a linear response theory for the information transmission through combined encoding analogously to the theory in section 3.2. Again, the theory rests on the assumption that the current modulations are only small perturbations to the spontaneous

[12] Compared to Eq. (3.1) and (3.2), the noise variance here appears as a prefactor to emphasize the splitting of the total variance. The tilde over $\tilde{\xi}$ indicates this difference.

activity. The trial averaged firing rate in LRT is then given by[13] (cf. [107]):

$$\tilde{r}(\omega) = \langle \tilde{\rho}(\omega) \rangle_{\mathrm{n}} = 2\pi\nu_0\delta(\omega) + \chi_{\mathrm{MM}}(\omega)\tilde{s}(\omega) + \chi_{\mathrm{VM}}(\omega)\tilde{s}(\omega) \tag{3.26}$$

$$= 2\pi\nu_0\delta(\omega) + \left(\chi_{\mathrm{MM}}(\omega) + \chi_{\mathrm{VM}}(\omega)\right)\tilde{s}(\omega). \tag{3.27}$$

Comparing this with Eq. (3.5), it is apparent that the linear response theory for combined encoding can be obtained from that of the single channels through the replacement $|\chi(\omega)| \to |\chi_{\mathrm{MM}}(\omega) + \chi_{\mathrm{VM}}(\omega)|$. It follows:

$$C_{\mathrm{auto}}^{\mathrm{com,lin}}(\omega) = C_{\mathrm{auto}}^{0}(\omega) + |\chi_{\mathrm{MM}}(\omega) + \chi_{\mathrm{VM}}(\omega)|^2 S_{ss}(\omega), \tag{3.28}$$

$$C_{\mathrm{cross}}^{\mathrm{com,lin}}(\omega) = |\chi_{\mathrm{MM}}(\omega) + \chi_{\mathrm{VM}}(\omega)|^2 S_{ss}(\omega). \tag{3.29}$$

As a technical aside, in these expressions, the response functions have to be evaluated at σ_n rather than at $\sigma_n/\sqrt{2}$ [107]. Inserting the linearly approximated correlation functions in Eq. (1.23) yields the LRT approximation $\mathcal{I}_{\mathrm{lin}}^{\mathrm{com}}(\omega)$ for the information transmission about $s(t)$ via combined encoding:

$$\mathcal{I}_{\mathrm{lin}}^{\mathrm{com}}(\omega) = \frac{1}{2}\log_2\left(1 + \frac{|\chi_{\mathrm{MM}}(\omega) + \chi_{\mathrm{VM}}(\omega)|^2 S_{ss}(\omega)}{C_{\mathrm{auto}}^{0}(\omega)}\right), \tag{3.30}$$

$$\mathcal{I}_{\mathrm{com}}^{\mathrm{lin}} = \frac{1}{2\pi\nu_0}\int_0^\infty \mathrm{d}\omega\ \mathcal{I}_{\mathrm{lin}}^{\mathrm{com}}(\omega). \tag{3.31}$$

This equation, again, can be used to develop an understanding for the role of linear response functions in combined encoding. In Eq. (3.30), one can rewrite,

$$|\chi_{\mathrm{MM}}(\omega) + \chi_{\mathrm{VM}}(\omega)|^2 = |\chi_{\mathrm{MM}}(\omega)|^2 + |\chi_{\mathrm{VM}}(\omega)|^2 + 2\mathrm{Re}\left[\chi_{\mathrm{MM}}^*(\omega)\chi_{\mathrm{VM}}(\omega)\right]. \tag{3.32}$$

Generally, the same arguments as given in section 3.2 and illustrated in figure 3.2 are also valid for the linear response theory for combined coding. However, this time both response functions χ_{MM} and χ_{VM} simultaneously contribute to $\mathcal{I}_{\mathrm{lin}}^{\mathrm{com}}(\omega)$. Moreover – and importantly – the term $\mathrm{Re}(\chi_{\mathrm{MM}}^*(\omega)\chi_{\mathrm{VM}}(\omega))$ reflects the (first order) effect of interactions between the two signal modulation schemes. Mathematically, this term can be positive or negative, and interactions can thus generally facilitate or depress information transmission within the linear framework.

[13] Technically, this results from the distributive property of the convolution in Eq. (1.12), which in turn is due to the linearity of the convolution operation.

Parameter	(unit)	Simulated values
σ_s	[1]	[0.02, 0.06, 0.1, 0.15, 0.2, 0.3, 0.5, 0.75, 1, 2]
Ω_0	(2πkHz)	[0, 0.25, 0.51, 1 , 2.54, 7.11]
$\hat{\sigma}_n$	[1]	$\hat{\sigma}_n^{(2)}$
τ_n	(ms)	[0, 5]
τ_s	(ms)	20

TABLE 3.3: **Signal and noise parameters used in simulations of section 3.8.** The column *Simulated values* lists all values that were sampled in the simulations for the respective parameter. The parameters that were used to produce the figures are annotated in their captions. Varied parameters used in inset figures are annotated in the plot themselves. Each $\hat{\sigma}_n^{(m)}$ was chosen to realize a fixed firing rate ν_m as explained above and shown in table 3.1.

An interesting question related to the effect of MM-VM-interactions is whether the combined application of mean and variance modulation can have synergetic effects, i.e. whether $\mathcal{I}_{\text{com}}^{\text{lin}}$ is larger than the sum of the information that is transmitted in the channels separately (at the same overall noise level in both channels). According to Eq. (3.8), the sum of the mutual information in mean and variance encoding in linear response theory can be written as:

$$\mathcal{I}_{\text{lin}}^{\text{sum}}(\omega) = \mathcal{I}_{\text{lin}}^{\text{MM}}(\omega) + \mathcal{I}_{\text{lin}}^{\text{VM}}(\omega) \tag{3.33}$$

$$= \frac{1}{2}\log_2\left(1 + \frac{\left(|\chi_{\text{MM}}(\omega)|^2 + |\chi_{\text{VM}}(\omega)|^2\right) S_{ss}(\omega)}{C_{\text{auto}}^0(\omega)} + \frac{|\chi_{\text{MM}}(\omega)|^2|\chi_{\text{VM}}(\omega)|^2 S_{ss}^2(\omega)}{(C_{\text{auto}}^0(\omega))^2}\right). \tag{3.34}$$

Comparing the arguments of the logarithms in $\mathcal{I}_{\text{com}}^{\text{lin}}$ and $\mathcal{I}_{\text{sum}}^{\text{lin}}$ and applying Eq. (3.32), it follows that the effect of combined encoding is synergetic, i.e. $\mathcal{I}_{\text{com}}^{\text{lin}} > \mathcal{I}_{\text{sum}}^{\text{lin}}$, when

$$\frac{2\text{Re}\left[\chi_{\text{MM}}^*(\omega)\chi_{\text{VM}}(\omega)\right]}{|\chi_{\text{MM}}(\omega)|^2|\chi_{\text{VM}}(\omega)|^2} \cdot \frac{C_{\text{auto}}^0(\omega)}{S_{ss}(\omega)} \geq 1. \tag{3.35}$$

This always holds in the limit $\sigma_s \to 0$, and hence a synergetic effect of combined encoding can be expected in the linear regime (cf. figure 3.22).

3.8.3 Simulations and parameters

I ran the simulations as described in section 3.3.2 but with signal modulated currents according to Eq. (3.24) in the respective cases. Here, I only considered LIF neurons. The range of the sampled parameters is given in table 3.3.

3.8.4 Results for information transmission through combined encoding

Here, I investigate whether and how combined encoding can benefit information transmission in LIF neurons. In particular, I compare the cases where the same signal is encoded in the mean and variance channel separately and in parallel. Synergy and redundancy are measured in terms of a relative difference $d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}})$ between $\mathcal{I}_{\text{com}}$ and $\mathcal{I}_{\text{sum}}$, where $d(x, y)$ is defined as

$$d(x, y) = \frac{x - y}{1/2(|x| + |y|)}. \tag{3.36}$$

The distance $d(x, y)$ is conceptually very similar to the synergy Syn in Eq. (1.30) and has some additional, desirable properties: the sign of d indicates general synergy or redundancy, and $|d|$ represents the strength of the difference of x and y relative to the average amplitude of both quantities. The definition further prevents numerical instabilities when either x or y are zero or close to zero.

3.8.4.1 Combined encoding can enhance information transmission

Figure 3.22 shows an overview of the influence of τ_n, σ_s and Ω_0 on combined coding information transmission. Figures 3.22A,C are similar to figures 3.13A and 3.18A (color coding as before) but additionally show curves of $\mathcal{I}_{\text{com}}$ and its linear approximation $\mathcal{I}_{\text{com}}^{\text{lin}}$ in green. The relative difference $d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}})$ can be seen in figures 3.22B,D. Figure 3.23 provides another visualization of the influence of signal and noise parameters on $d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}})$. It shows positive values of d, i.e. cases where combined coding creates an information-theoretic benefit, as green bubbles and negative values as purple bubbles.

Overall, both enhancement and reduction of the mutual information through combined encoding are possible. Thereby, an important result is that lower signal strengths and larger central signal frequencies benefit the information transmission through combined encoding, and vice versa. These two relations seem to apply independently and the relation between these parameters and the relative difference is monotonic (see figures 3.22B,D). The differences between combined and parallel coding are more prominent at white noise and decrease for finite τ_n, however, the noise correlation time does not influence whether synergetic or redundant effects dominate.

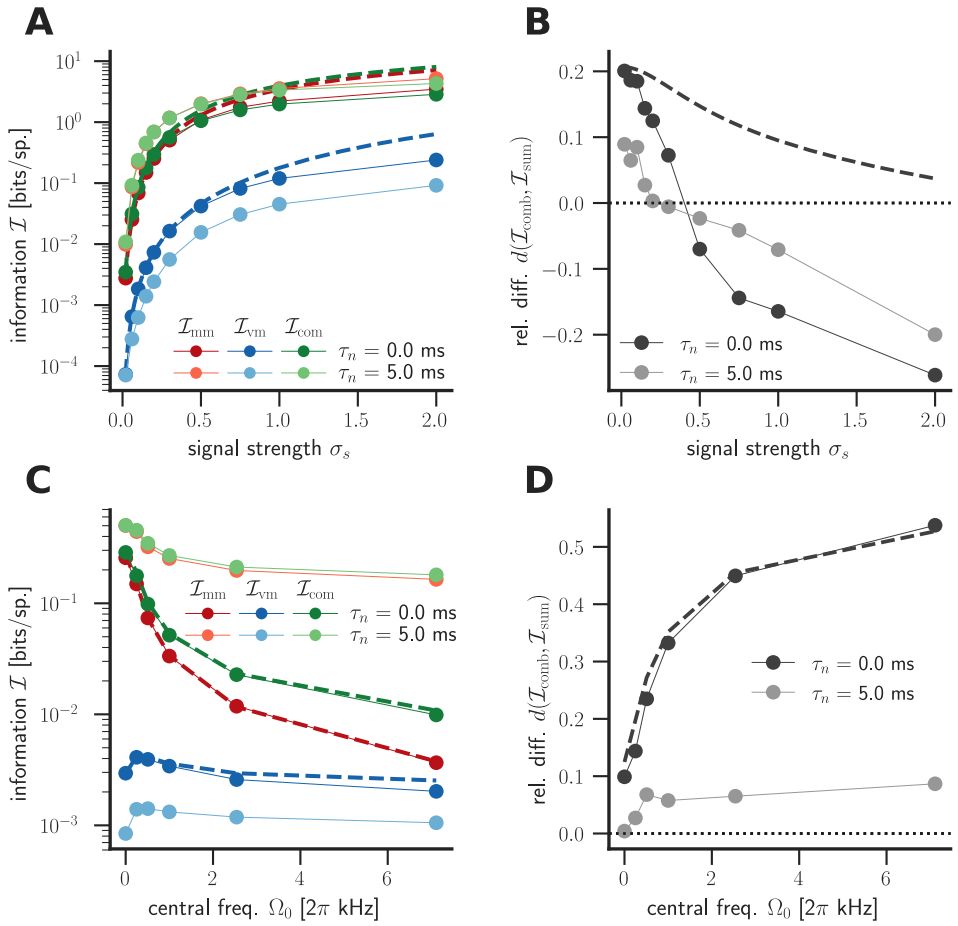


FIGURE 3.22: **Information transmission through combined encoding in leaky integrate-and-fire neurons.** MM and VM are colored red and blue, combined encoding is shown in green; circles denote exact information and dashed lines linear approximations thereof (the same applies for subsequent figures). **(A)** All information curves increase monotonically with σ_s (see also figure 3.24). **(B)** The relative difference $d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}})$ between combined encoding and summed separate encoding decreases with increasing signal strength. The dotted line marking $d = 0$ is crossed from above around $\sigma_s = 0.5$, which implies a synergetic advantage of combined encoding at lower signal strengths. **(C)** The difference between $\mathcal{I}_{\text{comb}}$ and $\mathcal{I}_{\text{MM}}$ increases with larger central signal frequencies Ω_0 even though $\mathcal{I}_{\text{VM}}$ remains almost constant. Therefore, $d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}})$ is a monotonically increasing function of Ω_0 (shown in **(D)**). At white noise (dark colors), the differences between $\mathcal{I}_{\text{comb}}$ and $\mathcal{I}_{\text{sum}}$ are more pronounced compared to $\tau_n = 5$ ms (light colors). The noise correlation time does not seem to influence the sign of $d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}})$. Curves for $\mathcal{I}_{\text{sum}}$ are not shown in (A) and (B) due to their high overlap with the curves for $\mathcal{I}_{\text{MM}}$ in many cases. Parameters are $\sigma_s = 0.15$ in (C) and (D); $\Omega_0 = 0.51 \times 2\pi$kHz in (A) and (B).

The linear theory $\mathcal{I}_{\text{com}}^{\text{lin}}$ is an expectedly good approximation for small σ_s but largely overestimates the difference $d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}})$ at moderate to large signal strengths (see figure 3.22B). Importantly, the linear theory always predicts positive values of d and is, therefore, not suitable to estimate the synergetic or redundant effect of combined encoding outside the linear regime.

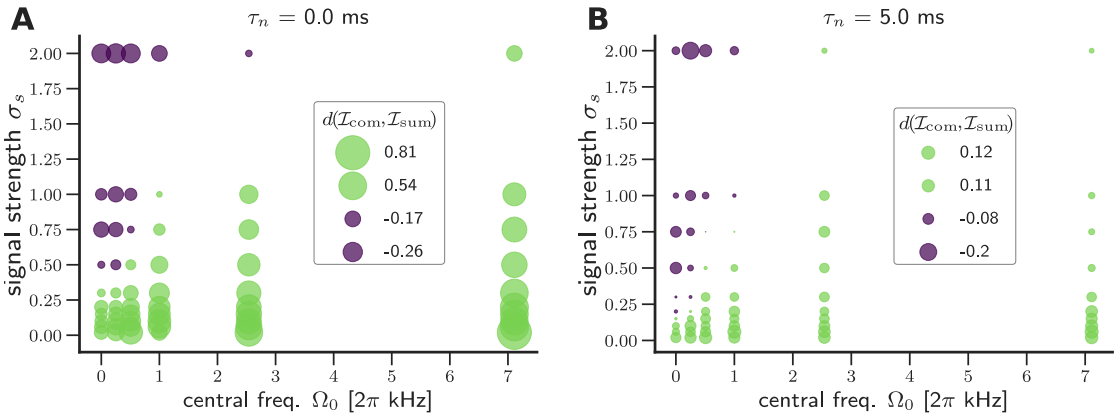


FIGURE 3.23: **Regions of synergetic and redundant effects in combined encoding as bubble plot.** For each combination of σ_s and Ω_0, a colored bubble represents the corresponding value of the relative difference $d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}})$. Green bubbles are cases where $d > 0$ (synergetic combined encoding) and purple bubbles where $d < 0$. The size of the bubbles represents the absolute value of d. In **(A)**, the noise correlation time is $\tau_n = 0$, and in **(B)** it is $\tau_n = 5$ ms. The visualization conforms the results of figure 3.22: smaller σ_s and larger Ω_0 promote information transmission through combined encoding.

3.8.4.2 Regions of maximal information gain

The applied methods have taken into account the metabolic cost of action potentials by expressing the mutual information in bits per spike (cf. discussion of efficient coding in section 1.3.2). Given that the signal strength σ_s represents the modulation depth of the synaptic currents it is reasonable to assume that larger signal strengths correspond to higher presynaptic spiking and hence to a higher energetic cost on the presynaptic side. Even though it is not obvious how this cost scales with σ_s, an investigation of the slope of $\mathcal{I}(\sigma_s)$ can provide information about regions where increases in the signal strength yield the maximum gain in information transmission.

Close-ups of figure 3.22A with linear scale and different y-axis scales are shown in figure 3.24. All information curves (MM, VM, combined) are sigmoidal, i.e. supralinear at small σ_s, followed by a roughly linear and then a sublinear curvature. In the supralinear region, the linear approximations fit the true information values well[14], and from Eq. (3.6) and (3.10) follows that in this regime the information is a quadratic function of σ_s, which is confirmed by figure 3.24.

[14]Remember that in this approximation linearity refers to a linear approximation of the firing rate and not of the mutual information itself.

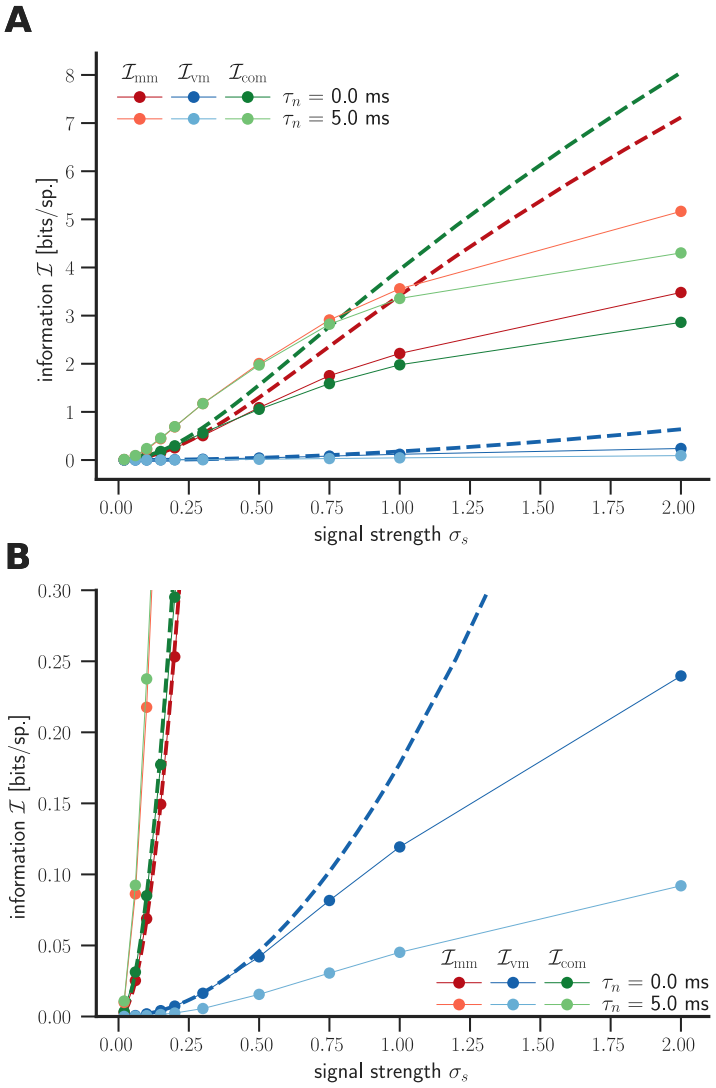


FIGURE 3.24: **Sigmoidal shapes of information curves imply regions of maximum information gain.** The figure shows close-ups of figure 3.22A with linear scale and y-axis ranges optimized to investigate $\mathcal{I}_{\text{MM}}$ and $\mathcal{I}_{\text{com}}$ in **(A)** and and $\mathcal{I}_{\text{VM}}$ in **(B)**. All curves of the exact information (dots) have sigmoidal shapes as a function of σ_s. This implies regions of a maximal slope, where increases in σ_s are accompanied by maximal information gain. These regions may thus be most relevant for biological optimization in terms of energy constraints (see text). The linear approximations (dashed lines) correctly predict the initial quadratic slope of the curves but become increasingly worse fits at larger σ_s. Here, it is $\Omega_0 = 0.25 \times 2\pi\text{kHz}$.

Further, I conclude that there is a region of maximal information gain where the quadratic behavior starts transitioning to linear, and the slope is maximal. Interestingly, comparing figures 3.22B and 3.24, these transition points for all informations (MM, VM, combined) and the point where the advantage of combined encoding over separate encoding vanishes (where $d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}})$ switches sign), are very similar; for $\tau_n = 0$ it is approximately between $\sigma_s = 0.3$ and $\sigma_s = 0.5$, for $\tau_n = 5$ ms between $\sigma_s = 0.25$ and $\sigma_s = 0.3$. However, the transitions is weaker and occurs slightly later for

VM (figure 3.24B). The coincidence of the transition points is in agreement with the intuition that the onset of saturation of information transmission is not only dependent on σ_s but generally occurs at higher information values. When the mutual information in mean and combined coding starts to saturate this is less so the case in variance coding, and hence $\mathcal{I}_{\text{sum}}$ keeps to increase stronger with σ_s than $\mathcal{I}_{\text{com}}$, giving rise to the aforementioned transition of $d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}})$.

3.8.4.3 Nonlinear effects can lead to synergetic advantages of combined encoding

In section 3.8.4.1, I argued that combined encoding can have synergetic effects on information transmission in the sense of a supralinear addition of the information found in the MM and VM channels if used separately. This phenomenon can best be understood by considering the frequency-resolved information and correlation functions. Figure 3.25 shows these functions for a synergetic case $[d(\mathcal{I}_{\text{comb}}, \mathcal{I}_{\text{sum}}) > 0]$ at $\sigma_s = 0.1$. Figures 3.25C,D reveal two mechanisms by which combined encoding can be enhanced in its information transmission capabilities. First, a supralinear addition of the information at the main maximum at Ω_0 is observable (figure 3.25D). The effect agrees with the linear theories $\mathcal{I}_{\text{lin}}^{\text{com}}(\omega)$ and $\mathcal{I}_{\text{lin}}^{\text{sum}}(\omega)$ and reflects the frequency coupling that is expressed by the term $\text{Re}\,[\chi_{\text{MM}}^{*}(\omega)\chi_{\text{VM}}(\omega)]$, which enters the cross-correlation Eq. (3.29) through Eq. (3.32).

Second, also nonlinear frequency coupling can have information increasing effects around the second order maximum at $2\Omega_0$. There, nonlinear input-output relations induce contributions to the information at higher order maxima (see section 3.7 for a more detailed discussion). This contribution is considerably larger for combined encoding (see inset of figure 3.25D). This is a particularly interesting effect because here $\mathcal{I}_{\text{VM}}$ does not itself show a contribution in this region (inset for figure 3.25C), and VM is hence not suitable to transmit information at these frequencies. Nonetheless, the presence of variance modulated input currents can positively influence the overall signal encoding in combination with mean modulation. This result is in agreement with previous studies on gain modulation [68] or enhancement of neural encoding through covarying fluctuations [256]. My finding provides an explanation at the level of the exact, frequency-resolved mutual information.

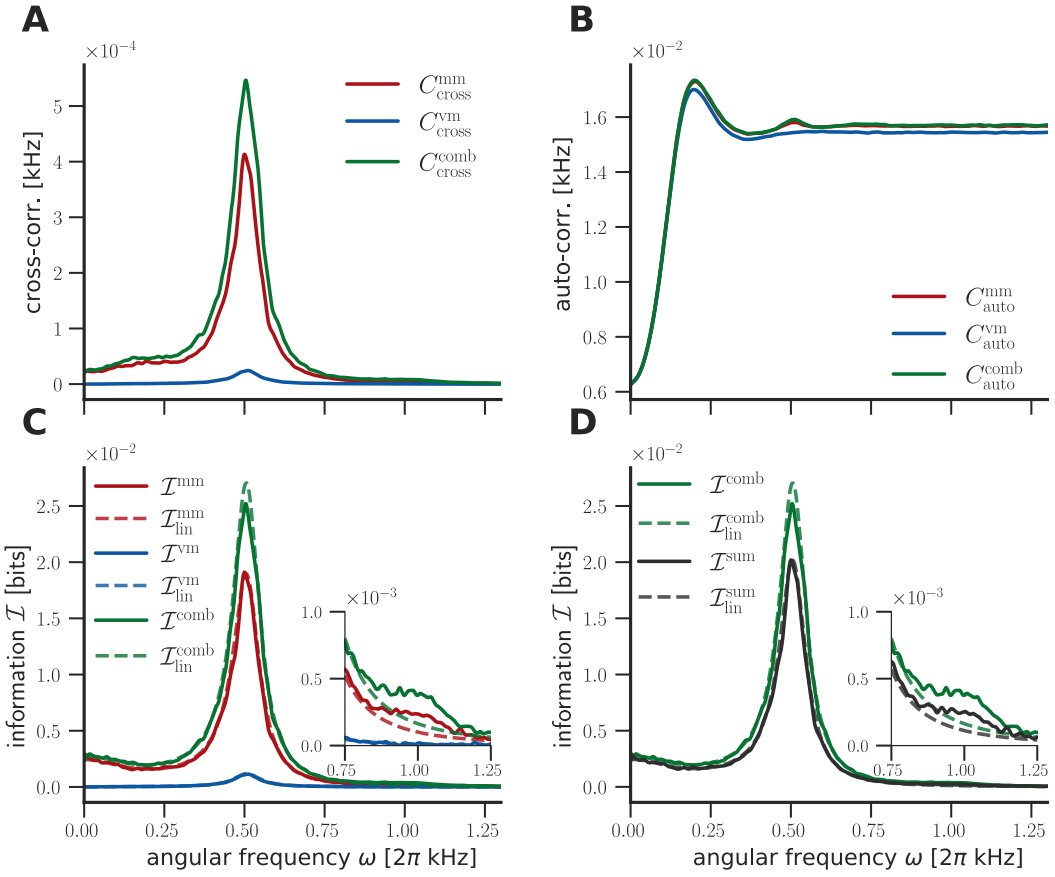


FIGURE 3.25: **Linear and nonlinear frequency coupling can promote combined encoding.** All functions are shown as functions of frequency. **(A)** Cross-correlation functions for MM, VM and combined encoding. Around $\Omega_0 = 0.5 \times 2\pi$Hz, a supralinear addition of MM and VM in combined encoding is observable by comparing the sum of the red and blue traces to the green trace. **(B)** All autocorrelation functions are similar and hence do not contribute to differences in information transmission. **(C)** Frequency-resolved information and their linear counterparts for MM, VM, and combined modulation. The inset shows a zoom at the region around $2\Omega_0$ where second order maxima in both $\mathcal{I}_{\text{comb}}$ and $\mathcal{I}_{\text{MM}}$ are visible. **(D)** Plot of $\mathcal{I}_{\text{comb}}$ and $\mathcal{I}_{\text{sum}}$ for direct comparison. The frequency-resolved information curves in (C) and (D) confirm supralinear addition in combined encoding at $\omega = \Omega_0$, which is also predicted within the linear theory $\mathcal{I}^{\text{lin}}$ and reflects frequency coupling, which is expressed in the term $\text{Re}\,[\chi^*_{\text{MM}}(\omega)\chi_{\text{VM}}(\omega)]$. Moreover, nonlinear frequency coupling can be observed around the second order maximum. Even though VM does not show any response in this region, the combined modulation of mean and variance currents enhances information transmission compared to pure mean encoding as can be seen in the insets. Parameters are $\sigma_s = 0.1$, $\Omega_0 = 0.51 \times 2\pi$kHz and $\tau_n = 0$.

Figures 3.25A,B show the cross- and autocorrelation functions for completeness. Around the linear regime, the cross-correlation functions mirror the properties of the information well. The autocorrelation functions are similar in all cases (MM, VM, combined) and do not considerably contribute to differences in information transmission.

3.9 Summary and discussion

In this chapter, I studied the mutual information of spike trains and mean and variance modulating signals. Modulations of the mean somatic current, due to a transient break in the excitation-inhibition balance, have been reported several areas, among which are auditory, visual and barrel cortex [70, 80, 249, 250, 319]. Similarly, experimental observations revealed that simultaneous modulation in the excitatory and inhibitory activity [71–73, 75, 80, 256] can induce signal-dependent changes in the strength of the somatic fluctuations. The encoding of sensory or internal signals via mean or variance modulations has been addressed experimentally [85, 115] and theoretically in the linear regime [107, 114, 261]. These studies pointed to the possibility that variance modulation may be faster or more efficient in signal encoding if it occurred on a white noise background. However, no comparative evaluation of the information transmission properties of mean and variance coding has existed beyond the linear regime. While the experimental level has been the focus of many studies, a number of mechanistic and computational questions related to how mean and variance modulating signals are represented at the spike level have remained open. In particular: *How many bits per spike are transmitted by mean and variance modulating signals? Does it make a difference whether a particular signal is encoded in the mean or variance? Is the information encoding linear?*

I tackled these questions by calculating the exact mutual information in spike trains about mean and variance modulating signals in LIF and EIF neurons in the fluctuation-driven, subthreshold regime. I used a recently proposed method [98] to calculate the exact information content in spike trains per spike as well as a function of frequency. Additionally, a linear response theory for the information and a linearly decodable lower bound for the information were considered in comparison. I generalized the findings by exploring many "what if"-scenarios by varying the firing rate, noise correlation statistics, signal strengths, and spiking mechanisms.

Moreover, I analyzed the statistical properties of the spike train Fourier coefficients in both modulation schemes, linked different information-theoretic measures, and analyzed the properties of signal encoding via combinations of mean and variance modulation. In the following, I summarize and discuss the results of this chapter.

3.9.1 Properties of spike train Fourier coefficients and information upper bound

Regarding section 3.4 on the statistical properties of spike train Fourier coefficients and the related section 3.5 on the (putative) upper bound I found:

- The Fourier coefficients of the spike trains resulting from both mean and variance modulated input currents are multivariate normal and independent. This means that the real and imaginary parts of these coefficients at all frequencies are Gaussian distributed across trials (for both repeated and varying signals) and are independent across frequencies (see figures 3.4 to 3.10).
- The independence and normality of the Fourier coefficients also hold for non-Gaussian input statistics such as the variance modulated input currents as long as the input is stationary (section 3.4.2 and figure 3.11).
- If a signal is presented repeatedly the means of the Fourier coefficients of the resulting spike trains encode the signal (section 3.4.3 and figure 3.12). This result can be expected from linear response theory and but is generalized here.
- In section 3.5, I show that a previously believed upper bound approximation of the mutual information is equivalent to the full information.

The independent Gaussianity of the Fourier coefficients is at the core of the correlation method [98]. While the property has been proven mathematically for stationary inputs [302], numerical evidence has been lacking, especially for variance modulated currents that are themselves not Gaussian. This supports the general applicability of the correlation method beyond the conventional setting with Gaussian inputs.

Here, the numerical evidence for multivariate, uncorrelated normality has been put forth by statistical tests in combination with sampling of "ground truth" data (termed surrogate or reference data). By nature, these methods cannot unequivocally prove statistical properties but rely on somewhat arbitrary choices of confidence intervals and criteria (a critique on p-values for hypothesis evaluation can be found in [320]). The results of section 3.4 are hence not technical proofs although they seem to be statistically solid. A particular limitation of testing for multivariate normality was given by the fact that, in principle, this property needs to be shown for all Fourier coefficients at once. This is

practically unfeasible because the number of data samples that is required for an informative statistical analysis scales exponentially with the number of Fourier coefficients which here is in the order of 10^5. The chosen analysis hence relies on a plausibility assumption: if there are dependencies among the Fourier coefficients they can be expected to be present among all coefficients and particularly visible among the coefficients at the most dominant frequencies. Based on this argument, overall multivariate normality was inferred from the multivariate normality of subsampled sets of Fourier coefficients (cf. section 3.4.1.2), and independence from the overall uncorrelatedness (section 3.4.1). The analysis was done for LIF neurons only but generalizes to other spiking neuron models with finite memory and stationary parameters, including EIF neurons [302].

In section 3.4.3, I concluded that in the given framework, the mean Fourier coefficients carry signal information. This statement can be refined to that all information about the signal is encoded in these means. As the distributions of the coefficients are Gaussians they are sufficiently determined by mean and variance, and all higher moments are functions thereof. The variances of the coefficients determine the mutual information and for trials with repeated, fixed signal describe the noise level. The means, on the other hand, are specific to a given signal, which they thus encode. For example, in the hypothetical noise-free case, all spike trains would have the same Fourier coefficients for trials with the same signal. These coefficients, also being the means, would then carry all signal information. Consequently, even though the mutual information is determined by the variances, the features that encode this information are the means. In turn, and somewhat unintuitively, the means do not enter the mutual information because the entropy of a Gaussian distribution only depends on its variance. This reflects that the variances of the Fourier coefficients mirror the overall spread of the mean values (dynamic range) across signals. Nonetheless, designing an optimal stimulus decoder may incorporate both these moments of the response distribution (see section 4.6 for a more thorough discussion).

The identity of the (putative) upper bound and the full information (correlation method) is a novel finding based on and closely linked to the specific role of the Fourier coefficients' mean and variance. In other words, the properties of the Fourier coefficients that were previously considered mere assumptions (for general encoding features [218, 224, 226]) naturally emerge under stationary inputs. The identity is not only remarkable in itself

but also allows for a re-evaluation of the conclusions that were drawn in previous studies that considered the upper bound only an approximation, e.g. in [218, 225, 317, 321].

Overall, my findings shed light on basic mechanisms of neural encoding and support a better understanding of the connections between signal encoding and information transmission in stationary situations. The analyses in sections 3.4 and 3.5 provide supportive evidence for the validity and generality of the correlation method[15] by means of a numerical confirmation of independent, multivariate Gaussianity and consistency of the method with the thought-to-be upper bound.

3.9.2 Information transmission through mean and variance modulations

The main results regarding the information transmission of mean and variance encoding are (cf. section 3.6):

- Mean coded signals have a larger information capacity than variance coded signals in almost all considered cases. This is always the case when noise has temporal correlations (see figures 3.13, 3.18, and 3.14). Only if white background noise and signals with very high frequencies are assumed, I find an advantage of variance modulation in information transmission (for LIF neurons).
- Analyzing the EIF spiking model, an advantage of mean coding over variance coding regardless of signal and noise parameters can be observed (see figure 3.14). Thereby, a finite spike initiation time, as incorporated in EIF models [35], limits the information containing frequency bandwidth in both modulation schemes, in agreement with linear response studies [36, 108].

The parameter dependence of the mutual information in the two coding schemes is summarized as follows (see figures 3.13 and 3.14):

- Information increases monotonically as the signal strength σ_s grows, for both mean and variance coding.
- Higher noise correlation times τ_n increase the information content in MM but decrease it for VM coding.

[15] The peer-review process of [3] had revealed sporadic reservations towards the correlation method and its underlying assumptions despite the clear evidence brought up in the original publication [98].

- The signal correlation time τ_s plays only a minor role for determining the information content in integrate-and-fire neurons.
- Increasing the noise amplitude σ_n – here equivalent to increasing the firing rate (see above) – while keeping the signal strength constant reduces the information content for both modulations.
- The information ratio $\beta = \mathcal{I}_{\mathrm{MM}}/\mathcal{I}_{\mathrm{VM}}$ increases with lower signal strengths, lower noise/firing rate levels, and longer noise correlation times.

It is important to note that the described results are generally not limited to the linear regime but are valid across all considered signal strengths. Even though many of the presented results are in line with what can be expected from known linear response functions, the results show in how far these response properties translate into the (complete) mutual information.

The information-theoretic benefit of VM over MM for fast fluctuating inputs in LIF neurons at white noise background is in agreement with the respective response functions of LIF neurons [107] and with the findings of Silberberg *et al.* [85]. Here, this result was shown to also hold beyond linearity assumptions. However, white noise is not biologically plausible because the biophysical processes causing the noise currents result from postsynaptic potentials that have finite rise and decay times [260]. I conclude that under realistic conditions and with biological neuron models, the information transmission through mean modulations is higher than through variance modulations. Consequently, in terms of mutual information the previously described (near) instantaneous signal transmission through variance modulations [85, 107, 114, 261] can be seen as an artifact of the white noise background.

Regarding the signal and noise parameters, experiments revealed that the statistics of input currents, both background- and signal-dependent, vary considerably across brain regions and depend on the brain state [322]. This indicates that the range of time constants and other parameters that I considered is likely to be realized in either different brain states or brain regions. Differences in information content may be indicative of the different computational capabilities in these brain regions. For example, the results in figures 3.18A,B indicate the existence of local optima for variance coding for specific signal frequencies. Since the location of these optima depends on neuronal spiking mechanism and background noise, they could be targeted by a brain region by enhancing

those synaptic time constants from the available pool that best match the properties of the signal. Moreover, the local maxima suggest that MM and VM encoding best operate on different signal frequency bands. If these bands are sufficiently well separated, this will allow for an encoding of two signals through MM and VM in parallel (see section 4.7.2).

The observed opposite effect of the noise correlation time τ_n on information transmission in MM and VM is rather surprising, and in terms of optimization implies a minimization or maximization of the synaptic time constants [323, 324], respectively (cf. figures 3.13 and 3.14). The effect cannot easily be understood in terms of response functions because these are not available in closed form for finite noise correlation times. However, interesting insights can be obtained by investigating the power spectra of $I_{\mathrm{MM}}(t)$ and $I_{\mathrm{VM}}(t)$. From Eq. (3.1), (3.2) and (3.11) follows for the power spectra of mean and variance modulated input currents:

$$S_{II}^{\mathrm{MM}}(f) = S_{nn}(f) + S_{ss}(f), \tag{3.37}$$

$$S_{II}^{\mathrm{VM}}(f) = S_{nn}(f)\left(1 + S_{ss}(f) + \mathcal{O}(s^4)\right), \tag{3.38}$$

with the signal and noise power spectra S_{ss} and S_{nn} (cf. section 3.1.2). These equations imply, loosely speaking, that in MM, the noise power is added to the signal power and hence has an interfering effect. For VM, on the other hand, noise is crucial as the carrier of the signal, and signal frequencies that are not present in the noise cannot be represented in the VM currents (signal power multiplicatively modulates noise power). Therefore, larger noise correlation times – i.e., noise power spectra with a smaller spread – reduce the interference of signal and noise in MM but provide less power across the band of signal-carrying frequencies in VM, agreeing with the described effect of τ_n. However, the described mechanism is rather heuristic and does not consider the highly nonlinear input-output transfer according to which neurons emit spikes.

The spiking model, signal and noise parameters in this study were chosen such that the irregular spiking statistics reproduces features of cortical recordings [44, 50, 83, 296, 297], and the selected exponential and leaky integrate-and-fire models have been proven to be fairly accurate for cortical neurons [32, 35, 38]. Within this framework, I recovered similar information values as in previous experimental studies [25, 218] ranging from $0.1 - 10$ bits/spike. This supports that the predicted information content is consistent with the *in vivo* and *in vitro* situations encountered in experiments. Nonetheless, the

used integrate-and-fire models are abstractions to the complex electrophysiology of real neurons, and the signal and noise parameters in *in vivo* conditions are not generally known. I discuss these limitations and possible extensions of the work in chapter 4.

3.9.3 Linear and nonlinear components of information encoding

I presented the analysis of linearly and nonlinearly encoded information in sections 3.6.1 and 3.7. I summarize and discuss the results here.

- In section 3.2, I developed a linear approximation $\mathcal{I}^{\text{lin}}$ for the mutual information, which is based on approximating the correlation functions under the assumption of a linear transfer function (asymptotically exact only for the cross-correlation).
- Whereas $\mathcal{I}^{\text{lin}}$ provides a good intuition for the role of neural response functions for the mutual information (see sections 3.2 and 3.8.2), violations of the linearity assumption at higher signal strengths lead to considerable prediction errors of $\mathcal{I}^{\text{lin}}$ (see figure 3.13). The most crucial approximation errors result from misestimations of the firing rate, unconstrained response amplitudes within linear response theory, and the occurrence of higher harmonics (see section 3.6.1 and figure 3.17).
- In section 3.7, I assessed the fraction of linearly encoded information by contrasting the full information $\mathcal{I}^{\text{tot}}$ with the linearly decodable information $\mathcal{I}^{\text{ld}}$ (different from $\mathcal{I}^{\text{lin}}$, among other things, by not approximation the autocorrelation). Using this approach, I found that the fraction of linearly encoded information is generally smaller for higher information content. Surprisingly, this generally holds regardless of the signal strength.
- However, the relation can be offset by nonlinearities at very low frequencies that occur in exponential integrate-and-fire neurons and variance coding. As another important type of nonlinear encoding, I identify higher harmonics in the spike cross-correlation functions (figures 3.19, 3.20 and 3.21).
- The linearly decodable information, alias information lower bound, is analytically shown to converge to the full information at small signal strengths in section 3.7.1.1.

Linear theories, including these for estimating the information in spike trains, have been very popular [1], and they can be useful in providing a tractable and intuitive picture

of the relevant mechanisms in neural coding as shown in this chapter. However, they are limited as nonlinear effects are known to be generally important (see also discussion section 4.2). The investigation of the linearly decodable portion of the transmitted information thus closes a critical gap left by previous studies by quantifying the mutual information that cannot be captured by linear approximations (see e.g. [25, 113, 129, 206, 218, 222, 223, 256]). The results thereby emphasize the importance of including nonlinear features in models of neural coding since they can be significantly relevant for information transmission. They also encourages a reevaluation of the results and conclusions that were obtained using linear estimations of the information only.

Previous attempts to estimate the nonlinear contributions to neural coding used a linearity index defined as the ratio of the signal-response coherence and the square root of the response-response coherence [225, 317, 321]. Applying this method indicated that an increasing signal strength leads to stronger nonlinear contributions [219, 225]. In agreement with these and other studies [288, 318], I have observed a linearizing effect of increased noise or firing rate in both coding channels (cf. figure 3.20).

Regarding the encoding linearity in both modulation schemes, variance coding, in contrast to mean coding, remains largely linear even for high signal strengths with an linearity index above 70% (see figure 3.19). This supports variance encoding as the means of signal transmission if only linear decoders are available. On the other hand, this is opposed by the fact that in many cases, the total linearly decodable information in MM encoding is larger even though its fraction of the total information is smaller than in VM encoding.

Another difference between both the coding schemes and neuron models lies in the kind of nonlinearities that they predominantly feature. I identified higher harmonics as the major source of nonlinearities for LIF neurons. They can be observed in mean coding already at rather small signal strengths. In EIF neurons, the most prominent nonlinearities are introduced at low frequencies when the signal is centered around high frequencies. This is best seen for variance modulation (figure 3.21). The differences of LIF and EIF models in producing nonlinearities (in the cross-correlation functions) are based on the spike initiation mechanism. The spike initiation time not only influences the linear response function but will also impact higher order response functions, especially at high frequencies. As a result, the finite spike initiation time of neurons can facilitate

the appearance of low-frequency power in the cross-correlation function, or equivalently PSTH, that stems from high frequency variance modulations. Hence, high frequency signals can have an effect outside the signals original frequency range and, in particular, modify the encoding at frequencies that are not in the range of the signal itself. This effect could be relevant for the investigation of neural signals that possess low-frequency power with elusive origins. [The role of nonlinearities with in combined encoding is discussed in the next section.]

3.9.4 Combined mean and variance encoding

The main results as for combined mean and variance modulation (section 3.8) are:

- I developed a linear response approximation for combined encoding at small signal strengths (sections 3.8.2 and 3.8.4.1).
- Combined encoding can have synergetic benefits. The regime of synergy is generally facilitated by small signal strengths and large central signal frequencies. The linear theory always predicts synergetic encoding and does not capture the redundant regime (section 3.8.4.1 and figures 3.22, 3.23).
- Mutual information curves are sigmoidal functions of the signal strength, which can be considered a proxy for presynaptic energy consumption through firing. Accordingly, there exist regions of maximum (and supralinear) information gain with respect to the signal strength (section 3.8.4.2 and figure 3.24), which may be important for energetic optimization.
- I identified nonlinear coupling of mean and variance modulations as the source of synergy in combined encoding. In particular, the presence of variance modulations can enhance information transmission at frequencies that are not accessible for VM itself (section 3.8.4.3 and figure 3.25).

The potentially synergetic effect of combined encoding has important implications for neural systems as they may exploit this mechanism for more efficient information transmission. In particular, nonlinear coupling between mean and variance changes renders a surprising candidate mechanism for optimized neural encoding which had not previously been discovered for information transmission. It remains open how this coupling is sustained or modified in more detailed neuron models and *in vivo* situations (see also

section 4.7). Response properties of neurons are commonly assessed for either of the input channels [85, 107, 108, 114, 115], and it seems particularly interesting to evaluate whether simultaneous activation of both modulations changes the dynamic properties of a neuron in a way that is not just a "superposition" of the known properties.

Furthermore, as proposed in section 4.7.3, *in vivo* experiments could help to estimate the modulation depth in both channels if they operate in parallel. The results could then be fed back into my analysis of combined encoding where I assumed identical modulation depths. Similarly, I have assumed that both channels operate at noise backgrounds with equal statistics and power. Changing the dynamic properties and strength of the noise for each channel independently could reveal additional mechanisms for efficient signal transmission.

Similar (synergetic) effects of neural coding enhancement have been proposed previously. For example, stimulus-dependent voltage fluctuations have been found to facilitate neural response gain modulations when they co-occur with changes in the mean membrane voltage [251]. Along similar lines, other studies investigated the influence of signal-modulated γ-fluctuations of the membrane voltage and discovered that those variance modulations help to increase information transmission compared to mean modulations only [255, 256]. These studies used more detailed biological models but are limited to the consideration of γ-activity, whereas in the present study I probed several stimulus parameters. These studies mostly addressed non-stationary, oscillatory stimuli. Furthermore, in [255, 256] only signal-response coherences and the lower bound $\mathcal{I}^{\mathrm{ld}}$ are considered as approximations for the mutual information. It is also noteworthy that the mentioned studies consider mean and variance changes of the membrane voltage rather than changes in the input current. Nonetheless, these quantities are closely related (the subthreshold voltage is a frequency filtered version of the input current; see section 1.1.4) and the results are transferable.

My results are different from findings according to which an increased input noise baseline can lead to linearized or amplified response functions and can thus serve an important computational role [40, 68, 225, 288, 318]. These noise levels result from signal-independent background activity, whereas combined encoding here refers to simultaneous, signal-dependent modulations of the current mean and variance.

The synergy in combined encoding, where existent, implies a higher encoding efficiency in terms of the information that can be transmitted per spike and hence per energy on the postsynaptic side (considering the energetic cost of spikes). However, it is not obvious whether the input currents for combined encoding and the MM and VM channels in sum involve identical energetic costs on the presynaptic site. This would have to be evaluated for a complete energetic comparison of the two coding scenarios and involves translating either current into presynaptic activity (cf. Appendix A.2). Similarly, it is unclear how changes in the modulation depth influence the energetic requirements. Furthermore, it remains to analyze whether the given scenario in which each channel (MM, VM, combined) possesses the same total noise power guarantees complete comparability of summed and combined encoding because the sum of noise in MM and VM is numerically higher.

Chapter 4

General discussion and outlook

In this chapter, I discuss the methods, results, limitations and achievements of this dissertation. I, thereby, address issues on a rather general level as I provided detailed, context-specific discussions of the results at the end of the respective chapters.

In section 4.1, I give a concise summary of the major results, discuss their meaning in a broader context, and mention limitations of their applicability. In sections 4.2–4.6, I discuss the issues of nonlinearities in neural encoding, the information-theoretic model of this dissertation, the assumption of independent neurons, the pros and cons of integrate-and-fire models, and different perspectives on the concept of the neural code. Finally, section 4.7 is devoted to short proposals of research projects that complement the presented study.

4.1 Understanding neural information transmission

Here, I review and discuss my work in a broader context, unifying the different parts of this work and at times referring to the "big picture" of neural coding. On this level, my work can be recapitulated as follows:

- The overall **goal** was to gain a better understanding of neural information transmission (chapter 1).
- The **methodological starting point** were the correlation method to compute the mutual information and its underlying assumptions (see section 4.3).
- The method was **applied to** information encoding through

... firing rate dynamics (chapter 2).

... mean modulation of the somatic current (chapter 3).

... variance modulation of the somatic current (chapter 3).

- Using a combination of analytical calculations and neuron simulations, I **found that**
  - ... spikes carry more information if they are temporally uncorrelated. I uncovered the underlying mechanisms for this, thereby providing a method to determine the effect of correlations on information transmission (chapter 2).
  - ... an expression for the information that accounts for rate correlations and only relies on the PSTH can be derived from the correlation method (section 2.1).
  - ... different popular methods to estimate the mutual information of stimulus and spikes can be reconciled and mapped to each other. However, I also found inconsistencies that suggest that the underlying assumptions may have to be revisited in some cases (sections 2.3, 3.5 and 3.7).
  - ... in stationary conditions, the first and second order statistics of the spike train Fourier coefficients (across trials) are the relevant features of the neural code for both encoding and determining mutual information (section 3.4).
  - ... in biologically relevant scenarios, modulations of the mean input are more proficient in transmitting signal information than variance modulations in both the linear and nonlinear regime (section 3.6).
  - ... intrinsic dynamic neuron properties, such as the spike initiation time, can crucially shape the information transmission capabilities of a neural population. Signal and noise parameters can have important influence, as well (section 3.6).
  - ... linear approximations to the neural code often provide a good intuition for the basic mechanisms at work, but nonlinear components can be considerably important. I elaborated on the nature of these nonlinearities (section 3.7).
  - ... simultaneous, combined modulation of mean and variance of the input current can have synergetic effects, i.e. information transmission is summed supralinearly (section 3.8).

The obtained results help to understand central aspects of the neural code and information transmission, e.g. by reconciling different methods and probing nonlinearities of

neural information transmission, which are subject to ongoing research efforts. Thereby, my study complements numerous studies on response functions and approximations to the mutual information with exact results. A number of recent publications on neural information transmission, including entire journal issues dedicated to the topic, confirm the prevailing relevance of the subject [325–328].

The applied methods and results constitute an effort of fundamental research that aims for a general description of relevant mechanisms and phenomena. Whereas this endeavor has been insightful, it is based on simplifying assumptions of which one has to be aware when interpreting the results; I discuss respective limitations below.

4.1.1 Neural correlations and information

Temporal correlations between spikes are abundant in observed neural activity. This is not surprising because sensory stimuli themselves exhibit correlated structure across time, e.g. visual scenes or sound sources, which results from the continuous dynamics of the underlying physical processes. However, correlations in neural activity not merely reflect those correlations of the sensory world but are more versatile and are crucial for transmitting and encoding signal information [184, 188, 196, 198, 281]. Temporal correlations can be classified as autocorrelations within single neurons and cross-correlations between neurons or trials. Moreover, these correlations are commonly attributed to the encoded signal or noise, respectively. As discussed, for example in sections 1.3.1.5 and 2.4.3, all these types of correlations influence the capacity of information transmission in neural circuits. They can introduce synergy or redundancy and can render the neural code more efficient or more robust. Shaping and processing spike correlations can thus be expected to be crucial for circuit design optimization. The fact that the correlation method itself [Eq. (1.23)] – as the name implies – relies on temporal correlation functions of spike trains further emphasizes the importance of correlations for neural coding. The findings of this dissertation contribute to a better understanding of the role of temporal correlations in signals that are encoded by neurons in different ways.

In line with the arguments of section 4.2 below, it is likely that not only linear but also higher order correlations are essential features of the neural code. Even though these correlations have not been investigated mechanistically here, the correlation method incorporates all types of correlations and compresses them into two-point correlations.

Of course, inter-neural correlations are particularly relevant for population encoding and hence for information processing in the brain. I give an outlook on respective extensions of the present work in section 4.7.4.

4.1.1.1 Limitations of Poisson neuron models

The calculations in chapter 2 were largely based on the assumption of Poisson neurons which do not possess any intrinsic correlations and, therefore, are suited for the analysis of rate correlations. From the experimental side, Poisson firing often seems to be a good approximation of cortical activity in many cases (see section 1.1.4.3). Theoretical work on balanced networks further suggests that neural populations implement ensembles of Poisson neurons in which the population rate can precisely track input signals, and single neurons spike as Poisson samples of that rate [44, 66, 69]. This Poisson rate coding hypothesis has been supported experimentally [329]. Within the framework of Poisson rate codes, noise correlations are naturally absent and neurons carry information independently. This is well in line with the model assumptions of chapter 2. However, even if Poisson spiking irregularity is given, which is not consistently observed [50], this spiking variability can also be a characteristic of certain spike timing codes on a population level. In this scenario, spike times are precisely distributed and orchestrated across cells in order to implement a signal-encoding population rate [69, 87]. This scheme arguably provides a more efficient signal encoding and, crucially, requires spike times across neurons to be highly dependent. This implies a strong violation of the assumptions that are commonly made for populations of Poisson neurons and that are also underlying a number of results in chapter 2. In conclusion, I think that the assumption of independent rate coding Poisson neurons is a good approximation that allows to derive fundamental insights but does not provide a generalizable framework so that the underlying assumptions need to be kept in mind.

4.1.2 Mean modulation versus variance modulation – the different effects of changing input baseline and fluctuations

The key findings that have motivated the information-theoretic investigation of signal encoding via current mean and variance modulations suggest that changes in the input variance can be translated to spiking output changes, on a population level, much faster

(instantaneously) than mean changes [85, 107, 114, 261]. Such very fast responses are particularly interesting given the rapidness in some cases of perception and decision making. These processes can occur within tenths of milliseconds despite likely involving polysynaptic signaling pathways [114, 330]. A conclusion of chapter 3 is that an information-based advantage of variance modulations over mean modulations does not exist in realistic settings with finite noise correlation and spike initiation times; rather, the opposite applies. Nonetheless, it needs to be emphasized that this result cannot be generalized beyond some innate properties of the used neuron models (see also section 4.5 below). For instance, it has been shown that morphology [17], active channel dynamics [331], or complex onset dynamics [332] can crucially influence a neuron's response properties. On the other hand, experiments probing the responsiveness to input mean and variance changes, such as [85, 114, 115], may have to be reevaluated at conditions that are closer to *in vivo* situations since the network activity that a neuron is embedded in crucially changes its response properties. Eventually, some uncertainty about the role of variance modulations in fast signaling remains and needs to be addressed by an interplay of theoretical and experimental efforts.

Variance modulations (VM) were subject to a number of studies in the past (e.g. in [36, 85, 107, 108, 114, 115, 261, 295]), whereas its formal representation therein varies. Here, I adopted the mathematical representation from [107]. Explicit evidence for variance modulations, as formalized in the present study, in *in vivo* situations seems to be lacking, whereas a considerable body of indirect evidence exists (see section 1.4.2). Conceptually, the two encoding schemes correspond to stimulus-induced changes of the baseline ($\simeq$MM) or the variability ($\simeq$VM) in presynaptic firing. These changes result in mean shifts and modulations of the fluctuations in the membrane potential, both of which are commonly recorded in electrophysiological studies, although with nonstationary properties [251, 255, 256]. Taking yet another perspective, in VM currents, the signal is encoded as the envelope of the fluctuating current trace. This scheme of encoding has been observed for example in weakly electric fish [123, 333]. The abundance and importance of variance modulations can be further accentuated if the concept is considered from a more general point of view. Eq. (3.1) and (3.2) for the MM and VM input currents correspond to special cases of *additive* and *multiplicative noise*, respectively. Whereas these concepts encompass a broader idea, they can be directly linked to this dissertation. Additive and multiplicative noise have been the focus of numerous

neuroscience studies [107, 115, 334], including very recent work regarding information transmission [327]. Moreover, recent studies regarding the role of mean and variance coding within the adaptive coding framework [335] or the sensitivity to rapid input fluctuations in cortical neurons [258] support the persistent relevance of the investigations carried out in the present study.

Apart from encoding a signal directly, neuron populations have been argued to use probabilistic codes that represent *Bayesian inference* of external variables[1]. In this context, it has been shown that uncertainty about external stimuli in some cases is directly encoded in the variability of cortical responses [338]. In postsynaptic, downstream circuits increased input fluctuations then lead to higher activity even if the baseline input is unchanged [336] (however, it needs to be noted that increased presynaptic spiking variability may not unequivocally result from higher input variance). This is well in agreement with the fact that pure variance modulations can induce informative rate modulations, the mechanisms on which encoding through VM is based on. Conceptually, the meaning of MM and VM can thus be conceptually extended to encode stimulus values and uncertainty thereof, respectively. This implies that the signals $s(t)$ in chapter 3 can be thought of representing these quantities, which introduces another dimension to the notion of mean and variance modulations. The information theory of this work then helps to answer the question "How much information do the spiking responses carry about the *posterior* probability of the presented stimulus?".

Moreover, beneficial interactions of signal-induced modulations of mean and fluctuation power of the membrane potential [255, 256] had been pointed out before, thereby indicating the importance of variance modulations on another level. Here, I have further followed that idea and found that the combined presence of MM and VM can have synergetic effects on information transmission (section 3.8). Whereas the notion of combined encoding seems to be a promising candidate for enhancing signal transmission, the mechanism requires further evaluation and elaboration by more detailed biological studies (see section 3.9.4).

[1] Probabilistic and predictive neural coding are important concepts in the brain sciences, according to which neural codes represent not just point estimates of external variables but also aspects of uncertainty and prior knowledge of these estimates (see e.g. [264, 336, 337]). A detailed discussion is outside the scope of this work.

4.1.2.1 Limitations of simulation studies

Most of the results in chapter 3 are based on simulations of the membrane voltage dynamics and spike times (explained in section 3.3.2), due to a lack of available analytic expressions in most considered cases. Naturally, simulation studies only allow for discrete sampling of the space of free parameters which constrains the parameter subspace that can be scanned in a reasonable time span. Accordingly, many of the conclusions in chapter 3 rely on a limited number of data points along each parameter axis that determine the signal and noise properties[2] [3]. Even though these points were chosen to sample the range of biological parameter values reasonably well, a few caveats remain. First, a constant trend in between neighboring data points is usually assumed, but technically it is not guaranteed that no local extrema, oscillations, etc., occur in between sampled points, which would not be visible. These scenarios become more likely if data points are relatively distant on a given parameter axis. Even though such scenarios cannot strictly be ruled out here, the overall consistency of the results and agreement with the analytic calculations (where possible) supports that none such effects are overseen. Second, as the simulations rely on stochastic sampling of signal and noise traces, they are subject to statistic variance that only vanishes in the virtual limit of infinite sequence times and trial numbers. This is the reason why curves are not perfectly smooth. Stochastic variability becomes most relevant when the values of the mutual information and correlation functions are close to zero, which is the case for very small signal strengths, in particular for EIF neurons. Despite this variability being observable in my results, the effects seem negligible, and I am positive that the overall conclusions of this study are unaffected.

4.2 The (non)linearities of information encoding

In chapter 3, I emphasized the importance of nonlinearities in neural coding for both single channel coding (figure 3.20 and 3.17) and combined encoding where nonlinearities can enhance encoding synergy (figure 3.25). In both cases, nonlinear inter-frequency

[2] For example, due to these constraints, I did not consider the suprathreshold regime $[\mu \cdot R > \Theta]$ which has been found to stronger facilitate nonlinear responses [275] but is less compatible with electrophysiological recordings in cortex (cf. section 3.3.1).

[3] Importantly, not all data points that were sampled for this work are shown in the figures but were used to generalize the results and conclusions.

coupling in the correlation functions has emerged as a potentially important mechanism for enhanced information transmission in neural systems compared to purely linear input-output relations. The correlation method used to determine the mutual information captures all linear and nonlinear contributions and, therefore, is eligible to express the importance of nonlinear effects quantitatively. As the correlation method is rather recent, the analysis that I presented in section 3.7 is a novel contribution to the understanding of nonlinear phenomena in neural coding.

Regarding chapter 2.1, no linearity assumption has strictly been made because the results apply for general (static) signal-to-rate transfer functions that are uniquely invertible, as mentioned in section 2.1.1. However, all calculations there rely on the assumption that the firing rate autocorrelation and distribution are known whereas the possible underlying signals are not explicitly considered. An investigation of the effects of nonlinear transfer functions on encoding redundancy is, therefore, open. Conceptually, such an analysis would inquire how the firing rate autocorrelation and distribution are determined by given signal distributions and dynamics for different nonlinear transfer functions. This would then yield the signal-dependent properties of the rate trajectories that are required to determine the mutual information from Eq. (2.3) and (2.4).

A priori, the importance of nonlinear features in the neural code is not surprising because complex computations require nonlinear operations (cf. figure 1.13). As explained in section 1.3, this is also manifested in the fact that artificial (deep) neural networks and perceptrons as their predecessors crucially rely on nonlinear activation functions. Similarly, the successful class of generalized linear models (GLM) are based on nonlinear filtering of linear signal integration. From a more biological viewpoint, nonlinearities in neural transfer functions are simply required by constraints such as physiological upper and lower limits of the firing rate.

On a more abstract, algorithmic level it is questionable whether the complex information processing in the brain can be understood exclusively from the nonlinear properties of single cells. Even though single cells are capable of remarkable signal processing, e.g. in terms of dendritic computations [18], it seems more plausible that most of the information processing is accomplished by neuron circuits and populations and the interactions therein. I think a major challenge for theoretical neuroscience is to identify the algorithmic building blocks of neural computation (is there something like the analogs

of logical gates in computers?), and to figure out how they are implemented in neural circuits. It is reasonable to assume that only circuits of nonlinear neurons can possibly implement these algorithmic units. However, it remains open and exciting to see what specific requirements there are on the properties of individual neurons, or whether most of the computation is lifted by intricately combining rather nonspecific, un-tuned nonlinearities (as in deep neural networks). Related to that is the question how biologically detailed a neuron model needs to be to well fit the nonlinear regime (cf. section 4.5), or conversely, whether the modeling focus needs to be on the connectivity pattern and network architecture. An answer to these questions could also help to better link the diversity of electrophysiological properties in real neuron populations to function (cf. Herfurth & Tchumatchenko 2017 [1]).

4.3 Remarks on the information-theoretic model

The major results of the present study are based on the *correlation method* for calculating the mutual information of signals and spike trains using spike auto- and cross-correlation functions (Dettner *et al.* [98], see section 1.3.2.2). The method provides a very general methodology that is applicable independent of many factors that determine the spiking output, such as the neuron model or noise properties, and thereby yield exact results (under the assumptions discussed below). In particular, the normality assumption for the input statistics that has been essential to many previous methods for computing the mutual information [25, 218] is not required for the correlation method (see section 3.4.2). Regarding the specific questions of this work, the correlation method allowed for a systematic and exact investigation of the information transmission properties of mean and variance modulated signals, the role of temporal signal correlations, and the linearity of the neural information encoding (chapters 2 and 3).

Moreover, the method exhibits reasonably good interpretability as the determining correlation functions are comprehensible spike train features. Especially in the linear regime, the method provides a suitable means to relate signal and neuron response properties to information transmission capabilities (section 3.2). Even the signal encoding features – i.e. the mean values of the spike train Fourier components – I could uncover in this work using the correlation method (section 3.4.3), and providing a deeper general understanding of the neural code.

In the present work I delivered (previously partly lacking) support for the theoretical foundations and validity of the correlation method: I showed that a previously derived putative upper bound for the information is indeed equivalent to the exact information and can, therefore, be equated with the correlation method (section 3.5). In the linear regime, i.e. at small signal strengths, the linearly decodable information (lower bound) and the correlation method converge as expected, which I show in sections 2.1.2 and 3.7. Moreover, in section 3.4, I confirmed numerically that the distribution of the spike train Fourier coefficients is multivariate normal with a diagonal correlation matrix as is required for the correlation method to be valid [98]. This distribution had previously been obtained analytically but could only partly be confirmed numerically [98, 302]. Overall, the present work has hence not only deployed the correlation model but also contributed to establish it further.

4.3.1 Underlying assumptions and limits

Despite its broad applicability, the correlation method relies on some basic assumptions about the information transmitting channel. In short, it requires a neuron model that nonlinearly transforms inputs, for example in the form of somatic currents, into spikes and that has finite memory, i.e. the dependence of spike occurrence only depends on a limited past. The input, both signal and noise component, needs to be stationary and generated from the same stochastic ensemble in each trial. Moreover, spike recording lengths need to be much larger than the intrinsic correlation time and integration times of current and neuron, respectively (see section 1.3.2.2).

For *in vivo* situations, the assumption of (very) long and stationary inputs seems to be the most arguable. Experiments have found that quickly decaying onset responses or even first spike latencies can be stimulus-encoding [70, 212] (cf. section 1.3). The correlation method is naturally not suited to address these scenarios. As thoroughly discussed in section 2.4.3, other methods better apply to situations with short signal lengths and where a few spikes carry the relevant information. Generally, however, it seems reasonable that stationarity is at least a good approximation, and due to the often short current correlation times, recording windows do not need to be very long for the correlation method to apply. As is common in neuroscience, the correlation method considers discrete, stereotyped spikes as the units of inter-neural communication. The

computational role of the shape of action potentials, or other communication channels apart from spike times, is debated but mostly considered secondary [26, 91]. Accordingly, the correlation method does not concern any potential components of the neural code that are features of action potentials themselves or not spike-based at all.

The present work focuses on information transmission in single neurons for which the correlation method has been derived. Obviously, information transmission through single neurons is only an approximation to the brain and a natural extension of the correlation method, and its applications here is an equivalent method to compute the mutual information in neuron populations (see section 4.7.4). Regarding the inputs to a neuron, they have been represented as the net somatic currents resulting from the presynaptic activity. Generally, though, the network scenario can easily be implemented since the correlation method allows for an explicit consideration of presynaptic spikes, as long as stationary spiking is warranted. Including single synapses merely requires an additional model for the postsynaptic potential generation and integration at each synapse, which, however, introduces higher computational demands.

4.4 Population codes and independent neurons

In this work, I considered single neurons across trials of repeated stimulus presentation. In the brain, neurons are embedded in networks, and it is widely assumed that much of the information processing and coding is accomplished by the combined activity of neuron ensembles (see [187] for a review; cf. sections 1.3.1.6 and 4.2). However, it has been argued that population codes can well be understood in terms of the joint activity of *independent* neurons [29]. This presumption is compatible with the correlation method used here: the trial averages in Eqs. (1.20) and (1.22) are equivalent to averages over identical neurons that are conditionally independent given a signal s.

Theoretical analyses according to which inter-neural correlations vanish in balanced networks support the notion of independent neurons in the cortex [66]. Nonetheless, inter-neural correlations do occur in biological networks [188]. Whereas the methods of this work well incorporate signal-dependent correlations (see chapter 2 in particular), noise correlations – i.e. co-variations of spiking variability – have been reported to potentially have crucial importance for coding in populations (they can introduce both

synergy and redundancy; see [184, 190, 196] for reviews). On the other hand, these correlations seem to be very small in many cases [198]. Therefore, the presented results can also help to understand mechanisms of network-level coding: in a recurrent network of N neurons where each neuron fulfills the stationarity and finite memory conditions, the information-relevant features are represented in the $N \times N$ correlation matrix of the spike trains' Fourier coefficients of these neurons. The diagonal elements of this matrix mirror single neuron encoding, which can be understood using my results. This provides a first order approximation of how mean and variance coding strategies impact network-level coding. The tightness of this approximation can be expected to be better if noise correlations are small. The off-diagonal elements which are determined by cross-neural correlations, such as noise correlations, can further reduce or increase the network's information content [196].

Generally, because my work represents the limit of conditionally independent neurons, it can promote more complete network level studies by providing a reference for encoding effects that are not mediated by inter-neural coupling but by temporal interactions only. In section 4.7, I explain how such network studies could look like.

4.5 Integrate-and-fire models as precision/simplicity trade-offs

Investigating inter-neural communication on a theoretical level requires neuron models. These are mathematical or algorithmic representations of the dynamic processes, such as the evolution of membrane voltage or ion concentrations, that are relevant for the questions under consideration. Generally, any neuron model portrays a trade-off between biological detailedness and computational demands and/or mathematical tractability. The design of neuron models that realize the desired trade-off has been an active field of research in computational neuroscience [9, 339]. A very reduced variant of neuron models are Poisson neurons as introduced in section 1.1.4.3 and extensively used in chapter 2.

In this context, it is important to note that there is no unique way to evaluate a given neuron model. The respective measure needs to be specified according to the particular demands for which the model is deployed. Generally, numerous features of real neurons

– among which are spike times, f-I curves, action potential shapes, firing rates, and spiking variability – can serve as objective functions. Examples for evaluating the spike timing precision of a model are spike coincidence rate [340] and explained temporal variance [31]. Another challenge in finding good neuron models is to find the set of stimuli that are used to obtain the reference data from real neurons. Again, the choice will generally be guided by the questions that are addressed with the model.

Whereas detailed neuron models – e.g. the Hodgkin-Huxley model or morphological compartment models – are capable of representing many biological variables, their large number of parameters introduces various problems: these models usually do not fit all evaluation criteria equally well, they are computationally expensive and are prone to overfitting if the parameters are not strongly constrained. Perhaps surprisingly, in many situations, the simpler integrate-and-fire (IF) models have been proven more successful in predicting spike times [31, 340]. In particular, the models used in this work – the leaky integrate-and-fire (LIF) and the exponential integrate-and-fire (EIF) model – realize a good compromise of simplicity and agreement with experimentally measured response properties of cortical neurons (see section 1.1.4 for details). These two models – in between which can be "interpolated" by tuning the spike initiation time – can be fitted to the linear response properties of real neurons. Moreover, LIF and EIF models are one-dimensional point neurons, i.e. they are described by only one state-variable (voltage) and have no spatial extent. This makes them relatively easy to handle, and the LIF model even allows for analytic treatment in some situations, which enabled a systematic and comprehensive investigation of the influence of the diverse model, signal and noise parameters (see chapter 3). More complex IF models incorporating dynamic thresholds and adaptive currents that adequately account for spike history have been demonstrated to be even more accurate [31, 151, 340–342]. However, even though adaptive integrate-and-fire models and GLM may better fit the response properties of real neurons, especially when rate adaptations play a role, the fitted parameter values of these neurons tend to be stimulus dependent. Therefore, these models often do not well generalize across (unseen) stimuli [9, 153, 340]. This is arguably less of a problem with the simpler LIF and EIF models.

Furthermore, more complex models do not permit analytical investigations and require even higher computational resources. This is important to note because already the computer simulations carried out for LIF and EIF neurons were fairly time-consuming.

In extensive studies such as the present one, the computational complexity of the utilized neuron model can thus limit – and has limited in the present study – the completeness at which the space of electrophysiological parameters can be analyzed. Discrete parameter scans have to be carried out whose time requirements increase with the model complexity.

Another benefit of the LIF and EIF neuron models is that their model parameters (input resistance, membrane time constant, spike initiation time) are rather intuitive and measurable electrophysiological cell properties. However, the biological spread of these quantities among and across cell types/areas/species is not negligible [289] and may crucially determine differences in the neurons' function. In the present study, cell parameters were kept constant for each model and chosen based on experimental findings to best represent real cortical neurons (see section 1.1.4). Nonetheless, the chosen parameters determine basic signal filtering properties of the neurons. Carrying out an equivalent analysis with alternative cell parameters could nicely complement the present study, and could help to assess the generality of my findings further.

4.6 Different views on the neural code: information theory, decoding, and computation

In this work, I based my investigations of the neural code on information theory, which has served as an essential pillar of neuroscience research [326]. Alternative approaches to investigate the neural code focus on decoding and the executed computations, which I link to information theory in this section.

4.6.1 Information theory vs decoding

As mentioned in the introduction of section 1.3.2, the analysis of the neural code is usually done through two complementary approaches asking for either the *how* or the *how much* of information encoding in neural activity. The latter corresponds to applying information theory to the neural code, as done here. The former approach often consists of the construction of a *decoder* that aims to infer a stimulus from given firing pattern [29]. According to their guiding questions, both approaches fundamentally differ in the aspects of the neural code they can shed light on. Information theory facilitates

quantification of the amount of transmitted information about a stimulus irrespective of the encoding scheme. Thereby, not having to know this scheme is a great strength of the information-theoretic approach and has made it very popular among neuroscientists. On the other hand, if one is explicitly interested in the utilized encoding scheme, information theory does not directly provide the particulars. Moreover, due to biological constraints, it is not generally clear whether the full amount of information as quantified by information theory is in practice accessible to the neural system under consideration [197].

Decoding methods, in contrast, can give insights into how information is encoded. This is achieved by inspecting which spike train features a functioning decoder, i.e. one that can successfully infer the stimulus from neural activity, relies on. However, it is noteworthy that not all functioning decoders allow for such an inspection of their inner workings. In particular, modern decoders based on artificial neural networks often lack this interpretability [343]. Another systematic limitation of decoding approaches is the impossibility of conceiving decoders for all possible ways of transmitting information: poor performance of a decoder may not be due to the lack of stimulus-relevant information in the neural activity but due to a decoder design that cannot capture the relevant encoding features. Therefore, decoding methods may miss important parts of the stimulus encoding mechanism.

Despite their differences, information theory and decoding methods are not unrelated. Intuitively, it is reasonable that higher mutual information between stimulus and neural response should correspond to a better decodability of the stimulus. This formally manifests in *Fano's inequality* and the *Cramér-Rao bound* [Eq. (4.4)] that both link higher information-theoretic measures (noise entropy and Fisher information) to a better performance of an optimal decoder [29, 204].

As demonstrated in section 3.4.3, under the given conditions, stimulus-information is encoded in the mean values of the spike train Fourier coefficients. Understanding the neural code in this context means to understand how these mean values arise as functions of the presented stimulus. For the linear regime, or more generally where λ^{ld} in figure 3.20 is close to 1, the nature of the neural code can easily be understood because $\lambda^{\mathrm{ld}} \approx 1$ implies that the signal can be decoded through linear filtering of the spiking output Fourier coefficients [128, 218]. In these cases the neural code can be seen as a rate

code in which the power of the different frequencies in the rate dynamics is proportional to the modulations of the signal at the same frequencies (cf. the linear response theory Eq. (3.5)). Importantly, the observed events are spikes, and the underlying rate dynamics can only be inferred by repeated presentation of a stimulus, or by simultaneously observing several "equivalent" neurons in a population. The discrimination between rate and timing code then depends on the temporal resolution at which the rate dynamics is analyzed. My results of chapter 2 imply that there is a natural cutoff for the required resolution of the rate dynamics for capturing all encoded information which is given by the inverse of the cutoff frequency of the linear response functions.

Specifying the neural code from my information-theoretic results in the nonlinear regime, however, is more challenging, in particular when inter-frequency couplings are present as shown in figure 3.17. These situations may require an effort to explicitly develop a suiting decoder to better understand the encoding mechanism at work. The design work of this decoder can be guided by the results that are obtained for the cross-correlation functions (figure 3.17C,D) because they provide information about how different frequencies interact in a nonlinear setting.

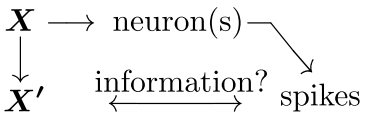
On a more philosophical note, it is not obvious what the decoder of neural activity is in the brain. Assuming the presence of such a decoder, it would have to rely on internal representations of the quantities it is decoding themselves, which in turn would have to be decoded from some neural activity, and so on. Given that the only output of the brain that physically interacts with the environment occurs as motor output via muscles (and glands), one may think of the respective brain cells as the final decoder of brain activity. However, in this concept, decoding is the process that results in behavior, and for intermediate steps of information processing in the brain no explicit decoding takes place. It has, therefore, been suggested that the concept of neural encoding and decoding are only sensible when they are regarded as the causal relations that govern the interaction of environment, motor output and brain activity (see [92]).

4.6.2 Information transmission vs computation

In the present work, I asked how well signals that are fed into a neuron are encoded in its spiking output, and optimization corresponds to maximizing the mutual information of signal and output spikes. Thereby, I made an implicit assumption: the neurons' task

is to represent the input signal dynamics in their digital output optimally. Obviously, this can only be conditionally true because the complex functionality of the brain fundamentally arises from filtering and combining different kinds of sensory and internal information streams. In other words, neural systems perform information processing and computation in which (irrelevant) information is discarded, and a mere passing of information between neurons does not account for the brain's complex functionality. Nonetheless, in many instances, it may be necessary to solely transport information in between different populations/areas of the brain where it can be further processed, and the respective neurons should be optimized for that.

A neural information processing stream can be schematically depicted as:



Here, $\boldsymbol{X}$ denotes, for example, a sensory variable or presynaptic activity in terms of spikes or synaptic currents. Now, $\boldsymbol{X'} = f(\boldsymbol{X})$ represents the function of the variable $\boldsymbol{X}$ to be computed, e.g. the extraction of edges if $\boldsymbol{X}$ is a visual scene. The biologically relevant optimization function then is the mutual information of the spikes and the relevant features $\boldsymbol{X'}$. This includes the special case of unprocessed information passing where $\boldsymbol{X'} = \boldsymbol{X}$. Linking this picture to the present work, the signal $s(t)$ in chapter 3, and similarly $r(t)$ in chapter 2, can be seen as representing the relevant features of an upstream stimulus (e.g. sensory) that have to be extracted from the input currents, which additionally contain different noise parts [cf. Eq. (3.1) and (3.2)]. Generally, one can think of $s(t)$ as being a variable that needs to be transmitted faithfully via the neurons under consideration. The mutual information of $s(t)$ and the spike trains is then the relevant quantity in order to evaluate the present information encoding. For the same reasons – namely the possibility to quantify the faithfulness of extracting and transmitting desired input features – information theory has traditionally been a major tool to analyze and develop principled theories of brain function [326].

However, in information-theoretic approaches, assessing the optimality of information transmission requires prior knowledge of the relevant computations. Otherwise, the interpretation of mutual information values is difficult. For example, low mutual information values can imply a channel with poor information transmission capacity. However, low values can also result from considering inappropriate input features that are not

actually transmitted, inaccessible, or irrelevant to the neuron. Likewise, high measured information about the input can mean a lack of desired computations which generally reduce information about the input.

4.7 Outlook

Finally, in this section, I give an overview of some possible extensions of the present study that I think would reasonably complement, expand, or enhance the presented results. Most of these proposals build upon the preceding discussion in this chapter.

4.7.1 Extended neuron models

In section 4.5, I explained that higher-dimensional neuron models that account for effects of spike history, as the adaptive exponential integrate-and-fire (AdEx) model [342], often capture the electrophysiological properties of cortical neurons better than LIF and EIF neurons. It would be straight-forward, although time-expensive, to rerun the simulations and analyses of chapter 3 with a model like the AdEx model. This could help to evaluate the generality, or conversely model specificity, of the major results of this chapter, e.g. the general information-theoretic benefit of mean modulation over variance modulation or the types of nonlinear contributions to the cross-correlation function. Moreover, using more detailed neuron models may be especially insightful regarding the coupling between variance modulated (fluctuations) and mean modulated (shifts) input currents, investigated in section 3.8. Along these lines, a previous study has investigated the effects of signal-modulated γ-fluctuations in Hodgkin-Huxley-like point neurons and discovered a crucial role of the interaction of fast fluctuations with mean changes in the membrane voltage [256]. Similar effects have been described in section 3.8, but a more detailed model could provide deeper insights into the underlying mechanisms.

4.7.2 Encoding two signals in parallel

I presented the advantages and properties of signal encoding via combined mean and variance modulation in section 3.8. A simultaneous modulation of input current mean and variance can also be conceived in a different version: modulations in the mean and

variance channel could encode two different signals s_1 and s_2 in parallel. In this concept, the signal $s(t)$ that appears twice in Eq. (3.24) is split into two signals $s_1(t)$ and $s_2(t)$. In this parallel encoding scheme, the input current mean $\mu_I(t)$ and variance $\sigma_I^2(t)$ are proportional to either modulating signal,

$$\mu_I(t) \propto s_1(t),$$
$$\sigma_I^2(t) \propto s_2(t).$$

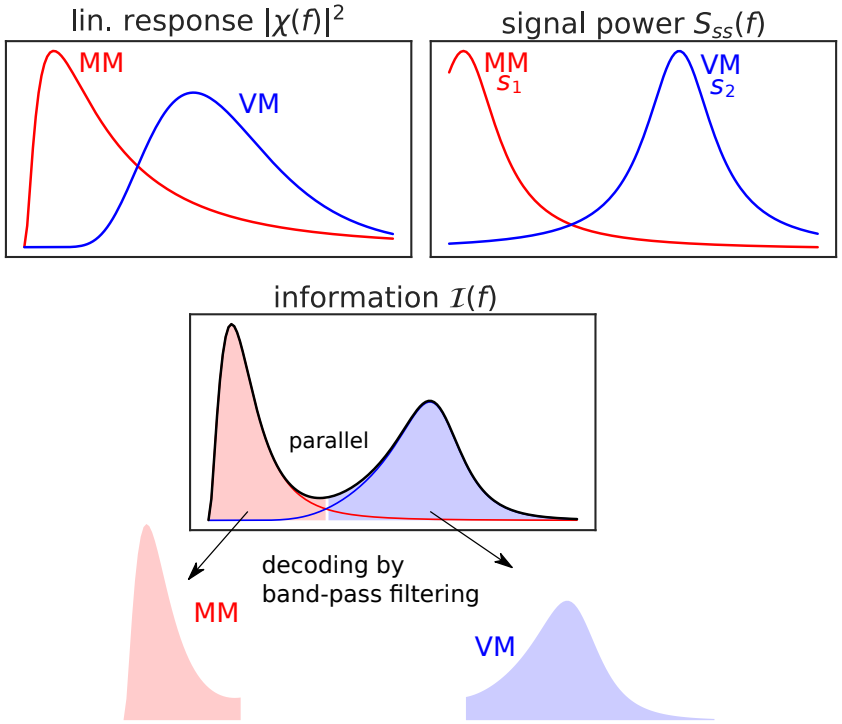


FIGURE 4.1: **Schematic of encoding of two signals by parallel mean and variance modulation.** The figure heuristically illustrates the basic effects that can be expected in parallel signal encoding in the linear regime. Top left: The squared linear response functions $|\chi(f)|^2$ for MM (red) and VM (blue) generally have different shapes and maxima at different frequencies. Top right: In an ideal situation the two signals that have to be encoded have power spectra that match the respective response functions and are located such that they decrease the overlap of the contributions of each signal in the cross-correlation. Signal s_1 modulates the input current mean, s_2 modulates its variance. Center/bottom: The frequencies in the overall information $\mathcal{I}(f)$ generally carry information about both s_1 and s_2. If the two contributions are well separated, however, they can easily be extracted by applying a respective band-pass filter. This yields two practically independent channels of information transmission. [Note that this depiction of the mutual information represents a sketch rather than an exact result; cf. Appendix A.4.]

Once more, an intuition for this coding scheme can be gained by considering the linear

regime (see figure 4.1). Following the logic of section 3.8, now with two signals, the frequency-resolved rate dynamics resulting from parallel encoding are a superposition of the dynamics of each channel separately, which, in turn, are proportional to their respective linear response function. Similarly, in the linear regime, the total mutual information can be pictured to be additively shaped by the contributions from both channels (in a rather heuristic picture; technically, the two channels do not carry independent information). Now, in an optimal setting, the cross-correlations caused by each signal would have as little overlap as possible because this means that both signals can be encoded (almost) independently without interference (see Appendix A.4). This situation is favored by the fact that the linear response functions $\chi_{\mathrm{MM}}(f)$ and $\chi_{\mathrm{VM}}(f)$ tend to be peaked around different frequencies (see for example Eq. (A.18), (A.19)). The overlap can further be decreased if the signals' power spectra are appropriately aligned with the respective peaks of the response functions and have little overlap themselves. As depicted in figure 4.1, the overall information can then be decomposed into two frequency bands which each carry the information about one of the signals. The first step of decoding would then consist of extracting these frequency bands by band-pass filtering. Overall, this yields two practically independent channels of information transmission.

Parallel encoding could, therefore, provide an efficient means to transmit information about two signals simultaneously and hence seems to be a scenario worthwhile investigating (the topic relates to a similar branch of research on multiplexing in neural systems [344]). In particular, it would be interesting to see whether *in vivo* neural systems have optimized the power spectra of mean and variance modulations to allow for reliable encoding/decoding of multiple signals; the opposite situation would be combined encoding where both channels are tuned to encode one signal. Moreover, the discussed mechanisms are largely based on the linearity assumption. Nonlinear effects may considerably change the picture, and hence a further analytic investigation of parallel encoding seems desirable.

4.7.3 *In vivo* recordings

The models and method used in the preceding chapters aim to represent and describe the mechanisms of neural coding in real biological neurons as well as possible. Practically, computational models can only approximate the effect of all relevant neurobiological

factors across different length and time scales (cf. section 4.5). A natural way to assess the validity of my model assumptions and results would be, generally speaking, to record spike trains from real neurons that are stimulated according to my model framework. These spike trains could then be analyzed equivalently to simulated spike trains, and the results could be compared. Thereby, in the ideal case these experiments would be done *in vivo*, and I propose some specific experimental situations in the following.

First, my results of chapter 3 could be tested by injecting mean and variance modulated currents with the properties of interest into current-clamped neurons (whole-cell patch clamp recordings [22]). The mutual information can then be computed according to Eq. (1.23) from trials with repeated and varying signals. It would be particularly interesting to investigate the high-frequency limit and linearity of the mutual information.

Second, in a living neural network, mean and variance encoding schemes are likely used in parallel (see above) because modulations of the excitation and inhibition due to arriving signals often coincide but are not perfectly balanced [68, 256]. Evaluating the neural code used in this situation requires to assess how mean and variance modulations interact intracellularly and on the network level. For example, simultaneously recording the changes of excitatory and inhibitory inputs in nearby neurons as in [74] as a function of sensory inputs *in vivo* could help to assess the amount of variance and mean coding present in the intact brain. This knowledge, in turn, could then be used for further simulations and analysis on parallel and/or combined encoding, depending on the experimental findings. Moreover, generalizing the present results and exploring how mean and variance coding arise and interact in a recurrent network may lead to further insights into the mechanisms of neural coding.

Furthermore, whole-cell recordings could help setting the parameter values in my neuron models (cf. section 4.5). Even though the parameters in this work are chosen according to strong experimental evidence (section 1.1.4), it is known that the response properties of neurons depend on various external variables such as global network states, or noise levels and can dynamically change, e.g. through plasticity. Therefore, recording subthreshold voltage dynamics and spiking responses in neurons *in vivo* under a chosen stimulation with mean and variance modulated currents could reveal crucial response properties in that specific setting. The results could be fed back into the neuron models of this study.

4.7.4 Extension to coding in recurrent networks

In section 4.4, I explained that inter-neural (noise) correlations can considerably influence neural coding on the population level. A natural extension to the presented work would be an analogous information-theoretic analysis for the joint activity of neurons in a population that are not independent but recurrently connected and/or non-trivially correlated. In principle, this requires to extend the correlation method such that it accounts for the off-diagonal terms in the neuron-neuron covariance matrix of the Fourier coefficients. Importantly, such a method for calculating the mutual information of joint neural activity and signals in recurrent neural networks has been devised and is currently under revision. I discussed other methods to calculate the information transmission in correlated neurons in detail in section 2.4.3. However, these are either approximations or of limited applicability under the assumption of long, stationary signals [181–184, 276].

Regarding chapter 2, an analysis that includes the effect of inter-neural correlations would help to understand the overall importance of correlations for information transmission better. In particular, it would be interesting to see whether the simultaneous presence of noise correlations can modify the described effects of signal-induced correlations. As for the results of chapter 3, it is intriguing to learn how noise correlations shape the information in MM and VM encoding in comparison, and how recurrent feedback with the special firing characteristics of each modulation scheme shapes the coding properties. The previously discussed studies could guide these efforts to find exact, analytical extensions of my results to the network case.

4.7.5 Encoding of signal parameters

I have focused on the mutual information in spike trains about signals whose dynamics are stochastic and statistically determined by their power spectrum and probability distribution. I have thereby addressed the *information about specific signal trajectories* that are generated from the same statistical ensemble. This implies that the relevant signal features are to be found in the time-evolution of the signal itself.

Here, I want to introduce another perspective according to which particular statistical parameters of the trajectory $s(t)$ carry relevant information and can thus be interpreted as the relevant, time-independent features. More technically, this approach asks for the

information about the generative factors of the stimulus dynamics. Examples of such factors in this work are the signal correlation time τ_s, amplitude σ_s, or central frequency Ω_0 (in the following, they are denoted in general notation by η). It is conceivable that stimulus features are represented by these parameters rather than by the specific signal trajectories. This scenario seems most likely for low-dimensional stimuli or features thereof. Then, the information about η in spike trains corresponds to the signal information and is the relevant quantity. I propose an investigation of the information transmission capabilities of mean and variance modulation for this scenario analogous to chapter 3 to obtain insight into this type of stimulus encoding on the one hand, and the specific properties of MM and VM on the other hand.

As a remark for a better understanding, the parameter η here determines the trajectories $s(t)$ and, at the same time, is the signal to be encoded. Mean and variance modulation refers to the process through which the time-dependent input current is generated from $s(t)$ and noise $n(t)$:

$$\eta \longrightarrow S = \{s(t;\eta)\} \longrightarrow I(t) = s(t;\eta) \odot n(t) \longrightarrow \text{spikes } R = \{\rho(t)\}, \tag{4.1}$$

where $\odot$ refers to either MM or VM. Then, $\mathcal{I}(R,\eta)$ is the desired mutual information.

Applying the correlation method for "encoding in stimulus parameters" is intricate, however. The spiking probabilities required to compute $\mathcal{I}(R,\eta)$ [cf. Eq. (1.15)] are given through the dependencies in Eq. (4.1):

$$p(R) = \int \mathrm{d}S \int \mathrm{d}\eta \; p(\eta)p(S|\eta)p(R|S), \tag{4.2}$$

$$p(R|\eta) = \int \mathrm{d}S \; p(S|\eta)p(R|S), \tag{4.3}$$

with $p(R|S) = \int \mathrm{d}I \;\; p(I|S)p(R|I)$. These functions and the respective entropies in $\mathcal{I}(R,\eta)$ are generally hard to compute but are numerically accessible.

Moreover, the setting of Eq. (4.1) allows for another information-theoretic analysis that provides knowledge about the decodability of the generative parameter η. The *Cramér-Rao bound* states that the mean squared error $\sigma^2_{\text{est}}(\eta)$ of any (unbiased) estimator of η is limited from below by [29]

$$\sigma^2_{\text{est}} \geq \frac{1}{\mathcal{I}_F(\eta)}, \tag{4.4}$$

where equality applies to an optimal decoder. Here, $\mathcal{I}_F$ is the *Fisher information* and defined as[4]

$$\mathcal{I}_F(\eta) = \left\langle -\frac{\partial^2 \ln p(R|\eta)}{\partial \eta^2} \right\rangle = -\int \mathrm{d}R \; p(R|\eta) \frac{\partial^2 \ln p(R|\eta)}{\partial \eta^2}, \tag{4.5}$$

where $p(R|\eta)$ again denotes the probability of the spiking responses at a given η. From section 3.4 we know that the spiking responses can be decomposed in independent Fourier components at frequencies f that each follow a normal distribution with frequency-dependent mean and variance. Hence, for each frequency it is $p(R|S(\eta)) = \mathcal{N}\left(\boldsymbol{\mu}(f,\eta), \boldsymbol{\sigma}^2(f,\eta)\right)$, and $p(R|\eta)$ can be calculated through Eq. (4.3). This is easier than calculating $p(R)$ and the respective entropies as is necessary for obtaining $\mathcal{I}(R,\eta)$. The complete Fisher information is obtained by applying Eq. (4.5) and integrating over all frequencies. The quantities required to compute the Fisher information are readily available from the spike trains, and a respective numerical analysis seems well applicable. The presented procedure could hence be a promising approach to analyze the properties of neural encoding of statistical signal parameters.

[4]The Fisher information has dimension $[1/\eta^2]$, whereas mutual information is given in bits.

Appendix A

Mathematical derivations and supplemental information

A.1 Spike correlation functions for Poisson neurons with dynamic rate modulation

Here, I illustrate how the correlation functions Eqs. (2.2) and (2.1) can be derived from the properties of Poisson spiking neurons. Following chapter 2, I assume time-varying rate trajectories $r(t)$.

The correlation functions are first considered in the temporal domain, i.e. as the inverse Fourier transforms of Eqs. (1.20) and (1.22). In the time domain, the cross-correlation function of two spike trains from trials m and k at time lag h, generated through a Poisson process (see section 1.1.4.3) with rate $r(t)$, is given by

$$\mathcal{F}^{-1}\{C_{\text{cross}}(f)\}(h) = \frac{1}{T}\int_0^T \mathrm{d}t\ \langle \rho_m(t)\rho_k(t-h)\rangle_{\text{tr}_{m\neq k}}, \tag{A.1}$$

with trial average $\langle\rangle_{\text{tr}}$, and as in the main text $\rho(t)$ denotes a spike train of δ-peaks. As the probabilities of spiking are conditionally independent between trials it follows

$$\begin{aligned}\mathcal{F}^{-1}\{C_{\text{cross}}(f)\}(h) &= \frac{1}{T}\int_0^T \mathrm{d}t\ \langle \rho_m(t)\rangle_{\text{tr}_m}\langle \rho_k(t-h)\rangle_{\text{tr}_k} &\text{(A.2)}\\ &= \frac{1}{T}\int_0^T \mathrm{d}t\ r(t)r(t-h). &\text{(A.3)}\end{aligned}$$

Here, $r(t) = \langle \rho(t) \rangle$ is the definition of the rate in the sense of a PSTH. Fourier transforming Eq. (A.3) yields Eq. (2.2) and is equivalent to the (Fourier transformed) autocorrelation of the PSTH (cf. section 1.16 or [98]).

Similarly, the autocorrelation function in the time domain is given by

$$\mathcal{F}^{-1}\{C_{\text{auto}}(f)\}(h) = \frac{1}{T}\int_0^T \mathrm{d}t \langle \rho_m(t)\rho_m(t-h)\rangle_{\text{tr}_m}. \tag{A.4}$$

Poisson spiking implies that the probability of spiking at time t only depends on $r(t)$ (by independence) and, therefore, at any time lag $|h| > 0$ different and equal trials are not distinguishable. Consequently, for finite time lags the trial average in the previous equation again factorizes. Hence, for $|h| > 0$ it is $\mathcal{F}^{-1}\{C_{\text{auto}}(f)\}(h) = \mathcal{F}^{-1}\{C_{\text{cross}}(f)\}(h)$. For $h = 0$ the autocorrelation reads

$$\mathcal{F}^{-1}\{C_{\text{auto}}(f)\}(0) = \frac{1}{T}\int_0^T \mathrm{d}t\ \langle \rho_m^2(t)\rangle_{\text{tr}_m}. \tag{A.5}$$

Here, the integrand represents for each t the spike autocorrelation at zero time lag for a homogeneous Poisson process with rate $r(t)$. This holds even though time averaging is not carried out because an averaging over many (infinite) trials is equivalent. The autocorrelation of a homogeneous Poisson process with constant rate r is known to be [29]

$$\frac{1}{T}\int_0^T \mathrm{d}t\ \rho(t)\rho_k(t-h) = r\delta(h) + r^2. \tag{A.6}$$

Inserting this for $h = 0$ in Eq. (A.5) and integrating over all times yields

$$\mathcal{F}^{-1}\{C_{\text{auto}}(f)\}(0) = \frac{1}{T}\int_0^T \mathrm{d}t\ r(t)\delta(0) + r^2(t) \tag{A.7}$$

$$= \nu\delta(0) + \sigma_r^2 + \nu^2. \tag{A.8}$$

Here, the last two terms are negligible as they are much smaller than $\nu\delta(0)$ and only present at $h = 0$, and thus do not contribute to the Fourier transform yielding $C_{\text{auto}}(f)$ as $T \to \infty$. Taken together, the autocorrelation in the time domain hence reads

$$\mathcal{F}^{-1}\{C_{\text{auto}}(f)\}(h) = \nu\delta(h) + \frac{1}{T}\int_0^T \mathrm{d}t\ r(t)r(t-h), \tag{A.9}$$

and Eq. (2.1) can be obtained as its Fourier transform.

Using similar arguments as above, I also find that

$$\frac{1}{T}\int_0^T \mathrm{d}t\ \langle \rho_m(t) s_m(t-h)\rangle_{\mathrm{tr}_m} = \frac{1}{T}\int_0^T \mathrm{d}t\ r(t)r(t-h), \tag{A.10}$$

because the rate is fixed across trials with the same signal and I assumed $r(t) \sim s(t)$. Therefore, the signal-response cross-correlation is equivalent to the spike autocorrelation. I used this relation in section 2.1.

A.2 Derivation of mean- and variance modulated input currents

Here, I provide additional information on how mean modulated (MM) and variance modulated (VM) somatic currents – introduced in Fig. 3.1 and Eq. (3.1), (3.2) – can be generated on the network level. To this end, I recapitulate and generalize the mechanisms put forward by previous publications in similar form [107, 114][1].

I assume the existence of a signal encoding population whose firing rates follow the signal $s(t)$ of interest. This population consists of N excitatory (E) and N inhibitory (I) Poisson neurons which fire independently and with a signal-modulated rate

$$\nu_E(t) = \nu_I(t) = \nu_0\left(1 + s(t)\right), \tag{A.11}$$

where ν_0 is the baseline firing rate.

Mean modulation. Within the diffusion approximation, a large number of presynaptic Poisson neurons with very small postsynaptic potentials leads to a Gaussian white or colored noise postsynaptic current $I(t)$ [56, 245, 246]. These excitatory and inhibitory currents have means

$$\mu_E(t) = \langle J_{E,i}\rangle N\nu_E(t), \quad \mu_I(t) = \langle J_{I,i}\rangle N\nu_I(t), \tag{A.12}$$

[1] These studies consider a different scaling of the synaptic strength with N or opposite modulation of excitation and inhibition for variance modulation, respectively.

and variances

$$\sigma_E^2(t) = \langle J_{E,i}^2 \rangle N \nu_E(t), \quad \sigma_I^2(t) = \langle J_{I,i}^2 \rangle N \nu_I(t), \tag{A.13}$$

where $J_{E,i}$, $J_{I,i}$ are the excitatory and inhibitory synaptic strengths of synapse i, respectively, and where $\langle\rangle$ denotes population average. The scaling of the synapses projecting to the downstream population is given by $J_{E,i} = \frac{J_0}{N}$, where J_0 is a constant. This synaptic scaling was also utilized in previous publications, e.g. in [345, 346]. If only N excitatory synapses project to the downstream population, I obtain the following statistics of the somatic current at the downstream population. The mean current is

$$\mu(t) = \mu_E(t) = J_0 \nu_E(t), \tag{A.14}$$

and the variance is

$$\sigma^2(t) = \sigma_E^2(t) + \sigma_{\text{bg}}^2 = J_0^2 \frac{\nu_E(t)}{N} + \sigma_{\text{bg}}^2 \overset{N\to\infty}{=} \sigma_{\text{bg}}^2. \tag{A.15}$$

Here, σ_{bg}^2 reflects signal-independent background fluctuations. The variance of the somatic current is constant in the limit of infinite N, and the signal modulates only the mean somatic current. The last two equations thus represent a possible realization of mean coding; a combination of both inhibitory and excitatory inputs with dominating excitation yields a similar implementation of mean modulation.

Variance modulation. In the second scenario, I consider a feedforward situation where downstream neurons receive inputs from N excitatory and N inhibitory encoding Poisson neurons whose firing rates again follow Eq. (A.11). Now, however, the synaptic scaling is $J_{E,i} = J_{I,i} = \frac{J_0}{\sqrt{N}}$. This type of synaptic scaling has been put forward by models describing the excitation-inhibition balance in cortex [66, 80, 243]. The mean somatic current in this situation reads

$$\mu(t) = \sqrt{N} J_0 \left(\nu_E(t) - \nu_I(t)\right) + \mu_{\text{bg}} = \mu_{\text{bg}},$$

with a constant background current μ_{bg}. For the variance I obtain with Eq. (A.13):

$$\begin{aligned}\sigma^2(t) &= \sigma_E^2(t) + \sigma_I^2(t) = J_0^2(\nu_E(t) + \nu_I(t)) \\ &= 2J_0^2\nu_0(1 + s(t)).\end{aligned}$$

Hence, signals modulating both the excitatory and the inhibitory currents simultaneously can selectively recruit variance modulation while leaving the current mean untouched. From the previous equations, I obtain the general forms of the mean and variance modulated currents that I introduced in Eq. (3.1) and (3.2).

A.3 Linear response functions for mean and variance modulation in LIF neurons with white noise

The quantities used in the linear response-based information approximation of section 3.2 for the LIF model with white noise background ($\tau_n = 0$) are given by [107, 124]:

$$\nu_0^{-1} = \tau_m\sqrt{\pi}\int_{V_r^*}^{\Theta^*} e^{s^2}\,(1 + \text{erf}(s))\,ds, \tag{A.16}$$

$$C^0_{\text{auto}}(\omega) = \nu_0\frac{|\mathcal{D}_{i\omega\tau_m}(\Theta^*)|^2 - e^{2\delta}|\mathcal{D}_{i\omega\tau_m}(V_r^*)|^2}{|\mathcal{D}_{i\omega\tau_m}(\Theta^*) - e^{\delta}e^{i\omega t_r}\mathcal{D}_{i\omega\tau_m}(V_r^*)|^2}, \tag{A.17}$$

$$\begin{aligned}\chi_{\text{MM}}(\omega) &= \nu_0\frac{\mu}{\sigma_n}\frac{i\omega\tau_m}{i\omega\tau_m - 1} \\ &\quad\times\frac{\mathcal{D}_{i\omega\tau_m-1}(\Theta^*) - e^{\delta}\mathcal{D}_{i\omega\tau_m-1}(V_r^*)}{\mathcal{D}_{i\omega\tau_m}(\Theta^*) - e^{\delta}e^{i\omega t_r}\mathcal{D}_{i\omega\tau_m}(V_r^*)}\ ,\end{aligned} \tag{A.18}$$

$$\begin{aligned}\chi_{\text{VM}}(\omega) &= \nu_0\frac{i\omega\tau_m(i\omega\tau_m - 1)}{2 - i\omega\tau_m} \\ &\quad\times\frac{\mathcal{D}_{i\omega\tau_m-2}(\Theta^*) - e^{\delta}\mathcal{D}_{i\omega\tau_m-2}(V_r^*)}{\mathcal{D}_{i\omega\tau_m}(\Theta^*) - e^{\delta}e^{i\omega t_r}\mathcal{D}_{i\omega\tau_m}(V_r^*)},\end{aligned} \tag{A.19}$$

here $\Theta^* = \frac{(\Theta/R-\mu)}{\sqrt{2}\sigma_n}$ and $V_r^* = \frac{(V_r/R-\mu)}{\sqrt{2}\sigma_n}$, and $\mathcal{D}_a(x)$ denotes the parabolic cylinder function [347].

A.4 Mutual information for encoding of two signals through parallel mean and variance modulation

According to section 4.7.2, two different signals can be simultaneously encoded if they modulate the mean and variance of the input current, respectively. In the following, these two signals are assumed independent and denoted by $s_1 = s_{\mathrm{MM}}$ and $s_1 = s_{\mathrm{VM}}$. The total input current then reads

$$I_{\mathrm{par}}(t) = \mu(1 + s_{\mathrm{MM}}(t)) + \frac{\sigma_n}{\sqrt{2}}\sqrt{1 + s_{\mathrm{VM}}(t)}\tilde{\xi}_1(t) + \frac{\sigma_n}{\sqrt{2}}\tilde{\xi}_2(t), \tag{A.20}$$

where the parameters are as in Eq. (3.24). Each signal can be considered independent noise with respect to the other signal. A linear response theory in the style of sections 3.2 and 3.8.2 then yields the following correlation functions:

$$C^{\mathrm{par,lin}}_{\mathrm{auto}}(f) = C^{0}_{\mathrm{auto}}(f) + |\chi_{\mathrm{MM}}(f)|S^{\mathrm{MM}}_{ss}(f) + |\chi_{\mathrm{VM}}(f)|^2 S^{\mathrm{VM}}_{ss}(f), \tag{A.21}$$

$$C^{\mathrm{par,lin}}_{\mathrm{cross},\alpha}(f) = |\chi_{\alpha}(f)|^2 S^{\alpha}_{ss}(f) \quad \text{with} \quad \alpha = \mathrm{MM\ or\ VM}. \tag{A.22}$$

Inserting the correlation functions in Eq. (1.23) yields for the information about s_{MM} ($\alpha = \mathrm{MM}$) and s_{VM} ($\alpha = \mathrm{VM}$):

$$\mathcal{I}^{\mathrm{par}}_{\mathrm{lin},\alpha}(f) = \frac{1}{2}\log_2\left(1 + \frac{|\chi_{\alpha}(f)|^2 S^{\alpha}_{ss}(f)}{C^{0}_{\mathrm{auto}}(f) + |\chi_{\beta}(f)|^2 S^{\beta}_{ss}(f)}\right), \tag{A.23}$$

where $\beta = \mathrm{MM}$ if $\alpha = \mathrm{VM}$, and vice versa. From this expression follows that the information about the signals is diminished when the cross-correlations overlap. The total mutual information is the sum of the information about each signal.

Appendix B

Symbols

Chapter 1

symbol	name/description	unit
$[x]$	molar concentration (of x)	M = mole/litre
P	membrane permeability	cm/s
I	electric current	A
g	conductance	Ω^{-1}
a	distance	m
R	input resistance	Ω
C_m	membrane conductance	F
$V(t)$	(time-dependent) membrane potential	V
$E = V_{\text{rest}}$	membrane (equilibrium) resting potential	V
V_{reset}	membrane reset potential (after spiking)	V
τ_m	membrane time constant	s
τ_{syn}	synaptic relaxation time	s
Θ	threshold voltage for the action potential initiation	V
t_r	refractory period	s
Δ_T	spike initiation time (spike slope factor)	s
$\rho(t)$	spike train of δ-peaks	Hz
$s(t)$	stimulus (signal)	(depends)
$r(t)$	instantaneous firing rate	spikes/s
CV_{ISI}	coefficient of variation of ISI dist.	1

$L(t)$	linear response function (Wiener kernel)	$\mathrm{A{\cdot}s^{-2}}$
Δ	discretized time bin/window size	s
T	length of signal/spike train sequences	s
f	frequency	$\mathrm{s^{-1}}$
ω	angular frequency $(2\pi \cdot f)$	$\mathrm{rad{\cdot}s^{-1}}$
$\tilde{x}(f) = \mathcal{F}(x(t))$	Fourier transform of $x(t)$	[units of $x(t)$]$^{-1}$
$\mathcal{I}$	mutual information (general symbol)	$\propto$ bits
$C_{xy}(f)$, $S_{xy}(f)$	(cross-)correlation (also spectrum) of x and y	$\propto$ units of $x \cdot y$
$C_{\mathrm{auto}}(f)$	spike auto-correlation in Fourier domain	Hz
$C_{\mathrm{cross}}(f)$	spike cross-correlation in Fourier domain	Hz
$S_{ss}(f)$	signal power spectrum	$\mathrm{A^2s^2}$
$I_0(t)$	input current during spont. activity	A
μ	constant part of input current	A
$\xi(t)$	noise part of input current	A
τ_n	noise correlation time	s
σ_n	noise standard deviation (amplitude)	A
ν	time/trial avg firing rate	Hz
ν_0	time/trial avg firing rate of spontaneous activity	Hz
$\mathcal{I}_{\mathrm{tot}}(f)$	exact, total mutual information per frequency	bits/Hz
$\mathcal{I}_{\mathrm{tot}}$	exact, total mutual information per spike	bits/spike
$\mathcal{I}_{\mathrm{ld/ub}}(f)$	mutual information lower/upper bound per frequency	bits/Hz
$\mathcal{I}_{\mathrm{ld/ub}}$	mutual information lower/upper bound per spike	bits/spike
$\mathcal{I}_{\mathrm{ind}}$	mutual information of independent spikes per spike	bits/spike
γ^2_{xy}	statistical coherence of x and y	1

Chapter 2

Previously listed symbols are not listed unless their use is modified.

symbol	name/description	unit
$r(t)$	here, rate trajectory representing signal $s(t)$	Hz
ν	rate process mean	Hz
σ_r	standard deviation of rate process	Hz
τ	rate process correlation time	s
$S_{rr}(f)$	rate power spectrum (here equiv. $S_{ss}(f)$)	Hz^{-2}
$S_{rr}^{\mathrm{exp}}(f)$	rate power spectrum for Ornstein-Uhlenbeck process	Hz^{-2}
$\mathcal{I}_{\mathrm{corr}}$	mutual information per spike at rate correlations only	bits/spike
$\mathrm{X}_{\mathrm{tele/uni/OU}}$	denotes case of telegraph, uniform and Ornstein-Uhlenbeck process	–
$\mathcal{I}_0$	leading order of expansion of $\mathcal{I}_{\mathrm{corr}}$ and $\mathcal{I}_{\mathrm{ind}}$ in σ_r	bits/spike

Chapter 3

Previously listed symbols are not listed unless their use is modified.

symbol	name/description	unit
$\mathrm{X}_{\mathrm{MM/VM}}$, $\mathrm{X}^{\mathrm{MM/VM}}$	labels MM and VM denote the mean and variance modulation case	–
$\mathrm{X}_{\mathrm{lin}}$, $\mathrm{X}^{\mathrm{lin}}$	label refers to analytic, linear approximation (e.g. $C_{\mathrm{cross}}^{\mathrm{lin}}(f)$)	–
$\mathrm{X}_{\mathrm{comb}}$, $\mathrm{X}^{\mathrm{comb}}$	label refers to case of combined MM-VM encoding	–
σ_s	signal strength as modulation depth, equiv. to std of signal	1
Ω_0	central signal frequency, frequency of maximum power in signal	Hz
τ_s	signal correlation time	s
$\hat{\sigma}_n$	normalized noise strength (c.f. table 3.1)	1
$\hat{\nu}$	firing rates corresponding to the different $\hat{\sigma}_n$	Hz
$C_{\mathrm{auto}}^0(f)$	spike train autocorrelation function at absence of signal ($\sigma_s = 0$)	s^{-1}
$\chi(f)$	linear response function in frequency domain	Hz
β	ratio of information in MM and VM channel	1

λ^{ld}	fraction of linearly decodable information	1
$\tilde{\xi}(t)$	normalized noise parts of input current with variance 1	1
$d(x, y)$	relative difference between x and y	1

Bibliography

[1] T. Herfurth and T. Tchumatchenko. How linear response shaped models of neural circuits and the quest for alternatives. *Current Opinion in Neurobiology*, 46, 2017.

[2] T. Herfurth and T. Tchumatchenko. Quantifying encoding redundancy induced by rate correlations in poisson neurons. *Physical Review E*, 99(4):042402, 2019.

[3] T. Herfurth and T. Tchumatchenko. Information transmission of mean and variance coding in integrate-and-fire neurons. *Physical Review E*, 99(3):032420, 2019.

[4] A. Plebe and M. Vivian. The computational units of the brain. In *Neurosemantics*, pages 9–35. Springer, 2016.

[5] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, A. J. Hudspeth, et al. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

[6] F. A. Azevedo, L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, R. Lent, S. Herculano-Houzel, et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541, 2009.

[7] M. A. Paradiso, M. F. Bear, and B. W. Connors. Neuroscience: exploring the brain. *Hagerstwon, MD: Lippincott Williams & Wilkins*, 718, 2007.

[8] A. Araque, G. Carmignoto, P. G. Haydon, S. H. Oliet, R. Robitaille, and A. Volterra. Gliotransmitters travel in time and space. *Neuron*, 81(4):728–739, 2014.

[9] W. Gerstner and W. M. Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002.

[10] E. G. Jones. Golgi, cajal and the neuron doctrine. *Journal of the History of the Neurosciences*, 8(2):170–178, 1999.

[11] O. Deiters. *Untersuchungen über Gehirn und Rückenmark des Menschen und der Säugethiere*. F. Veiweg, 1865.

[12] D. H. Hubel. *Eye, brain, and vision*, volume 22. Scientific American Library New York, 1988.

[13] T. Schwann. *Mikroskopische Untersuchungen über die Uebereinstimmung in der Struktur und dem Wachsthum der Thiere und Pflanzen*. Berlin: Sander, 1839.

[14] S. Ramón y Cajal. Textura del sistema nervioso del hombre y de los vertebrados. translation: Texture of the nervous system of man and the vertebrates. vol. 1. 1899, 1899.

[15] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Essential cell biology*. Garland Science, 2013.

[16] S. Richter, R. Loesel, G. Purschke, A. Schmidt-Rhaesa, G. Scholtz, T. Stach, L. Vogt, A. Wanninger, G. Brenneis, C. Döring, et al. Invertebrate neurophylogeny: suggested terms and definitions for a neuroanatomical glossary. *Frontiers in Zoology*, 7(1):29, 2010.

[17] S. Ostojic, G. Szapiro, E. Schwartz, B. Barbour, N. Brunel, and V. Hakim. Neuronal Morphology Generates High-Frequency Firing Resonance. *Journal of Neuroscience*, 35 (18):7056–7068, 2015.

[18] G. Stuart, N. Spruston, and M. Häusser. *Dendrites*. Oxford University Press, 2016.

[19] M. London and M. Häusser. Dendritic computation. *Annual Review of Neuroscience*, 28: 503–532, 2005.

[20] A. Tzilivaki and P. Poirazi. Challenging the point neuron dogma: Fs basket cells as 2-stage nonlinear integrators. *bioRxiv*, page 251314, 2018.

[21] E. Neher and B. Sakmann. Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature*, 260(5554):799, 1976.

[22] A. Molleman. *Patch clamping: an introductory guide to patch clamp electrophysiology*. John Wiley & Sons, 2003.

[23] D. E. Goldman. Potential, impedance, and rectification in membranes. *The Journal of general physiology*, 27(1):37–60, 1943.

[24] G. B. Ermentrout and D. H. Terman. *Mathematical foundations of neuroscience*, volume 35. Springer Science & Business Media, 2010.

[25] F. Rieke. *Spikes: exploring the neural code*. MIT press, 1999.

[26] R. Brette. Philosophy of the Spike: Rate-Based vs. Spike-Based Theories of the Brain. *Frontiers in Systems Neuroscience*, 9, 2015.

[27] A. L. Hodgkin and A. F. Huxley. A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerves. *J. Physiol.*, 117:500–544, 1952.

[28] J. Couto, D. Linaro, E. De Schutter, and M. Giugliano. On the firing rate dependency of the phase response curve of rat purkinje neurons in vitro. *PLoS Computational Biology*, 11(3):1–23, 2015.

[29] P. Dayan and L. F. Abbott. *Theoretical neuroscience*, volume 806. Cambridge, MA: MIT Press, 2001.

[30] F. Gabbiani and S. J. Cox. *Mathematics for neuroscientists*. Academic Press, 2010.

[31] C. Teeter, R. Iyer, V. Menon, N. Gouwens, D. Feng, J. Berg, A. Szafer, N. Cain, H. Zeng, M. Hawrylycz, et al. Generalized leaky integrate-and-fire models classify multiple neuron types. *Nature communications*, 9(1):709, 2018.

[32] A. Rauch, G. La Camera, H.-R. Lüscher, W. Senn, and S. Fusi. Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo–like input currents. *Journal of Neurophysiology*, 90(3):1598–1612, 2003.

[33] B. Ahmed, J. Anderson, R. Douglas, K. Martin, and D. Whitteridge. Estimates of the net excitatory currents evoked by visual stimulation of identified neurons in cat visual cortex. *Cerebral cortex (New York, NY: 1991)*, 8(5):462–476, 1998.

[34] D. A. McCormick. Membrane properties and neurotransmitter actions. *The synaptic organization of the brain*, pages 37–75, 1998.

[35] N. Fourcaud-Trocmé, D. Hansel, C. Van Vreeswijk, and N. Brunel. How spike generation mechanisms determine the neuronal response to fluctuating inputs. *Journal of Neuroscience*, 23(37):11628–11640, 2003.

[36] N. Fourcaud-Trocmé and N. Brunel. Dynamics of the instantaneous firing rate in response to changes in input statistics. *Journal of Computational Neuroscience*, 18(3):311–321, 2005.

[37] L. Badel, S. Lefort, T. K. Berger, C. C. Petersen, W. Gerstner, and M. J. Richardson. Extracting non-linear integrate-and-fire models from experimental data using dynamic i–v curves. *Biological cybernetics*, 99(4-5):361, 2008.

[38] L. Badel, S. Lefort, R. Brette, C. C. Petersen, W. Gerstner, and M. J. Richardson. Dynamic iv curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. *Journal of Neurophysiology*, 99(2):656–666, 2008.

[39] M. Pospischil, Z. Piwkowska, T. Bal, and A. Destexhe. Comparison of different neuron models to conductance-based post-stimulus time histograms obtained in cortical pyramidal cells using dynamic-clamp in vitro. *Biological cybernetics*, 105(2):167, 2011.

[40] N. Brunel, V. Hakim, and M. J. Richardson. Single neuron dynamics and computation. *Current Opinion in Neurobiology*, 25:149–155, 2014.

[41] E. M. Izhikevich. *Dynamical systems in neuroscience*. MIT press, 2007.

[42] A. Fairhall. Spike coding. In K. Doya, editor, *Bayesian Brain: Probabilistic Approaches to Neural Coding*, page 17. MIT Press, 2007.

[43] W. R. Softky and C. Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, 13(1):334–350, 1993.

[44] M. N. Shadlen and W. T. Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18 (10):3870–3896, 1998.

[45] M. N. Shadlen, K. H. Britten, W. T. Newsome, and J. A. Movshon. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience*, 16(4):1486–1510, 1996.

[46] M. N. Shadlen and W. T. Newsome. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4(4):569–579, 1994.

[47] D. J. Tolhurst, J. A. Movshon, and A. F. Dean. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research*, 23(8):775–785, 1983.

[48] E. D. Gershon, M. C. Wiener, P. E. Latham, and B. J. Richmond. Coding strategies in monkey v1 and inferior temporal cortices. *Journal of Neurophysiology*, 79(3):1135–1144, 1998.

[49] A. Compte, C. Constantinidis, J. Tegnér, S. Raghavachari, M. V. Chafee, P. S. Goldman-Rakic, and X.-J. Wang. Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *Journal of neurophysiology*, 90(5): 3441–3454, 2003.

[50] G. Maimon and J. A. Assad. Beyond poisson: increased spike-time regularity across primate parietal cortex. *Neuron*, 62(3):426–440, 2009.

[51] G. Söhl, S. Maxeiner, and K. Willecke. Expression and functions of neuronal gap junctions. *Nature reviews neuroscience*, 6(3):191, 2005.

[52] M. V. Bennett and R. S. Zukin. Electrical coupling and neuronal synchronization in the mammalian brain. *Neuron*, 41(4):495–511, 2004.

[53] B. Katz. Mechanisms of synaptic transmission. *Reviews of Modern Physics*, 31(2):524, 1959.

[54] A. Peters. The fine structure of the nervous system. *Neurons and their supporting cells*, 1991.

[55] A. Destexhe, Z. F. Mainen, and T. J. Sejnowski. Kinetic models of synaptic transmission. *Methods in neuronal modeling*, 2:1–25, 1998.

[56] N. Fourcaud and N. Brunel. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Computation*, 14(9):2057–2110, 2002.

[57] S. Ranson and S. Clark. The cerebral cortex. *The anatomy of the nervous system: its development and function, 10th edn. Saunders, Philadelphia*, pages 347–382, 1959.

[58] A. Rockel, R. W. Hiorns, and T. Powell. The basic uniformity in structure of the neocortex. *Brain: a journal of neurology*, 103(2):221–244, 1980.

[59] C. N. Carlo and C. F. Stevens. Structural uniformity of neocortex, revisited. *Proceedings of the National Academy of Sciences*, 110(4):1488–1493, 2013. doi: 10.1073/pnas.1221398110.

[60] M. A. Whittington, R. D. Traub, and J. G. Jefferys. Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature*, 373(6515): 612, 1995.

[61] H. Markram, M. Toledo-Rodriguez, Y. Wang, A. Gupta, G. Silberberg, and C. Wu. Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience*, 5(10):793, 2004.

[62] K. D. Miller. Canonical computations of cerebral cortex. *Current opinion in neurobiology*, 37:75–84, 2016.

[63] J. J. Knierim and D. C. Van Essen. Neuronal responses to static texture patterns in area v1 of the alert macaque monkey. *Journal of Neurophysiology*, 67(4):961–980, 1992.

[64] W. T. Newsome, K. H. Britten, J. Movshon, and M. N. Shadlen. Single neurons and the perception of visual motion. In *Neural mechanisms of visual perception. Proceedings of the Retina Research Foundation symposium*. Portfolio Publishing, 1989.

[65] J. Karbowski. Constancy and trade-offs in the neuroanatomical and metabolic design of the cerebral cortex. *Frontiers in neural circuits*, 8:9, 2014.

[66] C. van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293), 1996.

[67] C. van Vreeswijk and H. Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural computation*, 10(6):1321–1371, 1998.

[68] J. S. Isaacson and M. Scanziani. How inhibition shapes cortical activity. *Neuron*, 72(2): 231–243, 2011.

[69] S. Denève and C. K. Machens. Efficient codes and balanced networks. *Nature Neuroscience*, 19(3):375, 2016.

[70] M. Wehr and A. M. Zador. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, 426(6965):442–446, 2003.

[71] B. Haider, A. Duque, A. R. Hasenstaub, and D. A. McCormick. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience*, 26(17):4535–4545, 2006.

[72] Y. Shu, A. Hasenstaub, and D. A. McCormick. Turning on and off recurrent balanced cortical activity. *Nature*, 423(6937):288–293, 2003.

[73] M. Xue, B. V. Atallah, and M. Scanziani. Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature*, 511(7511):596, 2014.

[74] M. Okun and I. Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience*, 11(5):535–537, 2008.

[75] A. Tan, S. Andoni, and N. Priebe. A spontaneous state of weakly correlated synaptic excitation and inhibition in visual cortex. *Neuroscience*, 247:364–375, 2013.

[76] C. Poo and J. S. Isaacson. Odor representations in olfactory cortex:sparse coding, global inhibition, and oscillations. *Neuron*, 62(6):850–861, 2009.

[77] B. V. Atallah and M. Scanziani. Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. *Neuron*, 62(4):566–577, 2009.

[78] W. B. Wilent and D. Contreras. Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex. *Nature neuroscience*, 8(10):1364, 2005.

[79] A. Tan and M. Wehr. Balanced tone-evoked synaptic excitation and inhibition in mouse auditory cortex. *Neuroscience*, 163(4):1302–1315, 2009.

[80] G. K. Wu, P. Li, H. W. Tao, and L. I. Zhang. Nonmonotonic synaptic excitation and imbalanced inhibition underlying cortical intensity tuning. *Neuron*, 52(4):705–715, 2006.

[81] N. Dehghani, A. Peyrache, B. Telenczuk, M. Le Van Quyen, E. Halgren, S. S. Cash, N. G. Hatsopoulos, and A. Destexhe. Dynamic balance of excitation and inhibition in human and monkey neocortex. *Scientific Reports*, 6, 2016.

[82] G. K. Wu, R. Arbuckle, B.-H. Liu, H. W. Tao, and L. I. Zhang. Lateral sharpening of cortical frequency tuning by approximately balanced inhibition. *Neuron*, 58(1):132–143, 2008.

[83] A. Destexhe, M. Rudolph, and D. Paré. The high-conductance state of neocortical neurons in vivo. *Nature Reviews. Neuroscience*, 4(9):739–751, 2003.

[84] M. V. Tsodyks and T. Sejnowski. Rapid state switching in balanced cortical network models. *Network: Computation in Neural Systems*, 6(2):111–124, 1995.

[85] G. Silberberg, M. Bethge, H. Markram, K. Pawelzik, and M. Tsodyks. Dynamics of population rate codes in ensembles of neocortical neurons. *Journal of Neurophysiology*, 91 (2):704–709, 2004.

[86] R. Rubin, L. Abbott, and H. Sompolinsky. Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. *Proceedings of the National Academy of Sciences*, page 201705841, 2017.

[87] M. Boerlin, C. K. Machens, and S. Denève. Predictive coding of dynamical variables in balanced spiking networks. *PLoS computational biology*, 9(11):e1003258, 2013.

[88] S. B. Nelson and V. Valakh. Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. *Neuron*, 87(4):684–698, 2015.

[89] M. Selten, H. van Bokhoven, and N. N. Kasri. Inhibitory control of the excitatory/inhibitory balance in psychiatric disorders. *F1000Research*, 7, 2018.

[90] O. Yizhar, L. E. Fenno, M. Prigge, F. Schneider, T. J. Davidson, D. J. OShea, V. S. Sohal, I. Goshen, J. Finkelstein, J. T. Paz, et al. Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature*, 477(7363):171, 2011.

[91] D. Debanne, A. Bialowas, and S. Rama. What are the mechanisms for analogue and digital signalling in the brain? *Nature Reviews Neuroscience*, 14(1):63, 2013.

[92] R. Brette. Is coding a relevant metaphor for the brain? *bioRxiv*, page 168237, 2017.

[93] E. D. Adrian. *The basis of sensation.* Christophers; London, 22 Berners Steeet, W. 1, 1928.

[94] B. P. Bean. The action potential in mammalian central neurons. *Nature Reviews Neuroscience*, 8(6):451, 2007.

[95] A. A. Faisal, L. P. Selen, and D. M. Wolpert. Noise in the nervous system. *Nature reviews neuroscience*, 9(4):292, 2008.

[96] M. L. Schölvinck, A. B. Saleem, A. Benucci, K. D. Harris, and M. Carandini. Cortical state determines global variability and correlations in visual cortex. *Journal of Neuroscience*, 35(1):170–178, 2015.

[97] A. Renart and C. K. Machens. Variability in neural activity and behavior. *Current opinion in neurobiology*, 25:211–220, 2014.

[98] A. Dettner, S. Münzberg, and T. Tchumatchenko. Temporal pairwise spike correlations fully capture single-neuron information. *Nature Communications*, 7, 2016.

[99] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1962.

[100] A. P. Georgopoulos, J. F. Kalaska, R. Caminiti, and J. T. Massey. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, 2(11):1527–1537, 1982.

[101] J. Aljadeff, B. J. Lansdell, A. L. Fairhall, and D. Kleinfeld. Analysis of neuronal spike trains, deconstructed. *Neuron*, 91(2):221–259, 2016.

[102] S. Ostojic. Inter-spike interval distributions of spiking neurons driven by fluctuating inputs. *Journal of Neurophysiology*, 106(1):361–373, 2011.

[103] P. Hanggi and H. Thomas. Stochastic processes: Time evolution, symmetries and linear response. *Physics Reports*, 88:207–319, 1982.

[104] A. Kruscha and B. Lindner. Spike-count distribution in a neuronal population under weak common stimulation. *Phys. Rev. E*, 92(5), 2015.

[105] A. Kruscha and B. Lindner. Partial synchronous output of a neuronal population under weak common noise: Analytical approaches to the correlation statistics. *Phys. Rev. E*, 94 (2), 2016.

[106] N. Wiener. *Nonlinear problems in random theory,*. Wiley and MIT press, 1958.

[107] B. Lindner and L. Schimansky-Geier. Transmission of noise coded versus additive signals through a neuronal ensemble. *Physical Review Letters*, 86(14):2934–2937, 2001.

[108] M. J. E. Richardson. Firing-rate response of linear and nonlinear integrate-and-fire neurons to modulated current-based and conductance-based synaptic drive. *Phys. Rev. E*, 76(2), 2007.

[109] T. Tchumatchenko and F. Wolf. Representation of dynamical stimuli in populations of threshold neurons. *PLoS Computational Biology*, 7(10), 2011.

[110] J. M. Goldberg and P. B. Brown. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *Journal of neurophysiology*, 32(4):613–636, 1969.

[111] M. P. Touzel and F. Wolf. Complete firing-rate response of neurons with complex intrinsic dynamics. *PLoS computational biology*, 11(12):e1004636, 2015.

[112] G. Testa-Silva, M. B. Verhoog, D. Linaro, C. P. J. de Kock, J. C. Baayen, R. M. Meredith, C. I. De Zeeuw, M. Giugliano, and H. D. Mansvelder. High Bandwidth Synaptic Communication and Frequency Tracking in Human Neocortex. *PLoS Biology*, 12(11), 2014.

[113] J. Doose, G. Doron, M. Brecht, and B. Lindner. Cellular/Molecular Noisy Juxtacellular Stimulation In Vivo Leads to Reliable Spiking and Reveals High-Frequency Coding in Single Neurons. *Journal of Neuroscience*, 36(43):11120–11132, 2016.

[114] T. Tchumatchenko, A. Malyshev, F. Wolf, and M. Volgushev. Ultrafast Population Encoding by Cortical Neurons. *Journal of Neuroscience*, 31(34):12171–12179, 2011.

[115] C. Boucsein, T. Tetzlaff, R. Meier, A. Aertsen, and B. Naundorf. Dynamical Response Properties of Neocortical Neuron Ensembles: Multiplicative versus Additive Noise. *Journal of Neuroscience*, 29(4):1006–1010, 2009.

[116] T. O. Sharpee. Computational Identification of Receptive Fields. *Annual Review of Neuroscience*, 36(1):103–120, 2013.

[117] A. Ramirez, E. A. Pnevmatikakis, J. Merel, L. Paninski, K. D. Miller, and R. M. Bruno. Spatiotemporal receptive fields of barrel cortex revealed by reverse correlation of synaptic input. *Nature Neuroscience*, 17(6):866–75, 2014.

[118] L. M. Miller, M. A. Escabí, H. L. Read, and C. E. Schreiner. Spectrotemporal Receptive Fields in the Lemniscal Auditory Thalamus and Cortex. *Journal of Neurophysiology*, 87 (1), 2002.

[119] J. Touryan and J. A. Mazer. Linear and non-linear properties of feature selectivity in V4 neurons. *Frontiers in Systems Neuroscience*, 9(May), 2015.

[120] J. P. Jones and L. A. Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, 58(6):1187–1211, 1987.

[121] P. Z. Marmarelis and K.-I. Naka. White-noise analysis of a neuron chain: an application of the wiener theory. *Science*, 175(4027):1276–1278, 1972.

[122] S. Wienbar and G. W. Schwartz. The dynamic receptive fields of retinal ganglion cells. *Progress in Retinal and Eye Research*, 2018.

[123] M. G. Metzen, M. Jamali, J. Carriot, O. Ávila-Åkerberg, K. E. Cullen, and M. J. Chacron. Coding of envelopes by correlated but not single-neuron activity requires neural variability. *Proc. Natl. Acad. Sci.*, 112(15):4791–4796, 2015.

[124] S. Ostojic, N. Brunel, and V. Hakim. How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *Journal of Neuroscience*, 29 (33):10234–10253, 2009.

[125] T. Tchumatchenko, T. Geisel, M. Volgushev, and F. Wolf. Spike Correlations – What Can They Tell About Synchrony? *Frontiers in Neuroscience*, 5, 2011.

[126] T. Tchumatchenko and C. Clopath. Oscillations emerging from noise-driven steady state in networks with electrical synapses and subthreshold resonance. *Nature Communications*, 5, 2014.

[127] S. Blankenburg and B. Lindner. The effect of positive interspike interval correlations on neuronal information transmission. *Math. Biosci. Eng.*, 13(3):461–81, 2016.

[128] F. Gabbiani. Coding of time-varying signals in spike trains of linear and half-wave rectifying neurons. *Network*, 7(1):61–85, 1996.

[129] S. Blankenburg, W. Wu, B. Lindner, and S. Schreiber. Information filtering in resonant neurons. *Journal of Computational Neuroscience*, 39(3):349–370, 2015.

[130] X. Pitkow, S. Liu, D. E. Angelaki, G. C. DeAngelis, and A. Pouget. How Can Single Sensory Neurons Predict Behavior? *Neuron*, 87(2):411–423, 2015.

[131] X. Pitkow and M. Meister. Decorrelation and efficient coding by retinal ganglion cells. *Nature Neuroscience*, 15(4):628–635, 2012.

[132] M. Carandini. What simple and complex cells compute. *The Journal of physiology*, 577 (2):463–466, 2006.

[133] Y. Chen, S. Anand, S. Martinez-Conde, S. L. Macknik, Y. Bereshpolova, H. A. Swadlow, and J.-M. Alonso. The linearity and selectivity of neuronal responses in awake visual cortex. *Journal of vision*, 9(9):12–12, 2009.

[134] A. J. Bell and T. J. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Research*, 37(23):3327–38, 1997.

[135] Y. Yang, T. Solis-Escalante, M. van de Ruit, F. C. T. van der Helm, and A. C. Schouten. Nonlinear Coupling between Cortical Oscillations and Muscle Activity during Isotonic Wrist Flexion. *Frontiers in Computational Neuroscience*, 10, 2016.

[136] B. J. Fischer and J. L. Peña. Optimal nonlinear cue integration for sound localization. *Journal of Computational Neuroscience*, pages 1–16, 2016.

[137] D. R. Lyamzin, S. J. Barnes, R. Donato, J. A. Garcia-Lazaro, T. Keck, and N. A. Lesica. Nonlinear Transfer of Signal and Noise Correlations in Cortical Networks. *Journal of Neuroscience*, 35(21), 2015.

[138] W. Singer and A. Lazar. Does the Cerebral Cortex Exploit High-Dimensional, Non-linear Dynamics for Information Processing? *Frontiers in Computational Neuroscience*, 10:1–10, 2016.

[139] K. D. Miller and T. W. Troyer. Neural noise can explain expansive, power-law nonlinearities in neural response functions. *Journal of Neurophysiology*, 87(2):653–9, 2002.

[140] G. Lajoie, K. K. Lin, J.-P. Thivierge, and E. Shea-Brown. Encoding in Balanced Networks: Revisiting Spike Patterns and Chaos in Stimulus-Driven Systems. *PLoS Computational Biology*, 12(12), 2016.

[141] K.-I. Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3):183–192, 1989.

[142] P. Hitzler and A. Seda. *Mathematical Aspects of Logic Programming Semantics*. CRC Press, 2016.

[143] S. H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, 2014.

[144] L. E. Forsberg, L. H. Bonde, M. A. Harvey, and P. E. Roland. The Second Spiking Threshold: Dynamics of Laminar Network Spiking in the Visual Cortex. *Frontiers in Systems Neuroscience*, 10, 2016.

[145] O. Harish and D. Hansel. Asynchronous Rate Chaos in Spiking Neuronal Circuits. *PLoS Computational Biology*, 11(7), 2015.

[146] S. Ostojic and N. Brunel. From spiking neuron models to linear-nonlinear models. *PLoS Computational Biology*, 7(1), 2011.

[147] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–9, 2008.

[148] E. N. Brown, L. M. Frank, D. Tang, M. C. Quirk, and M. A. Wilson. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, 18(18):7411–7425, 1998.

[149] O. Schwartz, J. W. Pillow, N. C. Rust, and E. P. Simoncelli. Spike-triggered neural characterization. *Journal of vision*, 6(4):13–13, 2006.

[150] A. Calabrese, J. W. Schumacher, D. M. Schneider, L. Paninski, and S. M. Woolley. A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLOS ONE*, 6(1), 2011.

[151] C. Pozzorini, S. Mensi, O. Hagens, R. Naud, C. Koch, and W. Gerstner. Automated High-Throughput Characterization of Single Neurons by Means of Simplified Spiking Models. *PLoS Computational Biology*, 11(6), 2015.

[152] D. Campagner, M. H. Evans, M. R. Bale, A. Erskine, and R. S. Petersen. Prediction of primary somatosensory neuron activity during active tactile exploration. *Elife*, 5, 2016.

[153] A. I. Weber and J. W. Pillow. Capturing the dynamical repertoire of single neurons with generalized linear models. *Neural computation*, 29(12):3260–3289, 2017.

[154] S. Deneve and M. Chalk. Efficiency turns the table on neural encoding, decoding and noise. *Current opinion in neurobiology*, 37:141–148, 2016.

[155] J. I. Glaser, A. S. Benjamin, R. Farhoodi, and K. P. Kording. The roles of supervised machine learning in systems neuroscience. *arXiv preprint arXiv:1805.08239*, 2018.

[156] A. S. Benjamin, H. Fernandes, T. Tomlinson, P. Ramkumar, C. VerSteeg, R. Chowdhury, L. E. Miller, and K. Kording. Modern machine learning as a benchmark for fitting neural responses. *Frontiers in Computational Neuroscience*, 12:56, 2018.

[157] F. Theunissen and J. P. Miller. Temporal encoding in nervous systems: a rigorous definition. *Journal of computational neuroscience*, 2(2):149–162, 1995.

[158] C. Kayser, M. A. Montemurro, N. K. Logothetis, and S. Panzeri. Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron*, 61(4):597–608, 2009.

[159] S. Panzeri, R. S. Petersen, S. R. Schultz, M. Lebedev, and M. E. Diamond. The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron*, 29 (3):769–777, 2001.

[160] M. Vinck, B. Lima, T. Womelsdorf, R. Oostenveld, W. Singer, S. Neuenschwander, and P. Fries. Gamma-phase shifting in awake monkey visual cortex. *Journal of Neuroscience*, 30(4):1250–1257, 2010.

[161] M. N. Havenith, S. Yu, J. Biederlack, N.-H. Chen, W. Singer, and D. Nikolić. Synchrony makes neurons fire in sequence, and stimulus properties determine who is ahead. *Journal of neuroscience*, 31(23):8570–8584, 2011.

[162] A. L. Jacobs, G. Fridman, R. M. Douglas, N. M. Alam, P. E. Latham, G. T. Prusky, and S. Nirenberg. Ruling out and ruling in neural codes. *Proceedings of the National Academy of Sciences*, 106(14):5936–5941, 2009.

[163] D. A. Butts, C. Weng, J. Jin, C.-I. Yeh, N. A. Lesica, J.-M. Alonso, and G. B. Stanley. Temporal precision in the neural code and the timescales of natural vision. *Nature*, 449 (7158):92, 2007.

[164] T. Gollisch and M. Meister. Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina, 2010.

[165] T. Gollisch and M. Meister. Rapid neural coding in the retina with relative spike latencies. *Science*, 319(5866):1108–1111, 2008.

[166] M. Paoli, A. Albi, M. Zanon, D. Zanini, R. Antolini, and A. Haase. Neuronal response latencies encode first odor identity information across subjects. *Journal of Neuroscience*, pages 0453–18, 2018.

[167] T. J. Gawne, T. W. Kjaer, and B. J. Richmond. Latency: another potential code for feature binding in striate cortex. *Journal of neurophysiology*, 76(2):1356–1360, 1996.

[168] S. Furukawa and J. C. Middlebrooks. Cortical representation of auditory space: information-bearing features of spike patterns. *Journal of neurophysiology*, 87(4):1749–1762, 2002.

[169] A. R. Houweling and M. Brecht. Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, 451(7174):65, 2008.

[170] M. Brecht, M. Schneider, B. Sakmann, and T. W. Margrie. Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex. *Nature*, 427(6976):704, 2004.

[171] J. Wolfe, A. R. Houweling, and M. Brecht. Sparse and powerful cortical spikes. *Current opinion in neurobiology*, 20(3):306–312, 2010.

[172] T. L. Cheng-yu, M.-m. Poo, and Y. Dan. Burst spiking of a single cortical neuron modifies global brain state. *Science*, 324(5927):643–646, 2009.

[173] Z. F. Mainen and T. J. Sejnowski. Reliability of spike timing in neocortical neurons. *Science*, 268(5216):1503–1506, 1995.

[174] J. Sjöström and W. Gerstner. Spike-timing dependent plasticity. *Spike-timing dependent plasticity*, 35(0):0–0, 2010.

[175] D. Marr and T. Poggio. From understanding computation to understanding neural circuitry, 1976.

[176] B. N. Lundstrom and A. L. Fairhall. Decoding stimulus variance from a distributional neural code of interspike intervals. *Journal of Neuroscience*, 26(35):9030–9037, 2006.

[177] M. N. Insanally, I. Carcea, R. E. Field, C. C. Rodgers, B. DePasquale, K. Rajan, M. R. DeWeese, B. F. Albanna, and R. C. Froemke. Spike-timing-dependent ensemble encoding by non-classically responsive cortical neurons. *eLife*, 8:e42409, jan 2019. ISSN 2050-084X. doi: 10.7554/eLife.42409.

[178] J. Segundo, G. Moore, L. Stensaas, and T. Bullock. Sensitivity of neurones in aplysia to temporal pattern of arriving impulses. *Journal of Experimental Biology*, 40(4):643–667, 1963.

[179] L. Kostal, P. Lansky, and J.-P. Rospars. Neuronal coding and spiking randomness. *European Journal of Neuroscience*, 26(10):2693–2701, 2007.

[180] M. A. Montemurro, M. J. Rasch, Y. Murayama, N. K. Logothetis, and S. Panzeri. Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Current biology*, 18(5): 375–380, 2008.

[181] G. Pola, A. Thiele, K. Hoffmann, and S. Panzeri. An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network-Computation in Neural Systems*, 14(1):35–60, 2003.

[182] S. Panzeri and S. R. Schultz. A unified approach to the study of temporal, correlational, and rate coding. *Neural Computation*, 13(6):1311–1349, 2001.

[183] S. Nirenberg and P. E. Latham. Decoding neuronal spike trains: How important are correlations? *Proc. Natl. Acad. Sci.*, 100(12):7348–7353, 2003.

[184] E. Schneidman, W. Bialek, and M. J. Berry. Synergy, redundancy, and independence in population codes. *Journal of Neuroscience*, 23(37):11539–11553, 2003.

[185] P. E. Latham and S. Nirenberg. Synergy, redundancy, and independence in population codes, revisited. *Journal of Neuroscience*, 25(21):5195–5206, 2005.

[186] M. DeWeese. Optimization principles for the neural code. In *Advances in neural information processing systems*, pages 281–287, 1996.

[187] S. Saxena and J. P. Cunningham. Towards the neural population doctrine. *Current Opinion in Neurobiology*, 55:103 – 111, 2019. ISSN 0959-4388. doi: https://doi.org/10.1016/j.conb.2019.02.002. Machine Learning, Big Data, and Neuroscience.

[188] A. Kohn, R. Coen-Cagli, I. Kanitscheider, and A. Pouget. Correlations and Neuronal Population Information. *Annual Review of Neuroscience*, 39(1):237–256, 2016.

[189] E. Doi and M. S. Lewicki. A simple model of optimal population coding for sensory systems. *PLoS Computational Biology*, 10(8), 2014.

[190] M. Shamir. Emerging principles of population coding: in search for the neural code. *Current opinion in neurobiology*, 25:140–148, 2014.

[191] M. Shamir and H. Sompolinsky. Nonlinear population codes. *Neural computation*, 16(6): 1105–1136, 2004.

[192] R. Follmann, C. J. Goldsmith, and W. Stein. Multimodal sensory information is represented by a combinatorial code in a sensorimotor system. *PLOS Biology*, 16(10):e2004527, 2018.

[193] L. C. Osborne, S. E. Palmer, S. G. Lisberger, and W. Bialek. The neural basis for combinatorial coding in a cortical population response. *Journal of Neuroscience*, 28(50):13522–13531, 2008.

[194] R. A. da Silveira and M. J. Berry II. High-fidelity coding with correlated neurons. *PLoS computational biology*, 10(11):e1003970, 2014.

[195] F. Franke, M. Fiscella, M. Sevelev, B. Roska, A. Hierlemann, and R. Azeredo da Silveira. Structures of neural correlation and how they favor coding. *Neuron*, 89(2):409–422, 2016.

[196] B. B. Averbeck, P. E. Latham, and A. Pouget. Neural correlations, population coding and computation. *Nature Reviews. Neuroscience*, 7(5):358–366, 2006.

[197] R. Quian Quiroga and S. Panzeri. Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews. Neuroscience*, 10(3):173–85, 2009.

[198] M. R. Cohen and A. Kohn. Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819, 2011.

[199] R. S. Petersen, S. Panzeri, and M. E. Diamond. Population coding of stimulus location in rat somatosensory cortex. *Neuron*, 32(3):503–514, 2001.

[200] F. Montani, A. Kohn, M. A. Smith, and S. R. Schultz. The role of correlations in direction and contrast coding in the primary visual cortex. *Journal of Neuroscience*, 27(9):2338–2348, 2007.

[201] S. Nirenberg, S. M. Carcieri, A. L. Jacobs, and P. E. Latham. Retinal ganglion cells act largely as independent encoders. *Nature*, 411(6838):698, 2001.

[202] Y. Dan, J.-M. Alonso, W. M. Usrey, and R. C. Reid. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nature neuroscience*, 1(6):501, 1998.

[203] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948. ISSN 0005-8580. doi: 10.1002/j.1538-7305.1948.tb01338.x.

[204] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[205] H. B. Barlow et al. Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1:217–234, 1961.

[206] F. Rieke, D. Bodnar, and W. Bialek. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. Royal Soc. A*, 262(1365): 259–265, 1995.

[207] S. Laughlin. A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung c*, 36(9-10):910–912, 1981.

[208] J. J. Atick and A. N. Redlich. What does the retina know about natural scenes? *Neural computation*, 4(2):196–210, 1992.

[209] B. A. Brinkman, A. I. Weber, F. Rieke, and E. Shea-Brown. How do efficient coding strategies depend on origins of noise in neural circuits? *PLoS Computational Biology*, 12 (10), 2016.

[210] J. Gjorgjieva, H. Sompolinsky, and M. Meister. Benefits of pathway splitting in sensory coding. *Journal of Neuroscience*, 34(36):12127–44, 2014.

[211] K. Hardcastle, N. Maheswaranathan, S. Ganguli, and L. M. Giocomo. A multiplexed, heterogeneous, and adaptive code for navigation in medial entorhinal cortex. *Neuron*, 94 (2):375–387, 2017.

[212] S. M. Chase and E. D. Young. First-spike latency information in single neurons increases when referenced to population onset. *Proceedings of the National Academy of Sciences*, 104(12):5175–5180, 2007.

[213] L. Paninski. Estimation of entropy and mutual information. *Neural computation*, 15(6): 1191–1253, 2003.

[214] S. O. Voronenko and B. Lindner. Improved lower bound for the mutual information between signal and neural spike count. *Biological cybernetics*, pages 1–16, 2018.

[215] S. P. Strong, R. Koberle, R. R. d. R. van Steveninck, and W. Bialek. Entropy and information in neural spike trains. *Physical review letters*, 80(1):197, 1998.

[216] P. Reinagel and R. C. Reid. Temporal coding of visual information in the thalamus. *Journal of Neuroscience*, 20(14):5392–5400, 2000.

[217] S. Panzeri and A. Treves. Analytical estimates of limited sampling biases in different information measures. *Network: Computation in neural systems*, 7(1):87–107, 1996.

[218] A. Borst and F. E. Theunissen. Information theory and neural coding. *Nature Neuroscience*, 2:947–957, 1999.

[219] A. B. Neiman and D. F. Russell. Sensory coding in oscillatory electroreceptors of paddlefish. *Chaos*, 21(4):047505, 2011.

[220] R. D. Vilela and B. Lindner. Comparative study of different integrate-and-fire neurons: Spontaneous activity, dynamical response, and stimulus-induced correlation. *Phys. Rev. E*, 80(3), 2009.

[221] S. G. Sadeghi, M. J. Chacron, M. C. Taylor, and K. E. Cullen. Neural variability, detection thresholds, and information transmission in the vestibular system. *Journal of Neuroscience*, 27(4):771–781, 2007.

[222] H. Clague, F. Theunissen, and J. P. Miller. Effects of adaptation on neural coding by primary sensory interneurons in the cricket cercal system. *Journal of Neurophysiology*, 77 (1):207–220, 1997.

[223] C. Pozzorini, R. Naud, S. Mensi, and W. Gerstner. Temporal whitening by power-law adaptation in neocortical neurons. *Nature Neuroscience*, 16(7):942–948, 2013.

[224] C. J. Rozell and D. H. Johnson. Examining methods for estimating mutual information in spiking neural systems. *Neurocomputing*, 65:429–434, 2005.

[225] J. C. Roddey, B. Girish, and J. P. Miller. Assessing the performance of neural encoding models in the presence of noise. *Journal of Computational Neuroscience*, 8(2):95–112, 2000.

[226] D. Bernardi and B. Lindner. A frequency-resolved mutual information rate and its application to neural systems. *Journal of Neurophysiology*, 113(5):1342–1357, 2015.

[227] N. Brenner, S. P. Strong, R. Koberle, W. Bialek, and R. R. d. R. v. Steveninck. Synergy in a neural code. *Neural computation*, 12(7):1531–1552, 2000.

[228] W. E. Skaggs, B. L. McNaughton, K. M. Gothard, and E. J. Markus. An information-theoretic approach to deciphering the hippocampal code. In *Advances in neural information processing systems*, pages 1030–1037, 1993.

[229] E. J. Markus, C. A. Barnes, B. L. McNaughton, V. L. Gladden, and W. E. Skaggs. Spatial information content and reliability of hippocampal ca1 neurons: effects of visual input. *Hippocampus*, 4(4):410–421, 1994.

[230] W. E. Skaggs, B. L. McNaughton, M. A. Wilson, and C. A. Barnes. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6(2):149–172, 1996.

[231] B. C. Souza, R. Pavao, H. Belchior, and A. B. Tort. On information metrics for spatial coding. *Neuroscience*, 375:62–73, 2018.

[232] K. Mizuseki and G. Buzsáki. Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex. *Cell reports*, 4(5):1010–1021, 2013.

[233] E. T. Rolls, A. Treves, M. J. Tovee, and S. Panzeri. Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *Journal of computational neuroscience*, 4(4):309–333, 1997.

[234] J. J. Atick. Could information theory provide an ecological theory of sensory processing? *Network: Computation in neural systems*, 3(2):213–251, 1992.

[235] M. J. Wainwright. Visual adaptation as optimal information transmission. *Vision research*, 39(23):3960–3974, 1999.

[236] M. DeWeese. *Optimization principles for the neural code*. PhD thesis, Princeton University, 1995.

[237] A. Destexhe and D. Paré. Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *Journal of neurophysiology*, 81(4):1531–1547, 1999.

[238] E. V. Evarts. Temporal patterns of discharge of pyramidal tract neurons during sleep and waking in the monkey. *Journal of neurophysiology*, 27(2):152–171, 1964.

[239] A. Baranyi, M. B. Szente, and C. D. Woody. Electrophysiological characterization of different types of neurons recorded in vivo in the motor cortex of the cat. ii. membrane parameters, action potentials, current-induced voltage responses and electrotonic structures. *Journal of neurophysiology*, 69(6):1865–1879, 1993.

[240] M. Matsumura, T. Cope, and E. Fetz. Sustained excitatory synaptic input to motor cortex neurons in awake animals revealed by intracellular recording of membrane potentials. *Experimental brain research*, 70(3):463–469, 1988.

[241] C. Woody and E. Gruen. Characterization of electrophysiological properties of intracellularly recorded neurons in the neocortex of awake cats: a comparison of the response to injected current in spike overshoot and undershoot neurons. *Brain research*, 158(2): 343–357, 1978.

[242] M. Rudolph, M. Pospischil, I. Timofeev, and A. Destexhe. Inhibition determines membrane potential dynamics and controls action potential generation in awake and sleeping cat cortex. *Journal of neuroscience*, 27(20):5280–5290, 2007.

[243] A. Renart, J. De La Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris. The asynchronous state in cortical circuits. *Science*, 327(5965):587–590, 2010.

[244] E. M. Tartaglia and N. Brunel. Bistability and up/down state alternations in inhibition-dominated randomly connected networks of lif neurons. *Scientific reports*, 7(1):11916, 2017.

[245] A. Renart, N. Brunel, and X.-J. Wang. Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks. *Computational neuroscience: A comprehensive approach*, pages 431–490, 2004.

[246] A. N. Burkitt. A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological cybernetics*, 95(1):1–19, 2006.

[247] B.-h. Liu, P. Li, Y.-t. Li, Y. J. Sun, Y. Yanagawa, K. Obata, L. I. Zhang, and H. W. Tao. Visual receptive field structure of cortical inhibitory neurons revealed by two-photon imaging guided recording. *Journal of Neuroscience*, 29(34):10520–10532, 2009.

[248] D. D. Stettler and R. Axel. Representations of odor in the piriform cortex. *Neuron*, 63(6): 854–864, 2009.

[249] M. J. Higley and D. Contreras. Balanced excitation and inhibition determine spike timing during frequency adaptation. *Journal of Neuroscience*, 26(2):448–457, 2006.

[250] J. S. Anderson, M. Carandini, and D. Ferster. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *Journal of Neurophysiology*, 84(2): 909–926, 2000.

[251] J. A. Cardin, L. A. Palmer, and D. Contreras. Cellular mechanisms underlying stimulus-dependent gain modulation in primary visual cortex neurons in vivo. *Neuron*, 59(1): 150–160, 2008.

[252] A. Y. Tan, Y. Chen, B. Scholl, E. Seidemann, and N. J. Priebe. Sensory stimulation shifts visual cortex from synchronous to asynchronous states. *Nature*, 509(7499):226, 2014.

[253] I. M. Finn, N. J. Priebe, and D. Ferster. The emergence of contrast-invariant orientation tuning in simple cells of cat visual cortex. *Neuron*, 54(1):137–152, 2007.

[254] G. Hennequin, Y. Ahmadian, D. B. Rubin, M. Lengyel, and K. D. Miller. The dynamical regime of sensory cortex: Stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron*, 98(4):846–860, 2018.

[255] M. Volgushev, J. Pernberg, and U. T. Eysel. γ-frequency fluctuations of the membrane potential and response selectivity in visual cortical neurons. *European Journal of Neuroscience*, 17(9):1768–1776, 2003.

[256] T. Hoch, S. Volgushev, A. Malyshev, K. Obermayer, and M. Volgushev. Modulation of the amplitude of γ-band activity by stimulus phase enhances signal encoding. *European Journal of Neuroscience*, 33(7):1223–1239, 2011.

[257] C. D. Harvey, F. Collman, D. A. Dombeck, and D. W. Tank. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature*, 461(7266):941–946, 2009.

[258] S. Mensi, O. Hagens, W. Gerstner, and C. Pozzorini. Enhanced sensitivity to rapid input fluctuations by nonlinear threshold dynamics in neocortical pyramidal neurons. *PLoS Computational Biology*, 12(2):1–38, 2016.

[259] M. Arsiero, H.-R. Lüscher, B. N. Lundstrom, and M. Giugliano. The impact of input fluctuations on the frequency–current relationships of layer 5 pyramidal neurons in the rat medial prefrontal cortex. *Journal of Neuroscience*, 27(12):3274–3284, 2007.

[260] N. Brunel, F. S. Chance, N. Fourcaud, and L. F. Abbott. Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Physical Review Letters*, 86(10): 2186–2189, 2001.

[261] J. Pressley and T. W. Troyer. Complementary responses to mean and variance modulations in the perfect integrate-and-fire model. *Biological Cybernetics*, 101(1):63–70, 2009.

[262] B. Naundorf, F. Wolf, and M. Volgushev. Unique features of action potential initiation in cortical neurons. *Nature*, 440(7087):1060, 2006.

[263] D. Linaro, I. Biró, and M. Giugliano. Dynamical response properties of neocortical neurons to conductance-driven time-varying inputs. *European Journal of Neuroscience*, 47(1):17–32, 2018.

[264] W. J. Ma, J. M. Beck, P. E. Latham, and A. Pouget. Bayesian inference with probabilistic population codes. *Nature neuroscience*, 9(11):1432, 2006.

[265] M. E. Mazurek and M. N. Shadlen. Limits to the temporal fidelity of cortical spike rate signals. *Nature neuroscience*, 5(5):463, 2002.

[266] C. Geisler, N. Brunel, and X.-J. Wang. Contributions of intrinsic membrane dynamics to fast network oscillations with irregular neuronal discharges. *Journal of neurophysiology*, 94(6):4344–4361, 2005.

[267] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.

[268] K. Jacobs. *Stochastic processes for physicists: understanding noisy systems*. Cambridge University Press, 2010.

[269] T. A. Engel, N. A. Steinmetz, M. A. Gieselmann, A. Thiele, T. Moore, and K. Boahen. Selective modulation of cortical state during spatial attention. *Science*, 354(6316):1140–1144, 2016.

[270] G. Deco, D. Martí, A. Ledberg, R. Reig, and M. V. S. Vives. Effective reduced diffusion-models: a data driven approach to the analysis of neuronal dynamics. *PLoS computational biology*, 5(12):e1000587, 2009.

[271] D. T. Gillespie. Exact numerical simulation of the ornstein-uhlenbeck process and its integral. *Physical review E*, 54(2):2084, 1996.

[272] S. Primak, V. Lyandres, and V. Kontorovich. Markov models of non-gaussian exponentially correlated processes and their applications. *Physical Review E*, 63(6):061103, 2001.

[273] I. Hunter and R. Kearney. Generation of random sequences with jointly specified probability density and autocorrelation functions. *Biological cybernetics*, 47(2):141–146, 1983.

[274] B. Lindner, L. Schimansky-Geier, and A. Longtin. Maximizing spike train coherence or incoherence in the leaky integrate-and-fire model. *Physical Review E*, 66(3):031916, 2002.

[275] S. O. Voronenko and B. Lindner. Weakly nonlinear response of noisy neurons. *New Journal of Physics*, 19(3):033038, 2017.

[276] S. R. Schultz and S. Panzeri. Temporal correlations and neural spike train entropy. *Physical Review Letters*, 86(25):5823, 2001.

[277] D. G. Watts and G. Jenkins. *Spectral Analysis and its applications*. Oakland: Holden-Day, 1968.

[278] X.-J. Wang, Y. Liu, M. V. Sanchez-Vives, and D. A. McCormick. Adaptation and temporal decorrelation by single neurons in the primary visual cortex. *Journal of neurophysiology*, 89(6):3279–3293, 2003.

[279] M. S. Goldman, P. Maldonado, and L. Abbott. Redundancy reduction and sustained firing with stochastic depressing synapses. *Journal of Neuroscience*, 22(2):584–591, 2002.

[280] N. Spruston, P. Jonas, and B. Sakmann. Dendritic glutamate receptor channels in rat hippocampal ca3 and ca1 pyramidal neurons. *The Journal of physiology*, 482(2):325–352, 1995.

[281] S. Panzeri, S. R. Schultz, A. Treves, and E. T. Rolls. Correlations and the encoding of information in the nervous system. *Proc. Biol. Sci.*, 266(1423):1001–12, 1999.

[282] M. A. Montemurro, R. Senatore, and S. Panzeri. Tight data-robust bounds to mutual information combining shuffling and model selection techniques. *Neural Computation*, 19 (11):2913–2957, 2007.

[283] G. Pola, R. S. Petersen, A. Thiele, M. P. Young, and S. Panzeri. Data-robust tight lower bounds to the information carried by spike times of a neuronal population. *Neural Computation*, 17(9):1962–2005, 2005.

[284] A. S. Ecker, P. Berens, G. A. Keliris, M. Bethge, N. K. Logothetis, and A. S. Tolias. Decorrelated neuronal firing in cortical microcircuits. *Science*, 327(5965):584–7, 2010.

[285] C. Bauermeister, T. Schwalger, D. F. Russell, A. B. Neiman, and B. Lindner. Characteristic effects of stochastic oscillatory forcing on neural firing: analytical theory and comparison to paddlefish electroreceptor data. *PLoS computational biology*, 9(8):e1003170, 2013.

[286] R. Kubo. Statistical-mechanical theory of irreversible processes. I. general theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Japan*, 12(6): 570–586, 1957.

[287] H. Risken. Fokker-planck equation. In *The Fokker-Planck Equation*, pages 63–95. Springer, 1996.

[288] B. Lindner, M. J. Chacron, and A. Longtin. Integrate-and-fire neurons with threshold noise: A tractable model of how interspike interval correlations affect neuronal signal transmission. *Phys. Rev. E*, 72(2):021911, 2005.

[289] S. J. Tripathy, J. Savitskaya, S. D. Burton, N. N. Urban, and R. C. Gerkin. Neuroelectro: a window to the world's neuron electrophysiology data. *Frontiers in neuroinformatics*, 8: 40, 2014.

[290] L. G. Nowak, R. Azouz, M. V. Sanchez-Vives, C. M. Gray, and D. A. McCormick. Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. *Journal of Neurophysiology*, 89(3):1541–1566, 2003.

[291] F. Barbieri and N. Brunel. Can attractor network models account for the statistics of firing during persistent activity in prefrontal cortex? *Frontiers in neuroscience*, 2:3, 2008.

[292] D. J. Amit and N. Brunel. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 7(3):237–252, 1997.

[293] A. Roxin, N. Brunel, D. Hansel, G. Mongillo, and C. van Vreeswijk. On the distribution of firing rates in networks of cortical neurons. *Journal of Neuroscience*, 31(45):16217–16226, 2011.

[294] A. Kuhn, A. Aertsen, and S. Rotter. Neuronal integration of synaptic input in the fluctuation-driven regime. *Journal of Neuroscience*, 24(10):2345–2356, 2004.

[295] B. Naundorf, T. Geisel, and F. Wolf. Action potential onset dynamics and the response speed of neuronal populations. *Journal of Computational Neuroscience*, 18(3):297–309, 2005.

[296] H. Ko, L. Cossell, C. Baragli, J. Antolik, C. Clopath, S. B. Hofer, and T. D. Mrsic-Flogel. The emergence of functional microcircuits in visual cortex. *Nature*, 496(7443):96–100, 2013.

[297] J. de la Rocha, B. Doiron, E. Shea-Brown, K. Josić, and A. D. Reyes. Correlation between neural spike trains increases with firing rate. *Nature*, 448(7155):802–6, 2007.

[298] A. Compte, R. Reig, V. F. Descalzo, M. A. Harvey, G. D. Puccini, and M. V. Sanchez-Vives. Spontaneous high-frequency (10–80 hz) oscillations during up states in the cerebral cortex in vitro. *Journal of Neuroscience*, 28(51):13828–13844, 2008.

[299] A. Gupta, Y. Wang, and H. Markram. Organizing principles for a diversity of gabaergic interneurons and synapses in the neocortex. *Science*, 287(5451):273–278, 2000.

[300] P. F. Craigmile. Simulating a class of stationary gaussian processes using the davies–harte algorithm, with application to long memory processes. *J Time Ser Anal*, 24(5):505–511, 2003.

[301] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

[302] D. R. Brillinger. *Time series: data analysis and theory*, volume 36. Siam, 1981.

[303] A. Gut. *The Multivariate Normal Distribution*, pages 117–145. Springer New York, New York, NY, 2009. ISBN 978-1-4419-0162-0. doi: 10.1007/978-1-4419-0162-0_5.

[304] R. d'Agostino and E. S. Pearson. Tests for departure from normality. *Biometrika*, 60(3): 613–622, 1973.

[305] M. Natrella. *NIST/SEMATECH e-handbook of statistical methods*. NIST/SEMATECH, 2010.

[306] K. V. Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530, 1970.

[307] K. V. Mardia. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 115–128, 1974.

[308] H. C. Thode. *Testing for normality*, volume 164. CRC press, 2002.

[309] M. K. Cain, Z. Zhang, and K.-H. Yuan. Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior research methods*, 49(5):1716–1735, 2017.

[310] M. Zhou and Y. Shao. A powerful test for multivariate normality. *Journal of applied statistics*, 41(2):351–363, 2014.

[311] Z. Hanusz, R. Enomoto, T. Seo, and K. Koizumi. A Monte Carlo comparison of Jarque–Bera type tests and Henze–Zirkler test of multivariate normality. *Communications in Statistics-Simulation and Computation*, 47(5):1439–1452, 2018.

[312] N. Henze and B. Zirkler. A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, 19(10):3595–3617, 1990.

[313] R. Vallat. Pingouin: statistics in Python. *The Journal of Open Source Software*, 3(31): 1026, Nov. 2018.

[314] S. Korkmaz, D. Goksuluk, and G. Zararsiz. MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2):151–162, 2014.

[315] E. W. Weisstein. *Normal product distribution.* Wolfram Research, Inc., 2003.

[316] L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918.

[317] M. H. Rowe and A. B. Neiman. Information analysis of posterior canal afferents in the turtle, Trachemys scripta elegans. *Brain Research*, 1434:226–42, 2012.

[318] D. R. Chialvo, A. Longtin, and J. Müller-Gerking. Stochastic resonance in models of neuronal ensembles. *Phys. Rev. E*, 55(2):1798, 1997.

[319] L. I. Zhang, A. Y. Tan, C. E. Schreiner, and M. M. Merzenich. Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature*, 424(6945):201, 2003.

[320] D. J. MacKay. *Information theory, inference and learning algorithms.* Cambridge university press, 2003.

[321] M. J. Chacron. Nonlinear information processing in a model sensory system. *Journal of Neurophysiology*, 95(5):2933–2946, 2006.

[322] A. Y. Tan. Spatial diversity of spontaneous activity in the cortex. *Frontiers in Neural Circuits*, 9:48, 2015.

[323] M. C. Angulo, J. Rossier, and E. Audinat. Postsynaptic glutamate receptors and integrative properties of fast-spiking interneurons in the rat neocortex. *Journal of Neurophysiology*, 82(3):1295–1302, 1999.

[324] A. C. Flint, U. S. Maisch, J. H. Weishaupt, A. R. Kriegstein, and H. Monyer. Nr2a subunit expression shortens nmda receptor synaptic currents in developing neocortex. *Journal of Neuroscience*, 17(7):2469–2476, 1997.

[325] A. Pregowska, E. Kaplan, and J. Szczepanski. How Far can Neural Correlations Reduce Uncertainty? Comparison of Information Transmission Rates for Markov and Bernoulli Processes. *International Journal of Neural Systems*, 2019. doi: 10.1142/s0129065719500035.

[326] E. Piasini and S. Panzeri. Information theory in neuroscience. *Entropy*, 2019.

[327] J. Bauermann and B. Lindner. Multiplicative noise is beneficial for the transmission of sensory signals in simple neuron models. *Biosystems*, 2019.

[328] J. A. Berkowitz and T. O. Sharpee. Quantifying information conveyed by large neuronal populations. *Neural Computation*, pages 1–33, 2019.

[329] M. London, A. Roth, L. Beeren, M. Häusser, and P. E. Latham. Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*, 466(7302):123, 2010.

[330] T. R. Stanford, S. Shankar, D. P. Massoglia, M. G. Costello, and E. Salinas. Perceptual decision making in less than 30 milliseconds. *Nature Neuroscience*, 13(3):379–385, 2010.

[331] M. Huang, M. Volgushev, and F. Wolf. A small fraction of strongly cooperative sodium channels boosts neuronal encoding of high frequencies. *PLOS ONE*, 7(5):e37629, 2012.

[332] V. Ilin, A. Malyshev, F. Wolf, and M. Volgushev. Fast computations in cortical ensembles require rapid initiation of action potentials. *Journal of Neuroscience*, 33(6):2281–2292, 2013.

[333] M. G. Metzen, O. Ávila-Åkerberg, and M. J. Chacron. Coding stimulus amplitude by correlated neural activity. *Physical Review E*, 91(4):042717, 2015.

[334] A. Bulsara, E. Jacobs, T. Zhou, F. Moss, and L. Kiss. Stochastic resonance in a single neuron model: Theory and analog simulation. *Journal of Theoretical Biology*, 152(4): 531–555, 1991.

[335] W. F. Mynarski and A. M. Hermundstad. Adaptive coding for dynamic sensory inference. *eLife*, 7:e32055, jul 2018. ISSN 2050-084X. doi: 10.7554/eLife.32055.

[336] L. Aitchison and M. Lengyel. With or without you: predictive coding and bayesian inference in the brain. *Current opinion in neurobiology*, 46:219–227, 2017.

[337] K. Doya, S. Ishii, A. Pouget, and R. P. Rao. *Bayesian brain: Probabilistic approaches to neural coding*. MIT press, 2007.

[338] G. Orbán, P. Berkes, J. Fiser, and M. Lengyel. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, 92(2):530–543, 2016.

[339] W. Gerstner and R. Naud. How good are neuron models? *Science*, 326(5951):379–380, 2009.

[340] R. Naud and W. Gerstner. Can we predict every spike? *Spike Timing: Mechanisms and Function*, pages 65–76, 2013.

[341] R. Brette. What is the most realistic single-compartment model of spike initiation? *PLoS Computational Biology*, 11(4):1–13, 2015.

[342] R. Brette and W. Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of neurophysiology*, 94(5):3637–3642, 2005.

[343] M. Honegger. Shedding light on black box machine learning algorithms: Development of an axiomatic framework to assess the quality of methods that explain individual predictions. *arXiv preprint arXiv:1808.05054*, 2018.

[344] T. Akam and D. M. Kullmann. Oscillatory multiplexing of population codes for selective communication in the mammalian brain. *Nature Reviews Neuroscience*, 15(2):111, 2014.

[345] D. Hansel and H. Sompolinsky. Chaos and synchrony in a model of a hypercolumn in visual cortex. *Journal of computational neuroscience*, 3(1):7–34, 1996.

[346] D. Hansel and H. Sompolinsky. Synchronization and computation in a chaotic neural network. *Physical Review Letters*, 68(5):718, 1992.

[347] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1965.

Curriculum Vitae

Personal information

Name:	Tim Herfurth
Date of birth:	Dec 17th, 1987
Place of birth:	Berlin-Kreuzberg, Germany
Citizenship:	German
E-mail:	timherfurth87@gmail.com

Education

06/2007	Abitur, Adolf-Reichwein-Schule, Neu-Anspach, Germany.
10/2008-09/2011	Bachelor of Science in Physics, Goethe-University, Frankfurt, Germany.
10/2011-04/2014	Master of Science in Physics, Goethe-University, Frankfurt, Germany.
since 06/2014	Doctoral studies, Max Planck Institute for Brain Research, Frankfurt, Germany.
since 10/2015	International Max Planck Research School for Neural Circuits, Max Planck Institute for Brain Research, Frankfurt, Germany.