Mapping spiking network simulations to mathematically tractable rate models

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Pierre Ekelmans

aus

Brüssel, Belgien

Bonn 2023

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachterin/Betreuerin: Prof. Dr. Tatjana Tchumatchenko

Gutachter: Prof. Dr. Jan Hasenauer

Tag der Promotion: 19.04.2024

Erscheinungsjahr: 2024

Summary

Biological neural systems have evolved to give rise to complex cognitive functions that allow animals to interact with their environment, drive their behavior, and regulate physiological processes. These cognitive functions emerge from the processing of information fluxes in the form of discrete neuronal spikes. The operations on these signals are carried out by a complex network of interactions between neurons. Over the last century, extensive research efforts have led to a solid understanding of the morphology and physiology of neurons, the diversity of cell types, the detailed wiring of neural networks, and the main pathways along which information is transmitted across the brain. Despite this in-depth knowledge, the mechanisms underlying the emergence of cognitive functions remain poorly understood. In particular, an algorithmic characterization of the operations carried out by neural systems has only been achieved for the simplest circuits.

Because of the staggering complexity of the brain's circuitry, neuroscientists developed theoretical models and numerical simulations through which the emergent properties of networks can be studied. These models are diverse and reflect a trade-off between the level of detail in which the neural network is characterized and the interpretability of the framework. On the one hand, the activity regime in a neural network can be studied through a detailed spiking simulation of all neurons, given their electrophysiological properties, connectivity, and the external stimulation they receive. This approach can provide a realistic reflection of biological neural networks and generate precise predictions of the resulting firing patterns for any given parameters. Still, the high level of detail makes simulations heavy to perform and their outcome challenging to generalize. On the other hand, rate models characterize the global properties of a network mathematically and allow for a thorough analysis of the effect of its parameters. Nonetheless, the assumptions underlying the mathematical expression of the model impose constraints on its domain of validity and restrict the dynamics of its resulting firing activity.

In this work, I aim to bridge the gap between simulations of spiking networks and mathematically tractable rate models. In particular, the focus is set on simulations of Excitatory-Inhibitory (E-I) networks of Leaky Integrate-and-Fire (LIF) neurons. These simulations are commonly used in computational neuroscience to model neural systems. Yet, the mathematical study of such systems typically relies on intractable equations which require numerical methods to be

manipulated. I show that the activation function of LIF neurons can be well approximated with a power law within the range of biologically relevant firing rates. Following this, the LIF spiking network can be characterized by an adapted version of the Stabilized Supralinear Network (SSN) rate model, using parameters obtained by fitting the neurons' transfer function. The activity regimes of LIF spiking networks match the prediction of the SSN, both qualitatively and quantitatively. This is shown using parameters extracted from databases of state-of-the-art experimentally acquired data, from which biologically realistic E-I networks can be defined.

Besides the accuracy of its prediction, a major benefit of the SSN is the simplicity of its mathematical formulation. Thanks to this, it can be studied analytically to explore the diversity of nonlinear activity regimes accessible to two-population LIF systems. In particular, I show it can be used to derive exact conditions under which specific nonlinear transformations occur. These diverse nonlinear regimes are supported by the intrinsically supralinear activation function of LIF neurons and can be leveraged to perform information processing operations at the level of local circuits.

Finally, I extend the domain of validity of the SSN rate model by demonstrating how it can be modified to account for diversity in the properties of neurons within populations. This illustrates the flexibility of the SSN, as it can easily be tweaked to capture the activity of spiking networks with diverse properties.

Overall, due to its accuracy and mathematical tractability, the SSN appears to be a powerful tool to support the study of neural networks. It reduces the complexity of neural network analysis by shifting the focus to a higher level, towards the transformations operated in local networks rather than individual neurons. Following this approach, local E-I circuits can be treated as functional computing units whose transfer functions can be mathematically predicted from their connectivity.

Zusammenfassung

Deutsche Übersetzung

Biologisch-neuronale Systeme haben durch ihre Entwicklung komplexe kognitive Funktionen angenommen, die es Lebewesen ermöglichen, mit ihrer Umwelt zu interagieren, ihr Verhalten zu steuern und physiologische Prozesse zu regulieren. Diese kognitiven Funktionen ergeben sich aus Informationsverarbeitung in Form diskreter neuronaler Aktionspotentiale. Die Verarbeitung dieser Signale erfolgt über ein komplexes Netz neuronaler Interaktionen. Im Laufe des letzten Jahrhunderts haben umfangreiche Forschungsanstrengungen zu einem detaillierten Verständnis der neuronalen Morphologie und Physiologie, der Vielfalt neuronaler Zelltypen, der Verknüpfung neuronaler Netze und der Verbindungen, über die Informationen im Gehirn übertragen werden , geführt. Trotz dieses fundierten Wissens sind die Mechanismen, die der Entstehung kognitiver Funktionen zugrunde liegen, nach wie vor nur unzureichend verstanden. Insbesondere ist eine algorithmische Charakterisierung umgesetzter Rechenoperationen bisher nur für einfachste neuronale Schaltkreise möglich.

Zur Untersuchung der Eigenschaften verschiedenster neuronaler Schaltkreise haben Neurowissenschaftler theoretische Modelle und numerische Simulationen entwickelt. Diese Modelle sind äußerst vielfältig, wobei ein jedes einen Kompromiss zwischen Detailgrad und Interpretierbarkeit abbildet. Einerseits ist es möglich, das Aktivitätsregime eines neuronalen Netzes mittels einer detaillierten Spiking-Simulation aller Neurone unter Berücksichtigung ihrer elektrophysiologischen Eigenschaften, ihrer Konnektivität und ihrer externen Inputs zu untersuchen. Dieser Ansatz kann realistische und präzise Vorhersagen der sich ergebenden Aktivitätsmuster für beliebige Parameter liefern. Aufgrund des hohen Detailgrads sind diese Simulationen jedoch rechenintensiv und ihre Ergebnisse schwer zu verallgemeinern. Einfachere Feuerratenmodelle charakterisieren die globalen Eigenschaften eines neuronalen Netzes mathematisch und ermöglichen eine gründliche Analyse der Auswirkungen seiner Parameter. Die Annahmen, die der mathematischen Abbildung dieser Modelle zugrunde liegen, schränken jedoch ihren Gültigkeitsbereich ein und begrenzen die Dynamik der simulierten Aktivitätsmuster.

In dieser Dissertation strebe ich an, die Lücke zwischen Spiking-Modellen und mathematisch nachvollziehbareren Feuerratenmodellen zu schließen. Insbesondere widme ich mich der Simulation exzitatorisch-inhibitorischer (E-I) Netzwerke aus Leaky Integrate-and-Fire (LIF)

Neuronen. Diese Simulationen werden in der theoretischen Neurowissenschaft häufig zur Modellierung neuronaler Netzwerke verwendet. Die mathematische Untersuchung solcher Systeme beruht jedoch in der Regel auf schwer lösbaren Gleichungen, die nur mit numerischen Methoden bearbeitet werden können. Ich demonstriere, dass die Aktivierungsfunktion von LIF-Neuronen innerhalb des Bereichs biologisch realistischer Aktivitätsraten gut durch ein Potenzgesetz approximiert werden kann. So kann das LIF- Netzwerk durch eine angepasste Version eines Ratenmodells, des Stabilized Supralinear Networks (SSN), charakterisiert werden, wobei Parameter verwendet werden, die sich aus der neuronalen Transferfunktion ergeben. Anhand von Parametern, welche aus Datenbanken aktueller experimentell erfasster Daten extrahiert wurden, definiere ich biologisch realistische E-I LIF-Netzwerke und demonstriere, dass ihr Aktivitätsregime sowohl qualitativ als auch quantitativ mit Vorhersagen des SSN übereinstimmt. Neben der Genauigkeit der Vorhersage liegt ein großer Vorteil des SSN in der Einfachheit seiner mathematischen Formulierung. Dank dieser kann es analytisch untersucht werden, um die Vielfalt des nichtlinearen Aktivitätsregimes zu erforschen, die LIF-Systemen mit zwei Populationen zugänglich ist. Insbesondere zeige ich, dass sich genaue Bedingungen ableiten lassen, unter denen bestimmte nichtlineare Transformationen auftreten. Dieses vielfältige nichtlineare Regime wird durch die intrinsisch supralineare Aktivierungsfunktion der LIF-Neuronen gestützt und kann für Informationsverarbeitungsoperationen auf der Ebene lokaler Schaltkreise genutzt werden.

Weiterhin erweitere ich das SSN-Ratenmodell, indem ich Modifikationen aufzeige, welche die Vielfalt neuronaler Eigenschaften innerhalb neuronaler Populationen berücksichtigen. Dies verdeutlicht die Flexibilität des SSN, da es leicht modifizierbar ist, um die Aktivität verschiedenster Spiking-Modelle abzubilden.

Insgesamt erweist sich das SSN aufgrund seiner Genauigkeit und mathematischen Berechenbarkeit als ein leistungsfähiges Instrument zur Untersuchung neuronaler Netzwerke. Es reduziert die Komplexität der Analyse neuronaler Netze, indem es den Schwerpunkt auf eine höhere Ebene verlagert, nämlich auf die Transformationen, die in lokalen Netzwerken und nicht in einzelnen Neuronen stattfinden. Diesem Ansatz folgend können lokale E-I-Schaltkreise als funktionale Recheneinheiten behandelt werden, deren Transferfunktionen mathematisch aus ihrer Konnektivität vorhergesagt werden können.

Acknowledgements

I would like to thank everyone who has provided me with support for this work. First and foremost, my PhD advisor, Prof. Dr. Tatjana Tchumatchenko for the opportunity to work on this project, her academic guidance, and the freedom I had in my research. I am also very grateful to Dr. Nataliya Kraynyukova for her scientific supervision.

I want to thank the entire Tchumatchenko lab - past and present - for the friendly, stimulating, and supportive work environment I have enjoyed over the years. In particular, I want to highlight Dr. Andreas Nold whose sharp insight helped shape this project and Laura Bernáez Timón whom I had the pleasure of working closely with on multiple projects. I also want to point out Alexandra Vormberg, Dr. Max Eggl, Carlos Wert, Dr. Simon Renner and Dr. Tobias Mistele who have helped at various stages of this project.

I am very grateful for the many scientific institutions that have hosted me over the course of my nomadic PhD across Germany: MPIBR, FIAS, Uni Mainz and uni Bonn. I also want to acknowledge the graduate schools IMPRS and CMMS for the resources and training they provided me. I am especially grateful to the many funding sources that supported my research in a very practical sense: MPG, CRC1080 and LOEWE.

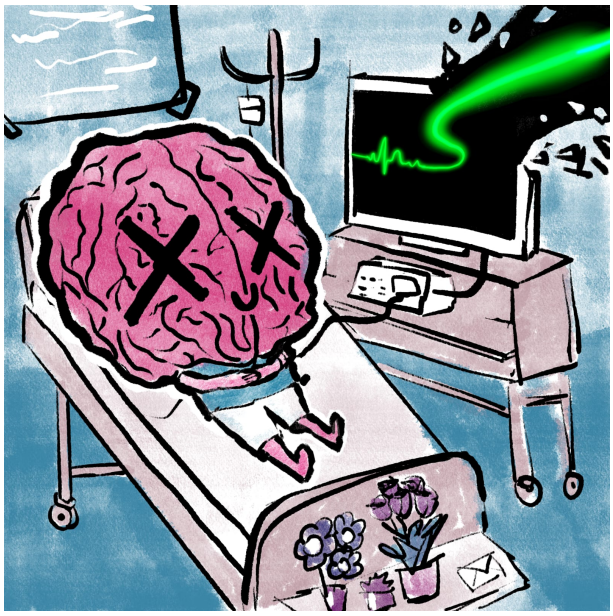
Regarding the preparation of my dissertation, I want to thank Dr. Danylo Batulin, Janko Petkovic, Petros Chalas, Petros Vlachos, Albert Miguel, Kanaan Mousaei, Ulzii Narantsasralt and Antoni Bertolín for their proofreading efforts. I am also very grateful to Marlene Aßfalg and Dr. Benjamin Kausch-Blecken von Schmeling for their help with translation. I want to acknowledge Jorg Wenniges from the Promotionsbüro for guiding me smoothly through all administrative processes.

On a more personal level, I want to express my gratitude to my family and friends. In particular, to my parents and my sister, who have continuously been supportive throughout all stages of my life. Finally, I would like to acknowledge the vital support I received from the many friends I made during my time in Germany. In particular, the inclusive Mont Blanc: Mahyar, Johanna, Ava, and Jan. The board games group: Ben, Fabio, Theo, Lisandro, Albert, Nicklas, Sigrid, Lucas, and Petros. And in no particular order, Danylo, Sophia, Michelle, Yağmur, and Mathieu.

You do not only write your PhD thesis for the doctoral committee. Maybe some years from now, a new student like you starting their PhD in Slovenia will find your thesis. This will be a gold mine of valuable information for them.

Prof. Dr. Franziska Matthäus

This thesis is dedicated to Slovenian Pierre.



Art by Mathieu Renard

Contents

1. Introduction

1.1. Biological neural networks

Single-celled organisms are capable of simple behaviors such as sensing the environment, directed movement, and metabolic regulation. These processes are controlled through complex regulatory pathways involving biochemical signaling cascades. In the case of multi-cellular organisms, coordination across the entire organism can still be carried out through chemical signaling, but the slow and untargeted diffusive transfer becomes limiting. In order to gain greater control and speed over this coordination, a new specialized cell type appeared in animals: the neuron. In fact, recent genetic analyses suggest that the apparition of neurons might have occurred multiple times over the history of animal evolution [157]. These early nervous systems could have been involved in the coordination of motor action, homeostatic regulation, sensory perception, or immune response [9].

The early forms of the nervous system were unstructured and formed neural nets across the surface of the organism. Throughout evolution, new species appeared whose anatomy was organized around specialized structures and organs. The process of cephalization led to the development of dense neural circuits and brains [148]. As brains grew more diverse and complex, they formed distinct anatomic structures or brain regions that specialize in particular functions. Thanks to this, the repertoire of cognitive functions realized by nervous systems grew much richer, including memory, learning, spatial navigation, planning, object recognition, problem-solving, emotional states, etc. The execution and interplay of such cognitive functions drive the behavioral output of animals in response to their environment and are therefore subjected to evolutionary pressure [178].

Ultimately, the operations carried out by nervous systems result from the interactions of neurons wired into complex networks that have been shaped throughout evolutionary history.

1.1.1. Neuronal physiology

Neurons are excitable and polarised cells as they can be stimulated to generate an electrical signal that propagates unidirectionally. The core of the cell is its cell body, also named soma, which contains the nucleus and most organelles necessary for cellular functions. A distinct characteristic of neurons is the presence of processes protruding out of the cell body: the dendrites and the axon. The dendrites form a complex branching structure that serves the function of integrating multiple signals received by the neuron and propagating them toward the cell body. There, if the summed contribution of these signals is sufficiently strong, the neuron will generate its own electrical impulse, which will propagate down a specialized process, the axon, leading to the stimulation of subsequent neurons (Fig. 1.1.A).

These signals take the form of changes in the electric potential across the cell membrane. In the absence of any stimulation, the Na^+/K^+ ion pumps maintain the inside of the cell at a lower electric potential than the extracellular medium, around $-65\,\mathrm{mV}$ [86]. Excitatory stimulations lead to a net influx of current, resulting in depolarization of the cell membrane, while inhibitory stimulations cause hyperpolarization. The overall sum of dendritic currents will modify the membrane potential in the soma and the axon initial segment (AIS), where a high density of voltage-gated ion channels can lead to the generation of an action potential, also named spike. If the membrane potential in the AIS reaches a threshold value near $-50\,\mathrm{mV}$ [86], voltage-gated Na^+ channels will open and lead to a large influx of current, which further amplifies the depolarization. As the membrane potential increases, the voltage-gated Na^+ channels inactivate while a second slower type of channel, the voltage-gated K^+ channels, cause an outflux of current, restoring a negative membrane potential. This mechanism allows for the amplification of the electric signal and the generation of a stereotypical response: the action potential (Fig. 1.1.B). Furthermore, since the generation of an action potential is triggered by an initial depolarization, it can propagate along the axon as nearby voltage-gated channels will likewise respond to the local change in potential caused by the adjacent depolarization. In order to accelerate the propagation of the action potential, the axon can be interspaced with insulating myelin coating to force a saltatory conduction of the signal. At the axon terminal, the electric signal caused by the spike triggers the activation of axonal boutons, which form synaptic connections with different neurons or other excitable cells.

1.1.2. Synaptic transmission

Synapses are specialized structures where the communication between neurons occurs. The typical synaptic structure consists of a presynaptic axonal bouton from which the signal is emitted if the presynaptic neuron spikes and a postsynaptic dendritic spine, which will receive the signal, inducing a change of postsynaptic membrane potential (PSP). A tight gap separates the cells: the synaptic cleft (Fig. 1.1.C). Once a presynaptic spike reaches the axon terminal, it activates voltage-gated Ca^{2+} channels, leading to the exocytosis of vesicles containing neurotransmitters into the synaptic cleft. These molecules can then diffuse across the cleft and reach specific receptors on the postsynaptic side. Once activated, these receptors enable the flow of ions across the membrane, either directly through a conformation change or indirectly by releasing secondary messengers, which then mediate the opening of specific ion channels. Depending on the nature of the neurotransmitter, different ion channels will be activated, leading to the influx of either a positive current (Na^+, depolarization) or a negative current (Cl^-, hyperpolarization). This local change in the electrical property of the cell will diffuse along the dendrite, and the summed effect of all synapses throughout the neuronal morphology will be sensed in the AIS.

As the current flows along the dendrite, diverse phenomena can alter the electric signal and modify the resulting effect at the AIS. One such effect is attenuation, which causes the voltage signal to weaken as it travels away from the synapse due to longitudinal electric resistance along the dendritic path. Attenuation makes distal synapses less effective than proximal ones. Similarly, the temporal patterns of synaptic activation will also matter as the signal originating from more distant dendrites will take longer to reach the soma. Since multiple depolarizing signals must arrive simultaneously in the soma to elicit a response, the temporal sequence in which synapses at different locations are activated will condition the emission of a spike. Another effect arising from the interaction of multiple dendritic signals is shunting inhibition, which occurs when one synaptic signal travels through a section where another synaptic activation had caused an opening of ion channels, resulting in an escape of current and a reduced signal. Finally, the presence of voltage-gated ion channels in the dendrite can lead to the generation of a localized dendritic spike, which enhances the synaptic input. Overall, the various transformations that can shape dendritic signals depending on spatial and temporal activation patterns constitute dendritic computations [162, 107].

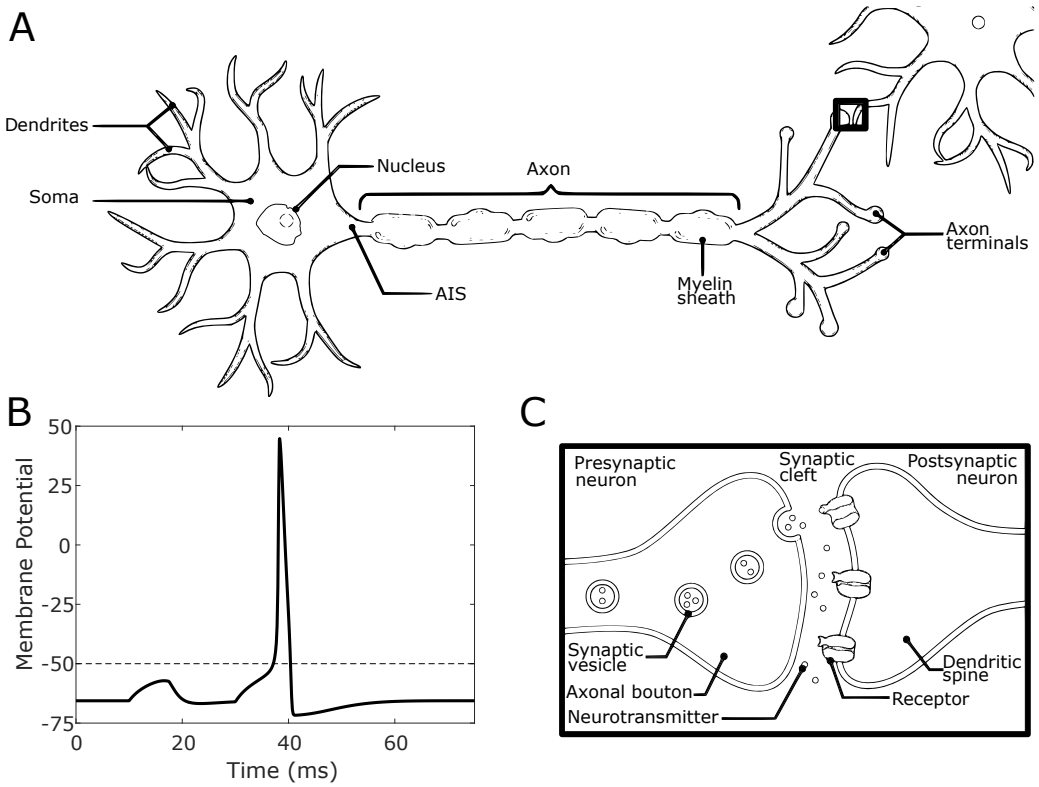


Figure 1.1.: **Neuronal structure and excitability. A:** Cellular structure of the neuron. The synaptic inputs are collected across the dendritic arbor and cause changes in the membrane voltage across the cell. At the AIS, the depolarization can trigger a spike, which will travel along the axon. Once the spike reaches the axon terminals, it will activate the synapses there, which will signal to other cells. **B:** Spike generation at the AIS. Upon excitatory stimulation, the voltage rises. If the depolarization does not reach the threshold where the voltage-gated channels open (dashed line), the voltage will return to the resting potential without eliciting a response (see depolarization at $10\,\mathrm{ms}$). If the depolarization crosses the threshold, it will trigger a positive feedback loop, which results in an action potential (see response at $30\,\mathrm{ms}$). The trace shown here is obtained using the Hodgkin-Huxley model (see section 1.2.1). **C:** Structure of the synapse (zoom-in from the highlighted box in panel A). The pre- and postsynaptic neurons are in close proximity. Communication is accomplished by the release of neurotransmitters in the synaptic cleft. Once specific receptors on the postsynaptic side bind to these molecules, they trigger the opening of ion channels, causing voltage changes in the postsynaptic compartment. Figures A and C are adapted from templates on `BioRender.com`.

While the majority of synapses connect on the dendrites of postsynaptic neurons, this is not always the case. Axosomatic synapses directly project on the cell body of the postsynaptic neuron, bypassing dendritic computations and leading to an immediate effect on the voltage in the soma. In the case of axoaxonic synapses, the postsynaptic side is the axon of another neuron. This circuit motif can modulate the signaling of the postsynaptic neuron by strengthening or weakening its response while bypassing the spike generation mechanism in the soma and specifically targeting parts of its axonal arbor [43]. Finally, some synapses bypass the need for neurotransmitters altogether. This is the case of electrical synapses in which the two cells form a gap junction where their cytoplasm is in direct contact. In this case, the electric current can directly flow from one cell to another, leading to a much faster transmission. Since electrical synapses lead to an effective synchronization of the activity of the two neurons, it does not allow for the incremental action and fine regulation present in neurotransmitter-mediated chemical synapses [113].

The phenomenon of strengthening or weakening of the connections in chemical synapses constitutes synaptic plasticity. The plasticity of a synapse can be under the control of diverse factors such as the precise timing of spikes, the firing rate of the pre- or postsynaptic neuron, or the activation of nearby synapses. It can take diverse forms depending on the timescales over which it occurs, as short-term plasticity (STP) is typically on the order of seconds, while long-term potentiation and depression involve molecular mechanisms that operate on much longer timescales. These mechanisms can include the synthesis of proteins, modifications in cellular trafficking, or changes in the morphology of the synapse. Overall, the mechanisms of synaptic plasticity shape neuronal connections in various ways based on the activity of the network, which can ultimately support the emergence of cognitive functions like memory or learning [172].

1.1.3. Network organization

Neurons can be broadly categorized into two classes: excitatory neurons (E) whose neurotransmitters trigger a depolarization of the postsynaptic neuron and promote the generation of an action potential, and inhibitory neurons (I) whose neurotransmitters cause hyperpolarization and push against postsynaptic firing [44]. However, more detailed classifications of neurons have revealed the existence of a rich diversity of cell types. These cell types differ by morphol-

ogy, electrophysiological properties, the brain regions they appear in, and the other cell types they form synaptic connections with [112]. Besides neurons, the nervous system contains a large number of glial cells, which are not electrically excitable but perform diverse supporting roles such as structure, immune function, metabolic supply, regulation of synaptic transmission, or electrical insulation.

Overall, the nervous system is organized in an elaborate network of interactions whose wiring enables the emergence of specific computations. The precise rules underlying the wiring of neural circuitry remain an active field of research. Nonetheless, these rules appear to be primarily determined genetically and mediated through the diffusion of signaling molecules and the interactions of protein markers on the cell surface [21].

One circuit motif that appears to be well-conserved and ubiquitous throughout the cortex is the so-called canonical cortical circuit. The cortex is structured in successive layers, where the external inputs primarily target the top layer, whereas the deeper layers project outward to other regions. Most connections occur within each layer, while a smaller fraction of the connections are made across cortical layers [77, 33]. At the same time, the cortex is organized in a succession of vertical minicolumns, which are driven by different stimuli and within which most of the synaptic connections are contained [121]. These minicolumns are therefore hypothesized to operate as distinct computational units. In this view, cortical function is carried out by a series of operational modules.

Ultimately, the architecture of the nervous system leads to the emergence of information pathways that transfer signals across specialized networks to process them and carry out diverse operations. At the macroscopic scale, different brain regions specialize in the processing of different information types, leading to the simultaneous execution of numerous various cognitive functions.

1.2. Neural Network models

While experimental manipulations and recording of neural networks are indispensable to investigate the mechanisms and substrates underlying the function of neural systems, the field of neuroscience also relies heavily on simulations and mathematical models to make sense of experimental findings, search for emerging phenomena, and help design testable hypotheses. Multiple neural network models exist, which differ by their degree of complexity, reflecting a trade-off between mathematical abstraction and biological realism (Fig. 1.2). The more complex models can incorporate detailed features and provide a more accurate reflection of their biological counterparts. On the other hand, simpler models have the advantage that they can be interpretable and easier to manipulate.

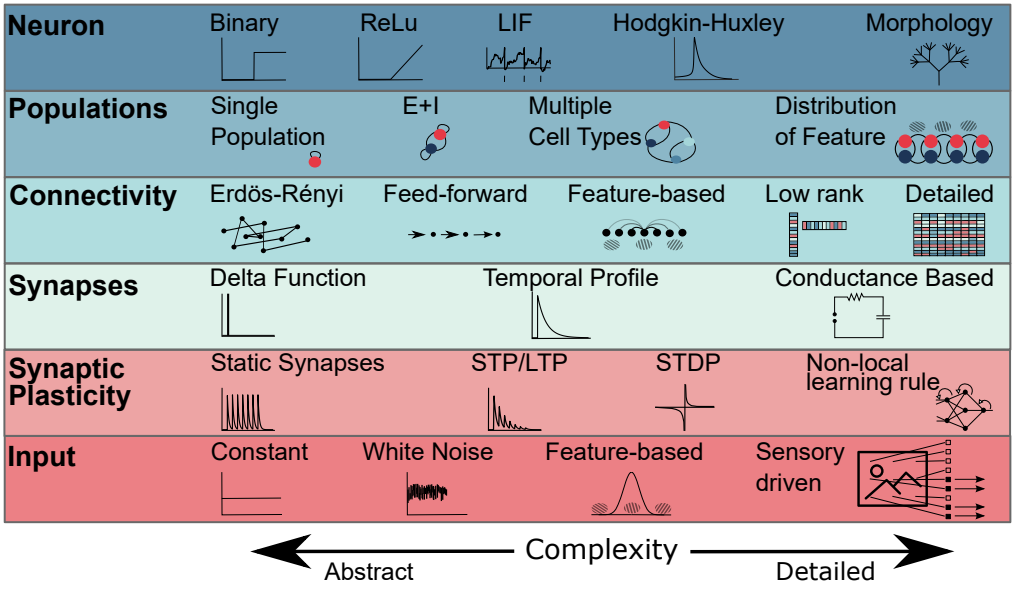


Figure 1.2.: **The essential building blocks of a neural network model ordered by degree of complexity.** A network model representing the activity of a neural circuit is characterized by its neuronal description, the categorization of neuronal properties into populations, the connectivity between neurons, the type of synapses and plasticity governing the synaptic strength, and the input that stimulates neurons. Each of these items can be modeled in different ways, with varying degrees of complexity. Figure adapted from [19].

1.2.1. Hodgkin-Huxley model

One prominent model of neuronal activation was proposed by Alan Hodgkin and Andrew Huxley in 1952. Based on the analysis of action potential transmission and measurements of membrane ionic conductance in squid giant axons [70, 71, 72, 74], they built a quantitative model

of the electrical properties of neuronal cell membrane [73]. This model can accurately recreate the generation of action potentials and their propagation along the axon. This high biological realism comes with a high computational cost, as the state of the membrane is defined by four variables coupled through ODEs. This cost is exacerbated in the case where neuronal morphology is considered, as the model must be applied to individual slices of neuronal processes. Because of this, the Hodgkin-Huxley model is mainly used to study sub-cellular phenomena [136] and is seldom used in the context of large networks [82]. Other conductance-based models have been proposed, such as the FitzHugh-Nagumo [54] or the Izhikevich model [81] which are both phenomenological simplifications of the Hodgkin-Huxley model.

1.2.2. Leaky Integrate-and-Fire model

The Leaky Integrate-and-Fire neuron (LIF) is another widely used neuron model [29]. It was initially proposed by Louis Lapicque in 1907 [97, 1] and later refined by Bruce Knight in 1972 [93]. The LIF model is based on an electric circuit analogy of the cell membrane [60]. It acts as a capacitor as it maintains a difference of electric potential between the intra- and extracellular sides. At the same time, it acts as an electrical resistance as it permits a passive leak of ions. Finally, the ion channels activated by synaptic inputs play the role of a source of current. Since the cell membrane plays these different roles simultaneously all over its surface, the equivalent electrical circuit is wired in parallel (Fig. 1.3.A). The analysis of the analogous circuit leads to the time evolution of the membrane potential

$$\frac{dV}{dt} = -\frac{C}{R}V + jC,$$

where C is the capacitance, R is the resistance, and j is the current source.

In the case of the LIF model, the equation is typically reformulated as

$$\frac{dV}{dt} = -\frac{(V - V_0)}{\tau} + I. \tag{1.1}$$

Here, the dynamics of the membrane potential is characterized by the membrane time constant τ, the input I which is given in units of $\mathrm{mV/s}$, and the resting potential V_0 which corresponds to the unexcited steady state of the cell resulting from the homeostatic activity of ion pumps.

In the LIF model, the spiking mechanism does not emerge from the electrical properties of

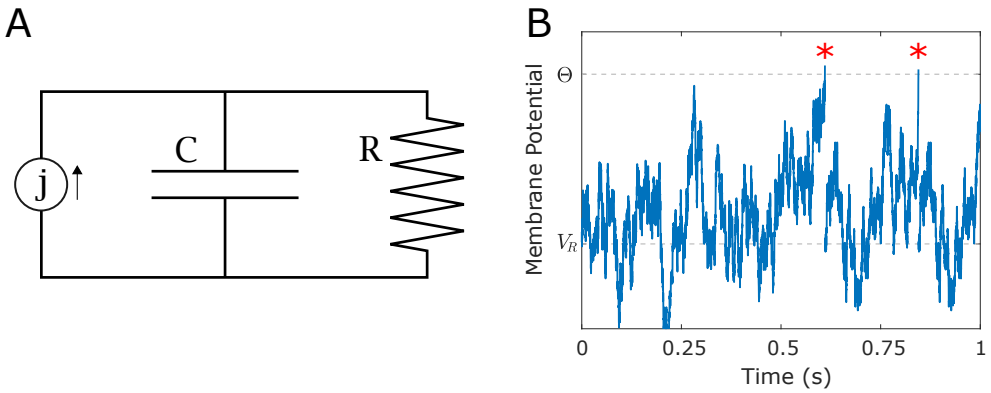


Figure 1.3.: **LIF neuron model. A:** The electrical properties of the neuronal cell membrane are analogous to RC circuits with a source of current in parallel. **B:** Example trajectory of the LIF membrane potential with white noise input I. Once the potential reaches the threshold Θ (red asterisk), the neuron emits a spike and is reinitialized at the reset potential V_R.

the circuit but comes in the form of additional boundary conditions. As soon as the membrane potential reaches a threshold value Θ, it fires. After the neuron spikes, it can enter a refractory period Δt_{ref} during which it cannot receive input and, therefore, cannot spike again. The membrane potential is then reinitialized to a reset value V_R and starts integrating input again (Fig. 1.3.B).

The activation function of a neuron, or FI curve, is the input-output relationship that characterizes the firing rate response ν of a neuron as a function of the level of input it receives. In the case of constant input of value μ, without a refractory period, the inter-spike interval (ISI) is obtained by solving the LIF equation (Eq. 1.1) with $V(t_0) = V_R$ and $V(t_0 + ISI) = \Theta$ [29]:

$$ISI = \tau \ln \left(\frac{\mu\tau + V_0 - V_R}{\mu\tau + V_0 - \Theta} \right). \tag{1.2}$$

The firing rate is then simply given by $\nu = ISI^{-1}$. In the situation where the input is not constant but follows a random process such as white noise, the activation function of the LIF neuron must be derived using stochastic calculus (section A.4).

In the case where the input follows Gaussian white noise, it is characterized by its mean μ and noise strength σ

$$I(t) = \mu + \sigma\, \eta(t).$$

with η a normally distributed random variable, such that

$$\langle \eta(t) \rangle = 0$$
$$\langle \eta(t) \times \eta(t') \rangle = \delta\left(t - t'\right).$$

With Gaussian white noise input, the membrane potential follows a drift-diffusion process. The steady state firing rate of a LIF neuron receiving white noise input corresponds to a first-passage-time problem and is given by the Ricciardi Φ equation [165, 32, 145, 25, 8]

$$\nu = \Phi(\mu, \sigma) = \left(\tau\sqrt{\pi} \int_{\frac{V_R - \mu\tau}{\sigma\sqrt{\tau}}}^{\frac{\Theta - \mu\tau}{\sigma\sqrt{\tau}}} e^{z^2} \left(1 + \operatorname{erf} z\right) dz \right)^{-1}. \tag{1.3}$$

The classic LIF neuron is widely used to model cortical activity, both *in vivo* and *in vitro* [60]. Nevertheless, its simple formulation does not reflect the complex nonlinear interactions between the membrane potential and its conductance. Because of this, models inspired by the LIF neuron have been developed to reflect the increased excitability as the membrane potential approaches the threshold Θ. Such models include the Quadratic Integrate-and-Fire (QIF), in which the dynamics depend on the membrane potential squared, [52] or the Exponential Integrate-and-Fire (EIF), which adds an exponential dependence on the membrane potential [56]. These neuron models have two steady states and can support more complex dynamics than classic LIF neurons.

1.2.3. Spiking network simulations

Neural networks can be studied by running simulations of spiking neurons. Each individual neuron is modeled and embedded into a network where it is connected to others. This approach offers a lot of flexibility as every feature of the network can be modeled at varying levels of detail. This is especially useful when recreating large-scale bio-realistic simulations of brain regions. In these cases, a detailed description of cell physiology, morphology, and connectivity can be directly based on state-of-the-art databases of experimentally reported parameters [20, 11, 111]. Such detailed simulations are valuable to make sense of the large amount of experimentally recorded data and reach a biologically realistic understanding of brain function. At the same time, the results of detailed spiking network simulations are difficult to interpret as

they result from the intricate interplay of multiple complex processes. It is therefore challenging to pinpoint the core mechanisms underlying diverse network activity patterns. Similarly, the link between specific parameter values and the resulting network activity cannot be established through simulations without an extensive search through parameter space, which is unfeasible for large networks and highly detailed simulations. This indefinite dependence on parameter values is especially problematic because experimentally reported parameters vary across broad ranges [94, 75, 42, 137, 92], and bio-realistic spiking network simulations typically require additional parameter tuning to output realistic activity patterns [20].

1.2.4. Rate models

Unlike spiking network simulations, which consider individual neurons, rate models consider populations as a whole and characterize their collective firing rate. This approach typically arises from a mean-field analysis of the system, in which populations are treated as single units, and the population-average firing rate is the solution of a mathematical equation. In order to hold, the mean-field description of neural networks relies on large population sizes, so the statistical variability across neurons is negligible compared to the mean [119]. For small networks, finite-size effects of spontaneous fluctuations and correlations cannot be ignored and corrections must be added to the rate model [25, 28, 159].

► Ricciardi Self-consistent solution

Using the Ricciardi Φ activation function of LIF neurons (Eq. 1.3), the steady state firing rate of a LIF network can be determined analytically under the assumption that the total input to neurons follows the distribution of Gaussian white noise [24]. The rate model takes the form of a self-consistent system of equations, which define the mean input μ_X, the input noise σ_X, and the firing rate ν_X of each population X. In the case of a 2-population E-I (Excitatory-Inhibitory)

network:

$$\begin{cases} \nu_E &= \Phi(\mu_E, \sigma_E) \\ \mu_E &= J_{EE}\,\nu_E - J_{EI}\,\nu_I + \mu_{extE} \\ \sigma_E^2 &= J_{EE}\,j_{EE}\,\nu_E + J_{EI}\,j_{EI}\,\nu_I + \sigma_{extE}^2 \\ \nu_I &= \Phi(\mu_I, \sigma_I) \\ \mu_I &= J_{IE}\,\nu_E - J_{II}\,\nu_I + \mu_{extI} \\ \sigma_I^2 &= J_{IE}\,j_{IE}\,\nu_E + J_{II}\,j_{II}\,\nu_I + \sigma_{extI}^2. \end{cases} \tag{1.4}$$

Where the parameter J_{XY} is the population-wise connection strength from population Y to population X. It is linked to the connection strength of the single synapse j_{XY} by the number of incoming synapses C_{XY}:

$$J_{XY} = j_{XY} \times C_{XY}.$$

The expression for the input noise σ_X is based on the assumption that the LIF spike train follows a Poisson process, which closely matches the firing patterns of simulated LIF neurons (see Fig. A.5).

► Balanced State

The Balanced State is a mathematical framework proposed by Carl van Vreeswijk and Haim Sompolinski [185, 186] to explain the generation of asynchronous irregular patterns of neuronal activity. It postulates that the mean excitatory and inhibitory inputs to neurons in large networks tend to cancel each other out, while the temporal input fluctuations remain and are the main drive of neuronal spiking. In particular, the mathematical framework of the Balanced State operates in the limit of infinitely large network size ($N \to \infty$). It assumes that the pairwise connection probability p is invariant with network size. Since the in-degree of neurons scales linearly with N, the synaptic strength must follow the scaling $j_{XY} = j^0_{XY}\sqrt{N}^{-1}$, so the input noise does not grow indefinitely with network size N:

$$\sigma_X^2 = \sigma_{ext}^2 + \sum_Y j_{XY}^2\, C_{XY}\, \nu_Y = \sigma_{ext}^2 + \sum_Y {j^0_{XY}}^2\, p_{XY}\, \nu_Y.$$

Therefore, the mean input to neurons scales with $\sqrt{N}$:

$$\mu_X = \mu_{ext} + \sum_Y j_{XY}\, C_{XY}\, \nu_Y = \sqrt{N} \left(\mu^0_{ext} + \sum_Y j^0_{XY}\, p_{XY} \right).$$

The external input here scales with $\sqrt{N}$ as well because it is considered to originate from an external population which follows the same scaling. In order for the mean input to remain finite for large network sizes, we must have

$$\left(\mu^0_{ext} + \sum_Y {j^0_{XY}}^2\, p_{XY} \right) = \mathcal{O}(\frac{1}{\sqrt{N}}).$$

This means that in the limit of large network sizes, the Balanced State imposes that the *E* and *I* inputs must cancel out. Meanwhile, multiple electrophysiological studies have confirmed that *E* and *I* synaptic inputs in the cortex tend to be balanced [163, 164, 67, 127, 151]. Furthermore, deviations in E-I balance have been shown to impair the signal-to-noise ratio and appear to be associated with neurological disorders [169].

Overall, the assumption of tight balance between *E* and *I* synaptic inputs suffices to exactly predict the firing rates in the network. For a 2-population network:

$$\begin{cases} \nu_E = (J_{II} - r J_{EI}) \det J^{-1} \times \mu_{ext} \\ \nu_I = (J_{IE} - r J_{EE}) \det J^{-1} \times \mu_{ext}, \end{cases} \tag{1.5}$$

where r is the ratio of external input to the *I* and *E* populations ($r = \mu_{extI}/\mu_{extE}$) and $\det J$ is the determinant of the population-wise connectivity matrix

$$\det J = J_{IE} J_{EI} - J_{EE} J_{II}. \tag{1.6}$$

Besides the assumption that N is large enough to achieve tight balance, the Balanced State rate model imposes additional conditions on its parameters. First of all, the J matrix must be invertible ($\det J \neq 0$). Secondly, the connection strengths J must be such that the firing rates

in Eq. 1.5 are non-negative. With positive feedforward inputs, this condition becomes

$$\begin{cases} 0 < r < \frac{J_{II}}{J_{EI}}, & \det J > 0 \\ r > \frac{J_{II}}{J_{EI}}, & \det J < 0. \end{cases} \tag{1.7}$$

The Balanced State rate model is unique in the simplicity of its mathematical formulation (Eq. 1.5) and the fact that it does not depend on the firing properties of neurons in the network. In particular, the model predicts that firing rates in balanced networks are linear functions of the feedforward input. This prediction contrasts with experimentally observed nonlinearities in network responses [34, 126]. Furthermore, from a computational point of view, nonlinearities are necessary to enable complex operations. In order to achieve nonlinear transformations in balanced networks, additional mechanisms such as synaptic plasticity [118] or detailed silencing of neuronal subpopulations [13] must be included.

► Stabilized Supralinear Network

The Stabilized Supralinear network (SSN) is a rate model developed by Kenneth Miller and Daniel Rubin [4, 150] to replicate the effect of surround suppression (reduced response to a broader stimulus of identical amplitude) and normalization (sublinear response to a sum of stimuli) observed in sensory cortical areas. The SSN is based on a power law neuronal activation function, as supported by experimental observations [139]. For a 2-population system, the dynamical equations of the SSN are given by

$$\begin{aligned} \tau_{P_E} \frac{d\nu_E}{dt} &= -\nu_E + a\left(J_{EE}\nu_E - J_{EI}\nu_I + \mu_{ext}\right)_+^n \\ \tau_{P_I} \frac{d\nu_I}{dt} &= -\nu_I + a\left(J_{IE}\nu_E - J_{II}\nu_I + r\mu_{ext}\right)_+^n, \end{aligned} \tag{1.8}$$

where τ_{P_E} and τ_{P_I} are the population time constants that characterize how fast the firing rates of each population evolve.

1.2.5. ANNs

Another unavoidable type of neural network model is the Artificial Neural Network (ANN) which has been trained through machine learning to perform specific tasks or generate particular activity patterns. These networks are comparable in their structure to Spiking Network models as they comprise individual neurons typically organized in layers and detailed connection rules, but they differ starkly in the conceptual approach. While Spiking Network models are used to generate activity patterns corresponding to a given network structure, ANNs are specifically trained and optimized to achieve a predefined transformation. The approach of neural network models is bottom-up (network structure $\rightarrow$ activity pattern) whereas ANNs are top-down (task performance/target activity pattern $\rightarrow$ network structure) [19]. Because of the goal difference, ANNs typically disregard biological realism in favor of algorithmic performance. The neuronal model is usually a simple Rectified-Linear [66] or sigmoidal unit [188], which are organized in successive processing layers [116, 57]. On the other hand, the connectivity of trained networks is generally very intricate and high-dimensional, making the interpretability of ANNs a tremendous challenge in the field of machine learning [38]. The ever-improving performances of ANNs have allowed breakthroughs in countless domains such as computer vision [96], competitive games [166, 167] protein folding [49] or text generation [129]. Ironically, even though ANNs are inspired by biological neural networks, they have not achieved this level of groundbreaking advances in neuroscience. Nevertheless, ANNs are a powerful tool and are being used increasingly often to help study neural systems in various ways. These approaches include decoding recorded neural activity [174, 17], providing insights into algorithmic solutions to perform specific tasks [87], or training biologically realistic spiking neural networks for behaviorally relevant tasks [62, 173].

1.3. Research goal

Elucidating the mechanisms underlying the function of biological neural networks remains one of the main challenges of neuroscience. Due to the high-dimensional and complex nature of biological neural networks, a proper mechanistic understanding can only be achieved with the use of theoretical principles, mathematical models, and simulations.

Spiking network simulations can provide valuable insights into the activity regimes of neural networks since they can produce the resulting activity pattern for any given network parameters and structure. However, the output of a simulation can be challenging to interpret and generalize. Which components of the network firing activity constitute the signal and which are just noise? Of all the interactions in the system, which ones are key to achieving the desired output? How does a slight change in parameter affect the overall network function?

On the other hand, rate models are powerful tools to study neural networks, and their mathematical analysis can lead to deep mechanistic insights into network function. These models are diverse and vary by the key assumptions under which they are valid, the degree of complexity of their mathematical formulation, and the properties of their predicted output. While the most complex rate models are impractical to use due to their mathematically intractable nature, simpler models can lead to erroneous conclusions if they are used outside of their narrow domain of validity.

In this work, I aim to bridge the gap between spiking simulations and mathematically tractable rate models. Using an adjusted SSN model designed to match LIF simulations, I explore the diversity of nonlinear activity regimes accessible to two-population E-I systems using biologically realistic parameters. I highlight how such regimes can play a crucial role in supporting complex computation at the network level and delimit the conditions under which the rate model fails to predict spiking network activity and how to address these limitations.

2. Rate model approximation of LIF networks

This chapter is adapted from the first result section and the supporting information of the article

Ekelmans Pierre, Nataliya Kraynyukova, and Tatjana Tchumatchenko. "Targeting operational regimes of interest in recurrent neural networks." *PLOS Computational Biology* 19.5 (2023): e1011097.

https://doi.org/10.1371/journal.pcbi.1011097

In this chapter, I derive the LIF-fitted SSN rate model which provides an advantageous trade-off between mathematical tractability and accuracy. Thanks to a simple mathematical formulation and its closed-form equation, it can be studied analytically and the occurrence of diverse activity features can be linked to conditions on its parameters. At the same time, it is designed to match the behavior of LIF neurons closely and I show how its parameters are chosen to reflect available data from experimental sources.

2.1. Single neuron FI curve approximation

As shown in section 1.2.2, the firing rate of a LIF neuron fed with white noise input can be determined exactly by the Ricciardi Φ equation (Eq. 1.3). While this expression is difficult to manipulate analytically, we can observe how the Φ function behaves for different parameters using numerical methods.

Overall, the Φ curve appears to show four different regimes as a function of the input μ (see Fig. 2.1):

- For very low input, the neuron is silent. If the input level is too low to bring the membrane potential to the spiking threshold and the fluctuations due to noise are insufficient to bridge this gap, the neuron will remain inactive and the firing rate will remain close to 0 Hz.

- For moderate input and sufficient noise level, neuron firing is a supralinear function of the input. This occurs due to random fluctuations, which can drive the membrane potential to the spiking threshold. The likelihood of such events increases as the mean input drives the membrane closer to its threshold.

- For high input, the contribution of random fluctuations becomes negligible and the threshold crossing is mainly driven by the integration of the mean input. Then, the firing rate grows linearly as a function of the input (see section A.5).

- For high firing rates, the FI curve grows sublinearly as the firing rates approach the limit imposed by a refractory period Δt_{ref}.

For firing rates in the biologically relevant 0-10 Hz range (see section 2.4.1), we can focus on the first two parts of the curve. In order to facilitate the analysis of LIF neuron dynamics, the activation function can be approximated with a supralinear power law

$$f(\mu) = a\,(\mu - b)_+^n\,, \tag{2.1}$$

where μ is the input to the neuron. Here, $(x)_+$ is equal to x when it is positive and zero otherwise. The b parameter acts as a threshold level of input at which the neuron starts firing. The parameter a is a scaling factor. The exponent n characterizes how supralinear the FI curve is and must be larger than 1. Using this approximation, the FI curve can be fitted for any LIF neuron to determine the three a, b, and n corresponding parameters (Fig. 2.2). In general, each cell type or neuronal population has different properties, so the activation function f_X for each population X must be defined separately.

2.2. Network rate model

For neurons embedded in a network, the input is the sum of feedforward input coming from outside the network and recurrent input, which is caused by the connections from other neurons within the network. In a mean-field framework, the mean input μ_X of neurons in population X can then be expressed as

$$\mu_X = \sum_Y J_{XY}\,\nu_Y + \mu_{extX}\,,$$

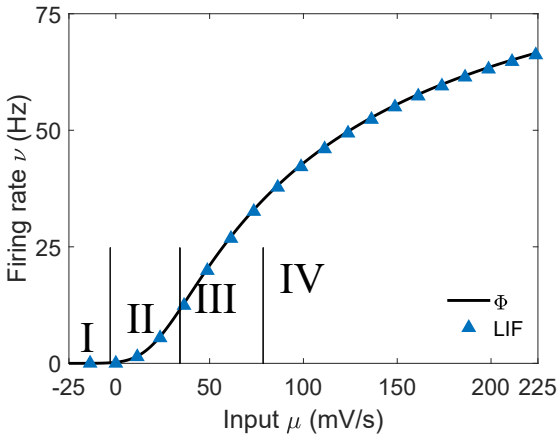


Figure 2.1.: **FI curve of Leaky Integrate-and-Fire neurons.** The theoretical response of LIF neurons to input (black line) is in agreement with the simulated LIF activation function (blue triangles). The response is first static and silent (I), then supralinear (II), linear (III), and finally sublinear as the firing rate approaches the upper bound caused by the refractory period (IV). Here, the refractory period is set at $10\,\mathrm{ms}$, so the firing rates cannot rise beyond $100\,\mathrm{Hz}$.

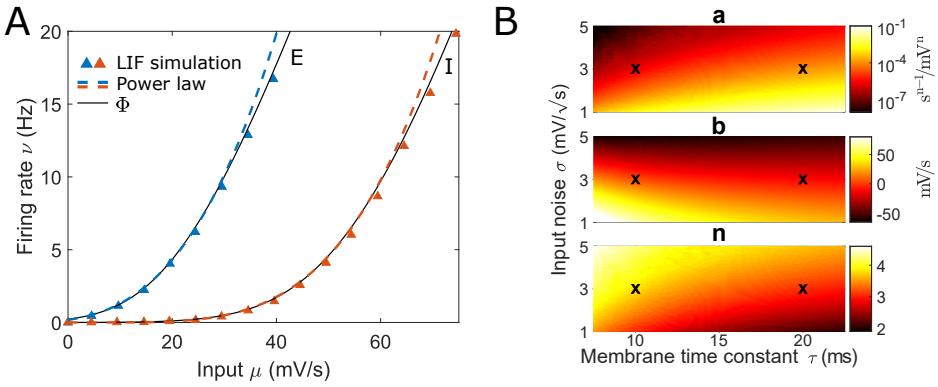


Figure 2.2.: **The FI curve of LIF neurons can be accurately fitted with a power law. A:** We consider two types of neurons: *E* and *I* (see Table 2.3). The activation function is fitted in the 0-10 Hz range and appears indistinguishable from the LIF theoretical and simulated FI curves. **B:** The fitted power law has three parameters, a, b and n. These parameters can be fitted for any LIF neuron with different parameters. The values of the three power law parameters are shown here for a range of input noise σ and membrane time constant τ of the LIF neuron. By default, the reset and resting potential are set to 0, and the firing threshold is set to $1\,\mathrm{mV}$. The parameters corresponding to the *E* and *I* neurons from panel A are indicated by a cross ($\tau_E = 20\,\mathrm{ms}$ and $\tau_I = 10\,\mathrm{ms}$); their numerical values are listed in Table 2.5.

where J_{XY} is the strength of the population-wise effective connection strength from population Y to population X and ν_X is the average firing rate of population X. Regarding the input noise σ, we assume that the contribution from recurrent inputs is negligible and that the input noise is mainly driven by the external source (i.e., $\sigma_X \approx \sigma_{extX}$). This is useful because the noise level σ does not appear explicitly in the power law approximation, but it is contained in the a, b, and n parameters. By assuming that the noise level is solely determined by the external noise σ_{ext}, the characteristics of the FI curve are decoupled from the activity levels of the populations in the system, and the closed-form network equation is simplified since a, b and n are considered constants. In the case of two populations, *E* and *I*, the steady states of the SSN are given by the closed-form equation:

$$\begin{cases} \nu_E = a_E \left(J_{EE}\, \nu_E - J_{EI}\, \nu_I + \mu_{ext} - b_E \right)_+^{n_E} \\ \nu_I = a_I \left(J_{IE}\, \nu_E - J_{II}\, \nu_I + r\mu_{ext} - b_I \right)_+^{n_I} , \end{cases} \tag{2.2}$$

where r is the ratio of external inputs to each population (μ_{extI}/μ_{extE}). The obtained system of equations is an extension of the classic Stabilized Supralinear Network (SSN) at the steady state (Eq. 1.8) [4, 150].

Following the method used in [95] to solve the classic SSN, we can define a characteristic function $\mathcal{F}$ to combine both equations into one. Starting from the total input to each population μ_X, which is given by

$$\begin{bmatrix} \mu_E \\ \mu_I \end{bmatrix} = \begin{bmatrix} J_{EE} & -J_{EI} \\ J_{IE} & -J_{II} \end{bmatrix} \begin{bmatrix} \nu_E \\ \nu_I \end{bmatrix} + \begin{bmatrix} \mu_{ext} \\ r\mu_{ext} \end{bmatrix} , \tag{2.3}$$

we can isolate the firing rates

$$\begin{bmatrix} \nu_E \\ \nu_I \end{bmatrix} = \frac{\begin{bmatrix} -J_{II} & J_{EI} \\ -J_{IE} & J_{EE} \end{bmatrix} \begin{bmatrix} \mu_E - \mu_{ext} \\ \mu_I - r\mu_{ext} \end{bmatrix}}{J_{EI}J_{IE} - J_{EE}J_{II}} .$$

Using the activation functions f_E and f_I to substitute the firing rates with a function of the input,

the system only has two variables μ_E and μ_I

$$\begin{cases} f_E(\mu_E) & = -\frac{J_{II}}{\det J}\,\mu_E + \frac{J_{EI}}{\det J}\,\mu_I + \mu_{ext}\,\frac{J_{II}-rJ_{EI}}{\det J} \\ f_I(\mu_I) & = -\frac{J_{IE}}{\det J}\,\mu_E + \frac{J_{EE}}{\det J}\,\mu_I + \mu_{ext}\,\frac{J_{IE}-rJ_{EE}}{\det J}, \end{cases}$$

where $\det J$ is the determinant of the weight matrix (Eq. 1.6). By isolating the μ_I term from the first equation and substituting it in the second one, we can obtain a single equation: the characteristic function $\mathcal{F}$. Its zeros are the solutions of the steady SSN equation (Eq. 2.2):

$$\mathcal{F}(\mu_E) = J_{EE} f_E(\mu_E) - J_{EI} f_I(\mu_I(\mu_E)) - \mu_E + \mu_{ext}. \tag{2.4}$$

The polynomial $\mu_I(\mu_E)$ is the total inhibitory input μ_I expressed in function of μ_E

$$\mu_I(\mu_E) = \frac{\det J}{J_{EI}} f_E(\mu_E) + \frac{J_{II}}{J_{EI}}\mu_E + \mu_{ext}\left(r - \frac{J_{II}}{J_{EI}}\right). \tag{2.5}$$

The zeros of the $\mathcal{F}$ function can be found with a nonlinear system solver, and the corresponding steady state firing rates are simply given by the activation functions $f_E(\mu_E)$ and $f_I(\mu_I(\mu_E))$. In the case of the adjusted power law approximation of the LIF FI curve, the characteristic function becomes

$$\mathcal{F}(\mu_E) = J_{EE} a_E (\mu_E - b_E)_+^{n_E} - J_{EI} a_I (\mu_I - b_I)_+^{n_I} - \mu_E + \mu_{ext}, \tag{2.6}$$

with

$$\mu_I = \frac{a_E \det J}{J_{EI}} (\mu_E - b_E)_+^{n_E} + \frac{J_{II}}{J_{EI}}\mu_E + \mu_{ext}\left(r - \frac{J_{II}}{J_{EI}}\right).$$

If the characteristic function $\mathcal{F}$ admits multiple zeros, the network possesses multiple steady state activity levels. The analysis of the characteristic function $\mathcal{F}$ alone cannot determine whether a steady state is stable. The stability of a fixed point will depend on the dynamic properties of the system, such as the population time constants τ_{P_X} in Eq. 1.8 (see Eq. 2.10). Nevertheless, it can be shown that a negative slope of the $\mathcal{F}$ function is a necessary condition

for a fixed point μ_E^0 to be stable [95]

$$\mu_E^0 \text{ is stable} \Rightarrow \left.\frac{d\mathcal{F}}{d\mu_E}\right|_{\mu_E^0} < 0. \tag{2.7}$$

In the case where the system comprises more than two populations, the closed form in Eq. 2.2 can be expanded by adding another equality for each additional population. Already with three populations, the set of equations can no longer be simplified into a single characteristic function of one variable because the development leads to a polynomial of the form $a\mu_{1+}^{n_1} + b\mu_{2+}^{n_2} + c\mu_1 + d\mu_2 + e = 0$ from which one variable cannot be algebraically isolated to express it as a function of the other. Nonetheless, the closed-form SSN in higher dimensions can still be solved numerically using algorithms such as the continuation method (section A.2).

2.3. Limitations

While the LIF-fitted SSN can provide a valuable framework to study firing rate regimes of spiking neural networks, it is worth keeping in mind the underlying assumptions of the model and its limitations.

2.3.1. Steady state assumption

Since the power law fit is based on the LIF response at steady state, this system is limited to studying neural networks at the steady state and cannot predict temporal dynamics or transient behaviors (see section 2.4.5). Furthermore, as mentioned previously, the temporal properties of the network do play a role in the linear stability of the system. If we assume a simple ODE of the form

$$\begin{cases} \tau_{P_E}\dot{\nu_E} &= -\nu_E + f_E\left(J_{EE}\nu_E - J_{EI}\nu_I + \mu_{ext}\right) \\ \tau_{P_I}\dot{\nu_I} &= -\nu_I + f_I\left(J_{IE}\nu_E - J_{II}\nu_I + r\mu_{ext}\right), \end{cases}$$

we can linearize around a fixed point $\nu_X = \nu_X^0 + d\nu_X$ to reach

$$\begin{bmatrix} d\dot{\nu_E} \\ d\dot{\nu_I} \end{bmatrix} = \begin{bmatrix} \frac{J_{EE}f'_E - 1}{\tau_{P_E}} & \frac{-J_{EI}f'_E}{\tau_{P_E}} \\ \frac{J_{IE}f'_I}{\tau_{P_I}} & \frac{-J_{II}f'_I - 1}{\tau_{P_I}} \end{bmatrix} \begin{bmatrix} d\nu_E \\ d\nu_I \end{bmatrix} + \begin{bmatrix} \frac{f'_E}{\tau_{P_E}} d\mu_{ext} \\ \frac{rf'_I}{\tau_{P_I}} d\mu_{ext} \end{bmatrix}. \tag{2.8}$$

The fixed point ν_E^0, ν_I^0 is linearly stable if the two eigenvalues of the 2×2 matrix $\mathcal{M}$ have negative real parts, which is equivalent to imposing that it has a positive determinant and a negative trace. Since the time constants τ_{P_X} are positive, they do not condition the sign of the determinant of the matrix $\mathcal{M}$. Therefore, the first stability condition is equivalent to $|T\mathcal{M}| > 0$, where T is the diagonal matrix containing the time constants so that the product $T\mathcal{M}$ is independent of the dynamic parameters. The first stability condition reads

$$\begin{aligned} |T\mathcal{M}| &> 0 \\ 1 - f'_E J_{EE} + f'_I J_{II} + f'_E f'_I \det J &> 0. \end{aligned} \tag{2.9}$$

It can be shown that $|T\mathcal{M}| = -\mathcal{F}'$, so the static stability condition is equivalent to the condition on the slope of the characteristic function (Eq. 2.7). The second stability condition, which requires a negative trace of $\mathcal{M}$, does impose a constraint on the time constants:

$$(J_{EE} f'_E - 1)\, \tau_{P_I} < (J_{II} f'_I - 1)\, \tau_{P_E}.$$

In the SSN, the neuronal activation functions f_X are given by Eq. 2.1. Their derivative can then be written as $f' = n a^{\frac{1}{n}} \nu^{\frac{n-1}{n}}$. The stability condition on the trace of $\mathcal{M}$ becomes:

$$\frac{\tau_{P_E}}{\tau_{P_I}} > \frac{J_{EE}\, n_E\, a_E^{\frac{1}{n_E}}\, \nu_E^{\frac{n_E-1}{n_E}} - 1}{J_{II}\, n_I\, a_I^{\frac{1}{n_I}}\, \nu_I^{\frac{n_I-1}{n_I}} + 1}. \tag{2.10}$$

This condition can be understood intuitively. If the excitatory activity is weak (negative numerator in Eq. 2.10), the steepness of the power law does not provide sufficient feedback amplification to sustain a perturbation and the fixed point is stable regardless of the time constants (i.e., the E population is below the ISN threshold, see section 3.1). On the contrary, if the E activity is high, the positive feedback of recurrent $E \to E$ connections can lead to runaway dynamics. In that case, the transient of the E population must be sufficiently slow to let the recurrent inhibition divert the trajectory away from this positive feedback loop and prevent instability.

2.3.2. Assumption of white noise input

The fitted power law is based on the theoretical FI curve of LIF neurons in response to white noise input (Eq. 1.3). The LIF-fitted SSN is therefore only applicable to situations where the

statistics of the input received by neurons are well approximated with a white noise process.

In our framework, the external input is assumed to be white noise. This input models two distinct processes. The first process is the sum of all synaptic inputs received by neurons outside of the considered E-I network. The source can be a connection across layers within the same cortical minicolumn as well as inputs from long-range projections from other brain regions. The mean input from this process is assumed to carry most of the information received by the neural network. For instance, in the context of a neural network involved in processing sensory stimuli, the mean external input is assumed to be a function of the features of the stimulus. The second process that is modeled by the external input is the intrinsic fluctuation in the membrane potential. It is caused by the stochastic properties of ionic pumps and channels or transient inhomogeneity within and around the cell [135, 3]. While spontaneous membrane fluctuations can realistically follow a Gaussian noise, this assumption is more delicate in the case of the external synaptic inputs.

First of all, we must assume that the input is uncorrelated in time. This is a strong simplification of the mechanisms underlying the firing activity in neurons outside of the considered network, given that their own activity could be driven by shared inputs or that feedback connections could lead to longer correlation timescales. Furthermore, the synaptic input is triggered by spikes which are discrete events. Spike trains of LIF neurons are typically modeled as Poisson processes (Fig. A.5). Nevertheless, based on the Central Limit Theorem, the overall contribution from all synaptic inputs will approach a normal distribution if the number of synapses is large. Considering that the number of synaptic connections typically approaches 10^4 per neuron [182, 158, 46], the normal approximation is justifiable. Finally, the Ricciardi activation function (Eq. 1.3) and its power law approximation rely on a drift-diffusion approximation of the input. This assumption breaks if individual synaptic inputs have a high weight, in which case neuronal spiking will be driven by punctual events rather than a continuous integration process.

Overall, the sum of postsynaptic potentials (PSPs) received by a neuron is equivalent to a Wiener process with drift -or white noise- if the number of incoming PSPs is sufficiently large, if multiple PSPs are required to trigger spiking, and if the activity is temporally uncorrelated or Poissonian [47, 176]. This applies to both the recurrent and the feedforward connections.

2.3.3. Recurrent noise

In the fitted SSN (Eq. 2.2), the contribution of the noise σ in the input is contained in the power law parameters a, b, and n. In order to keep the mathematical model tractable, the noise σ is considered independent of the level of recurrent activity. Since the total noise is due to the combined effect of the external and recurrent inputs, the recurrent noise should be negligible:

$$\sigma^2 = \sigma_{ext}^2 + \sigma_{rec}^2 \approx \sigma_{ext}^2.$$

In a spiking network with Poissonian spike trains, the recurrent input noise of neurons in population X is given by $\sigma_{recX}^2 = \sum_Y j_{XY}^2 C_{XY} \nu_Y$, where j_{XY} is the synaptic strength of an individual connection $Y \to X$ and C_{XY} is the number of connections a neuron in population X receives from population Y. For the recurrent noise to be negligible compared to the external noise, the strength of individual PSPs must be small since their contribution is squared. Likewise, the recurrent noise grows linearly with the activity level of the network and cannot be neglected for high firing rates.

In situations where the contribution of recurrent noise cannot be ignored, it is still possible to use the SSN framework, but the power law parameters must be adjusted with respect to the activity level

$$\begin{aligned}
\nu_E &= a_E\left(\nu_E, \nu_I\right)\left(J_{EE}\nu_E - J_{EI}\nu_I + \mu_{ext} - b_E\left(\nu_E, \nu_I\right)\right)_+^{n_E(\nu_E,\nu_I)} \\
\nu_I &= a_I\left(\nu_E, \nu_I\right)\left(J_{IE}\nu_E - J_{II}\nu_I + r\mu_{ext} - b_I\left(\nu_E, \nu_I\right)\right)_+^{n_I(\nu_E,\nu_I)}.
\end{aligned}$$

However, this approach misses the main benefit of the SSN as it is not suitable for algebraic analysis. A possible compromise could be to define the activation function as a piecewise power law, so that a tractable system can be recovered on separate domains

$$\begin{aligned}
\nu_E &= a_{E_1}\left(J_{EE}\nu_E - J_{EI}\nu_I + \mu_{ext} - b_{E_1}\right)_+^{n_{E_1}}, \quad \nu_E + \nu_I < \nu_T \\
\nu_E &= a_{E_2}\left(J_{EE}\nu_E - J_{EI}\nu_I + \mu_{ext} - b_{E_2}\right)_+^{n_{E_2}}, \quad \nu_E + \nu_I > \nu_T,
\end{aligned}$$

where ν_T is a threshold value above which the power law parameters are chosen to correspond to a higher value of σ, which reflects the higher noise level due to the recurrent activity.

In this work, we restrict our analysis to situations where the recurrent noise is negligible, as

this regime is supported by experimentally reported parameters (see section 2.4).

2.3.4. Mean-field framework

Another essential feature of the LIF-fitted SSN is that it is a mean-field approach [130]. As such, the behavior of individual neurons is not considered and we instead study the global state of the population which comprises a large number of individual neurons. We focus on the population-averaged activity and ignore the possible contribution of each cell. Since individual events are not considered, the possible correlations between them cannot be incorporated into the model. This includes correlations in the spike time of neurons, which can arise due to the shared recurrent inputs within the network. This phenomenon can amplify fluctuations in network activity as a spontaneous increase in activity can lead to a temporary rise in recurrent input, which further fuels the increased activity. In order to capture such effects as correlations and fluctuations, rate models must consider the finite-size deviations from the mean-field [25, 28, 27, 159]. These effects are deviations from the population mean and become negligible in the limit of infinite size N.

Another interesting feature missing from mean-field approximations is the effect of connectivity motifs in the network. The local structure of the connectivity at the level of a few neurons has been shown to significantly condition the statistics of neuronal activity and its effects are thought to support complex computation at the network level [79, 170].

2.3.5. Fitting inaccuracy

Since the power law activation function of neurons approximates the theoretical stationary firing rate of the LIF neuron, the accuracy of the SSN prediction is limited by its goodness-of-fit. Although the power law provides a good approximation of the Φ function (Eq. 1.3), even minor deviations can be amplified through the effect of recurrent connections. If the firing rate is over/under-estimated, the resulting recurrent inputs to both populations will also be modified, leading to a recursive error propagation. The amplitude of the error depends on the domain over which the FI curve is fitted. In order to maximize accuracy, the fit should be made locally in the range where the network activity will be studied (Fig. A.7). In particular, the accuracy of the power law fit deteriorates when the range considered includes the linear high-input regime (Fig. 2.1).

2.4. Biologically realistic network parameters

2.4.1. Firing rates

In this work, we restrict our analysis to the low firing rate range of 0-10 Hz. This is the range of firing rates that is typically recorded *in vivo* across various brain regions and species [110, 143, 69, 59, 84, 175, 149, 12]. It should be clarified that this does not ignore the fact that firing rates can reach much larger values than 10 Hz, but such situations mostly fall outside of the context of this study as the mean-field network model operates at the steady state and does not distinguish individual neurons. In the case of transient response, the recorded firing rates can reach values far beyond 10 Hz, but this increase does not last more than a few tens of ms [64, 53, 59]. Such sharp responses should not be considered in a steady state network as they can result from dynamic properties of the network in response to an abrupt perturbation. They can be caused as the system transitions from one equilibrium point to another and overshoots the target. Similarly, the observation of exceptionally high firing rates over a short time window can reflect a synchronization event in response to a perturbation. For instance, with a 10 ms resolution, an event that can consistently trigger the emission of a single spike will result in a punctual 100 Hz anomaly regardless of the baseline firing rate. Such correlative effects are outside the scope of mean-field models.

Firing rates beyond the 0-10 Hz range can also be observed when considering the characteristics of individual cells [69]. This can occur even if all neurons do effectively fire at a rate lower than 10 Hz, due to the stochastic nature of the Poisson process leading some cells to randomly emit a higher number of spikes over the recording time window. Beyond the statistic fluctuations of spike trains, it has been observed that neurons within a population can fire at vastly different firing rates [78, 125] following a long-tail distribution approaching a log-normal curve [149]. This phenomenon can be caused by the heterogeneity in the input to neurons, leading the most highly stimulated neurons to fire at much higher rates than the population average. Similarly, a diversity of cell types associated with different physiological properties can result in heterogeneity in the firing patterns [41]. Nevertheless, even in experiments that report high heterogeneity in single-cell firing patterns, the population averages still appear to fall within the 0-10 Hz range [149]. Since the mean-field model does not deal with the fine structure of the network but only considers the average behavior, it is justified to focus on this range of firing

rates.

It should be mentioned that restricting our study to the low firing rates provides two substantial advantages. Firstly, the FI curve remains in the supralinear region (see Fig. 2.1). Secondly, the recurrent noise remains limited, which facilitates the analytical study of the SSN model.

Nonetheless, the SSN framework is not intrinsically limited to the analysis of low activity regimes, as long as the FI curve is appropriately fitted and the recurrent noise is accounted for.

2.4.2. Connection probability

In the rate model, the effective connectivity is determined by population-wise connectivity parameters J. They depend on three factors: the number of neurons in a population, the strength of individual synapses, and the probability of connection for a pair of neurons.

In order to characterize the connectivity in an E-I network, four parameters must be determined: p_{EE}, p_{EI}, p_{IE}, and p_{II}. The parameter p_{XY} is the pair-wise probability that a neuron in population Y projects to a neuron in population X. As a target network, we consider layer 2/3 of the mouse V1 region. This choice is only motivated by the fact that this brain region is the focus of much research and is therefore well documented in the literature. In particular, we use the Allen Institute Synaptic Physiology database [6]. The data compiled in this database is acquired through octopatching, in which 8 different cells are patched simultaneously to probe 56 candidate synaptic connections. When one of the cells is driven to spike, the response of the other 7 cells is recorded, and the cells that respond to the spike (EPSP or IPSP) are reported to be connected to the spiking neuron. In the database, layer 2/3 comprises four cell types: Pyramidal, PV, SST, and VIP neurons. In a 2-population E-I network, we agglomerate all interneurons (PV, SST, and VIP) into a single *I* population. The resulting connection probabilities are given in Table 2.1.

p_{EE}	p_{EI}	p_{IE}	p_{II}
6.5 %	20 %	27.5 %	10 %

Table 2.1.: Connection probabilities between populations *E* and *I* collected from the Synaptic Physiology database of the Allen Institute [6] for Mouse V1 layer 2/3.

2.4.3. Synaptic strength

The next step in determining the effective connectivity J_{XY} of the network is to obtain the strength of the synaptic connection j_{XY} in a pair of connected neurons. For this, we use again the synaptic physiology database of the Allen Institute [6]. During the octopatching measurements, different features of the Post Synaptic Potential (PSP) are recorded. We focus here on three: The mean peak amplitude (A_{PSP}), the mean rise time (t_R), and the mean decay time constant (τ_D). These three parameters, along with the synaptic transmission delay D, can fully characterize the average waveform of a single PSP:

$$PSP(t) = \begin{cases} V_0 & \text{, if } t < D \\ V_0 + \frac{A_{PSP}}{t_R}\ (t - D) & \text{, if } D < t < t_R + D \\ V_0 + A_{PSP}\, e^{-\frac{t - t_R - D}{\tau_D}} & \text{, if } t > t_R + D. \end{cases}$$

Here, V_0 is the resting potential and t is time such that the presynaptic spike occurs at $t = 0$. Since the only input received by the neuron in this experiment is from the synaptic connection, we can characterize it as a function of the PSP signal using the LIF equation (Eq. 1.1)

$$I_{syn}\,(t) = \frac{dV}{dt} + \frac{V - V_0}{\tau}.$$

In the static mean-field approximation, the temporal evolution of the synaptic signal is irrelevant and we can simply characterize the synaptic strength with the time-independent overall change in potential caused by a single spike ($\int I_{syn} dt = j_{XY}$). From this, the synaptic strength is given by

$$j_{XY} = \int_0^\infty \left(\frac{dV}{dt} + \frac{V - V_0}{\tau_X} \right) dt = (V_0 - V_0) + \frac{\int_0^\infty (V - V_0)\ dt}{\tau_X} = A_{PSP_{XY}} \frac{t_{R_{XY}}/2 + \tau_{D_{XY}}}{\tau_X}.$$

	EE	**EI**	**IE**	**II**
A_{PSP} (mV)	0.102	-0.362	0.465	-0.53
t_R (ms)	3.8	6.6	3.4	5.8
τ_D (ms)	26	145	25	88

Table 2.2.: Properties of the PSP collected from the Allen Institute database [6] for Mouse V1 layer 2/3.

Neuronal parameters	τ (ms)	$\mathbf{V_0}$ (mV)	$\mathbf{\Theta}$ (mV)
Excitatory population	20	-75.7	-36.4
Inhibitory population	10	-72.9	-35.2

Table 2.3.: Electrophysiological properties of the *E* and *I* neurons collected from the Allen Institute database [6] for Mouse V1 layer 2/3.

It is the area under curve of the PSP divided by the membrane time constant of the postsynaptic neuron. The contribution of τ_X can be interpreted as the pull that the synaptic input must overcome to elicit a PSP in the postsynaptic neuron. Therefore, the stronger the leak, the stronger j_{XY} must be to achieve the observed change in the postsynaptic potential.

The membrane time constant τ, the resting potential V_0, and the firing threshold Θ can be obtained from the Cell type database of the Allen institute [6] (Table 2.3). We use the difference between the threshold and resting potentials to rescale the synaptic weights j_{XY}. With this, the resting potential of the LIF can be set at 0 and the firing threshold at $1\,\mathrm{mV}$. Similarly, we set the reset potential after spiking at the resting potential $V_R = V_0 = 0\,\mathrm{mV}$.

2.4.4. Network size

The last parameter we need to characterize the effective population-wise connection strength is the number of neurons in each population. In the SSN framework, we assume that a network consists of populations of neurons which share a similar external input and connect together with a homogeneous connection probability and strength. All connections originating from outside this circuit are considered to be feedforward input. In biological circuits, it is difficult to determine what can constitute a single network since the brain exhibits a high degree of complexity and does not consist of well-separated circuits. Nonetheless, the analysis of cortical regions where projection columns can be anatomically identified leads to a consistent order of magnitude for network sizes (see section A.7). Overall, experimental data supports that $N_E = 3000$ and $N_I = 1000$ neurons constitute a realistic size for a homogeneous E-I network. This network size corresponds to a fundamental functional network unit such as a minicolumn found in diverse cortical regions [101, 123, 45, 63, 171].

Given the number of neurons, the connection probabilities (Table 2.1), and the strengths of individual synapses (Table 2.2 and Table 2.3), we can define the population connectivity of the

$\mathbf{J_{EE}}$	$\mathbf{J_{EI}}$	$\mathbf{J_{IE}}$	$\mathbf{J_{II}}$
0.672	13.2	23.7	11.8

Table 2.4.: Effective population-wise connectivity parameters for 2D network of the mouse V1 layer 2/3 cortex derived from the Allen institute database [6], the connection weights are given in mV.

mouse V1 cortex in layer 2/3 following

$$\begin{aligned} J_{XY} &= j_{XY}\, p_{XY}\, N_Y \\ &= A_{PSP_{XY}} \frac{t_{R_{XY}}/2 + \tau_{D_{XY}}}{\tau_X\,(\Theta_X - V_{0X})}\, p_{XY}\, N_Y, \end{aligned}$$

where the normalization by $(\Theta_X - V_{0X})$ is used such that the LIF reset and threshold can be set to 0 and 1 respectively without affecting the dynamics. The resulting connectivity parameters are given in Table 2.4.

2.4.5. Temporal constraints

Although the static SSN assumes that the system is at steady state, this assumption can be too limiting to model biological neural systems as only dead brains display time-invariant activity patterns [98]. The steady state framework can however be applied in situations where the rate dynamics evolve on sufficiently slow timescales to assume that the network operates at the equilibrium state.

In the case of an abrupt perturbation, a biological neural network will undergo a transient response on a timescale of approximately $20\,\mathrm{ms}$ [64]. At the same time, *in vivo* recordings of multiple cortical areas have reported autocorrelation on much slower timescales, in the order of hundreds of ms [37, 36, 40]. This discrepancy can be explained by the interplay of different timescales. On the one hand, the fast intrinsic timescales of the network (τ_{P_X}) are responsible for the dynamics observed upon sudden changes and drive the system towards its equilibrium. On the other hand, slower dynamics can drive the network while it maintains its state near equilibrium. For instance, if the feedforward input to the network evolves on timescales much slower than the intrinsic time constants (τ_{P_X}), the network dynamics will maintain the rates at equilibrium throughout the dynamic trajectory of the external input. Similarly, changes in excitability due to neuromodulation or adaptation can be modeled as a slow change in the steady state network parameters, while the rate dynamics fulfill the static SSN equation. Overall, the

steady state assumption is not limited to strictly static networks. The steady state SSN is suitable for analyzing processes that unfold over slow timescales.

2.4.6. External input

Finally, the characteristics of the external input must be determined. As previously mentioned, the external input is assumed to be white noise with a mean μ_{ext} and a standard deviation σ_{ext}. Regarding the mean input, we leave this as a free parameter so the network firing rates are studied as a function of the external input. The parameter r, which characterizes the relative external input to the inhibitory and excitatory populations, is by default set to 1. Regarding the noise level σ_{ext}, it is set to $3\,\mathrm{mV}/\sqrt{\mathrm{s}}$. This choice is justified by the corresponding FI curve of the LIF neuron, as this noise level leads to a power law exponent n near 3, which is in line with the experimentally recorded FI curves [140, 175]. The corresponding power law parameters for the *E* and *I* populations are given in Table 2.5.

Given all parameters of the LIF equivalent mouse V1 layer 2/3 network, the spiking network can be simulated and compared to its mean-field SSN counterpart (Fig. 2.3). Overall, we find that the SSN provides an accurate, mathematically tractable prediction of LIF network activity. Small quantitative discrepancies in the predicted firing rates can be caused by the power law approximation of the neuronal activation function or the disregard of recurrent noise.

SSN parameters	**a**	**b**	**n**
Excitatory population	1.08×10^{-4}	-11.1	3.08
Inhibitory population	2.21×10^{-6}	4.8	3.82

Table 2.5.: Power law parameters of the fitted FI curve of LIF neurons, with the input μ in $\mathrm{mV/s}$ and the firing rate ν in Hz. The resulting curves are shown in Fig. 2.2.A.

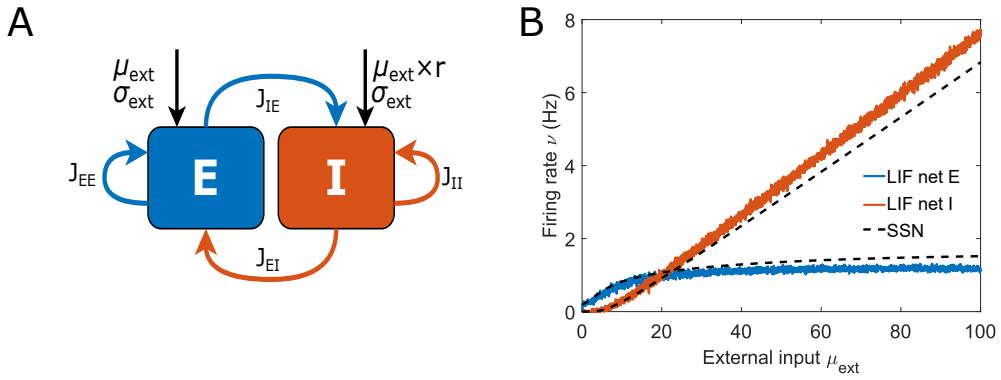


Figure 2.3.: **Simulation of the mouse V1 network based on experimentally reported parameters. A:** Schematic diagram of the network architecture. The E-I network consists of two homogeneous populations recurrently connected with four effective connection strengths. The blue arrows are excitatory (positive connection strength) and the red arrows are inhibitory (negative connection strength). The two populations receive feedforward input in the form of white noise. **B:** The firing rates of the two populations are determined as a function of the external input μ_{ext}. The results obtained from simulating a LIF network matches closely the mean-field prediction obtained with the SSN. The *E* activity saturates with increasing external input, which illustrates supersaturation (see section 3.2).

3. Diversity of network regimes

This chapter is adapted from the result section and the supporting information of the article

Ekelmans Pierre, Nataliya Kraynyukova, and Tatjana Tchumatchenko. "Targeting operational regimes of interest in recurrent neural networks." *PLOS Computational Biology* 19.5 (2023): e1011097.

https://doi.org/10.1371/journal.pcbi.1011097

In this chapter, I study diverse activity regimes accessible to the SSN. For this, I consider two populations (E-I) with their respective power law activation function and with a focus on the $0-10\,\mathrm{Hz}$ firing rate range. While changes in the connection parameters J_{XY} are allowed compared to the values derived from the mouse V1 network (Table 2.4), biological verisimilitude is maintained by imposing that all effective connection strengths remain within the same range.

3.1. Inhibition Stabilized Network

In a neural network with excitatory neurons, recurrent connections can lead to a positive feedback loop, leading to runaway dynamics. As excitatory activity increases, the recurrent E→E feedback grows stronger and drives further increase in excitatory activity. In the absence of inhibition, this mechanism poses a limit to the level of stable activity a network can achieve.

Considering a 1-population system of excitatory neurons with an activation function f_E

$$\tau_{P_E}\dot{\nu_E} = f_E\left(J_{EE}\nu_E + \mu_{ext}\right) - \nu_E,$$

the linearized system around a fixed point follows

$$\dot{\delta\nu_E} = \frac{f'_E J_{EE} - 1}{\tau_{P_E}}\delta\nu_E + \frac{f'_E}{\tau_{P_E}}\delta\mu_{ext}.$$

Since the population time constant τ_{P_E} is positive by definition, the system becomes linearly unstable when

$$f'_E J_{EE} > 1. \tag{3.1}$$

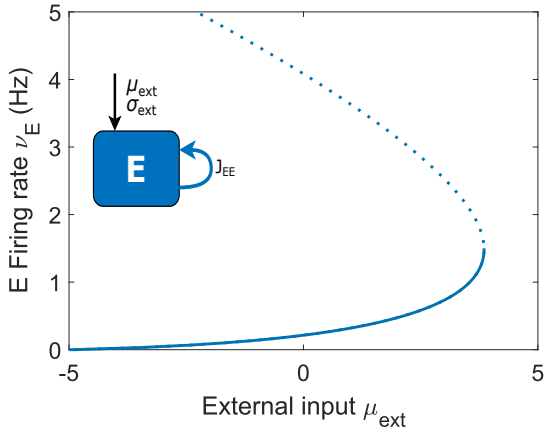


Figure 3.1.: **Instability in the 1D subnetwork.** In the absence of recurrent inhibition, the excitatory network is only stable until a point where excitatory feedback is sufficient to amplify perturbations and lead to runaway dynamics. Higher excitatory activity can only be dynamically unstable (dotted line).

In the SSN, the activation function is given by the power law (Eq. 2.1), such that $f'(x) = na^{\frac{1}{n}}\nu^{\frac{n-1}{n}}$. The stability condition for the 1D network is then an upper limit on the excitatory firing rate:

$$\nu_E < \frac{1}{(J_{EE}^{n_E} n_E^{n_E} a_E)^{\frac{1}{n_E - 1}}}. \tag{3.2}$$

Beyond this level, the fixed point is unstable and perturbations can lead to an unbound self-feeding growth of the excitatory activity (Fig. 3.1).

In the case of an E-I network, high levels of excitatory activity can be sustained thanks to the recurrent inhibition which prevents runaway dynamics [181, 131]. In such cases, the system is in the Inhibition Stabilized Network (ISN) state. For a neural network, the benefits of high recurrent feedback excitation are to amplify feedforward signals [105, 104] and ensure a robust response despite deterioration of incoming signals [152, 103, 16]. In Inhibition Stabilized Networks, recurrent inhibition allows such high gains by ensuring the stability of the system.

The Paradoxical Effect is a feature of the ISN and was first characterized in models of brain rhythms [180, 181]. It was observed that the *E* and *I* populations oscillate in phase when an oscillating external drive only targets the inhibitory neurons. This phenomenon was paradoxical because a phase opposition was expected, where the excitatory activity would be weakest when inhibition is strongest. This phenomenon has also been confirmed experimentally where the theta rhythm of pyramidal cells is in phase with interneurons [168]. Similarly, optogenetic studies have demonstrated a paradoxical response of inhibitory neurons whose activity decreases in response to direct stimulation in multiple brain regions [154]. The Paradoxical

Effect can be defined as a decrease of the inhibitory firing rate as its external input increases: $\frac{d\nu_I}{d\mu_{ext\ I}} < 0$. For a 2-population E-I system with activation functions f_E and f_I for each population, the linearization around a fixed point is [115]

$$\begin{cases} \delta\nu_E &= f'_E \times (J_{EE}\delta\nu_E - J_{EI}\delta\nu_I + \delta\mu_{extE}) \\ \delta\nu_I &= f'_I \times (J_{IE}\delta\nu_I - J_{II}\delta\nu_I + \delta\mu_{extI}) \,, \end{cases}$$

which leads to the following response upon change of external input

$$\begin{bmatrix} \delta\nu_E \\ \delta\nu_I \end{bmatrix} = \frac{\begin{bmatrix} 1 + J_{II}f'_I & -J_{EI}f'_E \\ J_{IE}f'_I & 1 - J_{EE}f'_E \end{bmatrix} \begin{bmatrix} f'_E\delta\mu_{extE} \\ f'_I\delta\mu_{extI} \end{bmatrix}}{|-T\mathcal{M}|}. \tag{3.3}$$

Where $|-T\mathcal{M}|$ is the determinant of the matrix

$$\begin{bmatrix} 1 - J_{EE}f'_E & J_{EI}f'_E \\ -J_{IE}f'_I & 1 + J_{II}f'_I \end{bmatrix}.$$

If the fixed point around which the system is linearized is stable, $|-T\mathcal{M}|$ must be positive [115]. Since the eigenvalues of the matrix $\mathcal{M}$ characterize the stability of the fixed point (Eq. 2.8), the eigenvalues λ_1, λ_2 of $\mathcal{M}$ must have negative real parts if the fixed point is stable. In that case, their product is positive and $|\mathcal{M}|$ is positive. Since T, is a diagonal matrix of positive values (the time constants) and $\mathcal{M}$ has two dimensions, $|-T\mathcal{M}|$ is positive for a stable steady state. This also holds in higher dimensions $d > 2$ as $|-T\mathcal{M}| = (-1)^d \prod_{i=1}^{d} \tau_{P_i}\lambda_i$, which is always positive if all $\mathrm{Re}\,(\lambda_i) < 0$ (following the complex conjugate root theorem).

Around a stable steady state, the Paradoxical Effect condition ($\frac{d\nu_I}{d\mu_{ext\ I}} < 0$) is equivalent to $f'_E J_{EE} > 1$ (Eq. 3.3), which is the condition at which the excitatory subnetwork is unstable. The Paradoxical Effect is therefore a signature of the inhibition stabilized state of an E-I network. It should be pointed out that the *E* activity necessarily decreases as the *I* population is stimulated ($-f'_I f'_E J_{EI}$ is always negative), and it is this suppressed excitation that can lead to the decrease of *I* activity. The Paradoxical Effect and the instability of the *E* subnetwork occur strictly under the same condition in the 2-population network with constant synapses. This is not as straightforward in networks with plastic synapses [190] or networks consisting of more than

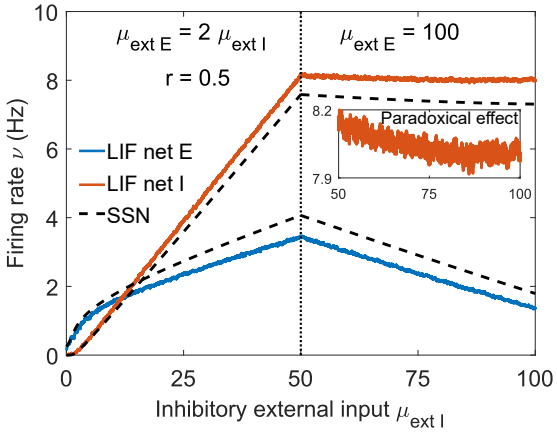


Figure 3.2.: **Demonstration of the Paradoxical Effect in an Inhibition Stabilized Network.** The paradoxical effect is the decrease of inhibitory activity as the inhibitory population is specifically stimulated. In order for the Paradoxical effect to occur, the system must be in the inhibitory stabilized state, so the excitatory activity must be brought sufficiently high. On the left side of the figure until the vertical dashed line, both μ_{extE} and μ_{extI} are increased to reach this point. Beyond the vertical dashed line, only the external input to the *I* population is increased, which results in a reduced *I* activity (see inset) until the *E* firing rate drops below the ISN threshold. The decrease in *I* activity is small because it operates close to the onset of the ISN.

two neuronal populations [152] (see section A.8).

In spiking network simulations, we can predict when the system is in the ISN state by using the condition on the excitatory firing rate derived in the SSN (Eq. 3.2). Once the system is in a state where the *E* subnetwork is unstable, we can demonstrate the paradoxical effect by selectively stimulating the inhibitory population. Using the parameters derived for the mouse V1 network [6], we find that the *E* subnetwork becomes unstable when ν_E is larger than $27.5\,\mathrm{Hz}$. That activity level is difficult to reach in this network because the *E* firing rate tends to plateau unless the feedforward input is strong and primarily targeted at the *E* population, $r \ll 1$ (see Fig. 2.3 or Fig. A.9.B). Furthermore, such high activity levels are beyond the 0-10 Hz range over which we fit the LIF FI curve, making the SSN approximation much less accurate. It appears the considered V1 network does not operate in the ISN state for realistic activity and input levels. In order to enter the ISN at lower firing rates, the gain J_{EE} must be increased. Here, we impose that the ISN threshold occurs at $\nu_E = 1.5\,\mathrm{Hz}$ and find J_{EE} must be increased to $4.75\,\mathrm{mV}$. With this adjusted network, we can probe the Paradoxical Effect, and find that the LIF spiking simulation precisely matches the rate model prediction (Fig. 3.2).

The Inhibition Stabilized state is an important feature of the SSN since it characterizes its stability. Because of the supralinear activation function of the *E* population, recurrent excita-

tion can induce a strong positive feedback loop which causes instability. As long as the ISN threshold is not crossed, the *E* subnetwork is not unstable and the 2D network is guaranteed to be stable (Eq. 2.10 and Eq. 2.9). Once the threshold is crossed, however, the network can only be stabilized through the control of recurrent inhibition. In that case, the stability of the system can still be broken if the dynamic properties of the network (τ_{P_E}, τ_{P_I}) can amplify the *E* positive feedback faster than the *I* stabilizing feedback (Eq. 2.10).

3.2. Supersaturation

While the activity of a neural network is typically driven up by its external drive, the response of a neural network to its external drive can show a range of nonlinear behaviors [31]. The *susceptibility* to external input, introduced in [177], characterizes the relative change of excitatory firing rate to relative changes in external input:

$$\delta = \frac{d\nu_E}{d\mu_{ext}} \frac{\mu_{ext}}{\nu_E}.$$

The susceptibility δ is defined around a fixed point and provides a local power law description of the relationship between ν_E and μ_{ext}:

$$\nu_E = \nu_E^0 \left(\frac{\mu_{ext}}{\mu_{ext}^0} \right)^{\delta}.$$

The response of the network to external input is supralinear if $\delta > 1$ and sublinear if $0 < \delta < 1$. In the case where the response is sublinear, the response is saturating, since further increments of μ_{ext} lead to diminishing increases of ν_E. For $\delta = 0$, the network is fully saturated as the *E* population is impervious to the external drive. Finally, when δ is negative, the *E* population is suppressed by feedforward input. This phenomenon of decreasing *E* activity in response to increase in external input ($\frac{d\nu_E}{d\mu_{ext}} < 0$) is generally referred to as supersaturation [39].

This phenomenon has been widely observed in the visual cortex [39, 183, 100, 133]. In mouse V1, for instance, a substantial number of pyramidal neurons show reduced firing in response to enhanced stimulus contrast [146], while the average activity of thalamic neurons - primarily targeting V1 neurons - is an increasing function of the stimulus contrast [138]. Therefore, *E* neurons can be suppressed despite the increase in external input.

In the SSN, the network has been shown to have a supersaturating regime if $r > \frac{J_{II}}{J_{EI}}$ [4, 134].

Using the linearized system (Eq. 3.3), the supersaturation condition $\frac{d\nu_E}{d\mu_{ext}} < 0$ yields

$$\frac{f'_E\left(1 + f'_I\left(J_{II} - rJ_{EI}\right)\right)}{|-T\mathcal{M}|} < 0,$$

which does require $r > \frac{J_{II}}{J_{EI}}$ if the considered fixed point is stable ($|-T\mathcal{M}| < 0$) and the neuronal activation functions are monotonically increasing ($f'_X > 0$). Using the power law activation function $f'_X = a^{\frac{1}{n_X}} n_X \nu_X^{\frac{n_X-1}{n_X}}$, the supersaturation condition becomes a condition on the inhibitory firing rate [4]

$$\nu_I > \sqrt[n_I-1]{\frac{1}{a_I\left(n_I\left(rJ_{EI} - J_{II}\right)\right)^{n_I}}}. \tag{3.4}$$

Supersaturation occurs because of recurrent inhibition. If the direct effect of the external stimulation on the *E* population is weaker than the increased suppression due to the external stimulation on the *I* population, an increase of feedforward input to the network will lead to a net reduction of input to the *E* population. Due to this, the existence of supersaturation is primarily controlled by the r parameter which characterizes the relative strength of feedforward inputs to both populations.

The existence of stable supersaturation in the SSN is intrinsically linked to the nonlinearity of the neuronal activation function. Without this nonlinearity, the supersaturation condition would be independent of the network activity. In that case, ν_E would be steadily decreasing with μ_{ext}, inevitably leading to negative firing rates. For this reason, the Balanced State cannot support supersaturation: a negative slope is synonymous with negative rates. The nonlinearity of the SSN allows the firing rates to grow until the network reaches the point where supersaturation occurs.

3.2.1. Modulation of the peak *E* firing rate

Recent work [155] compared the responses of LIF and SSN models, pointing out that the peak *E* activity in supersaturating spiking networks is small, and in particular, it is smaller than the SSN peak. As shown in Fig. 3.3.A, the peak firing rates obtained with the two methods are in agreement.

Furthermore, it is possible to control the height of the *E* firing rate peak in both networks such that it can be made arbitrarily high. We begin by analyzing how the value of maximal *E*

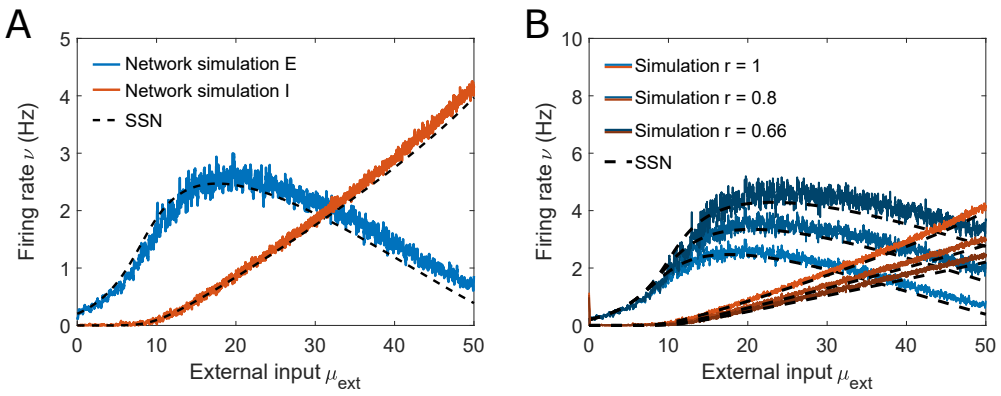


Figure 3.3.: **The SSN-derived condition for supersaturation is confirmed in a LIF spiking network. A:** In the supersaturating regime, the *E* firing rate decreases with increasing external input μ_{ext}. In this network, the SSN predicts that both rates grow until ν_I reaches $0.65\,\mathrm{Hz}$. Beyond that point, the *E* firing rate decreases and eventually reaches 0. This prediction is confirmed in the LIF spiking neural network. **B:** Modulation of the peak *E* activity. By simultaneously modifying r, J_{IE} and J_{EI}^{-1} by a common factor, the amplitude of the *E* curve can be arbitrarily increased.

firing rate depends on SSN parameters for supersaturating networks. A characterizing property of the maximal *E* firing rate ν_E is that it satisfies $\frac{d\nu_E}{d\mu_{ext}} = 0$. This occurs for $\nu_I = (a_I n_I^{n_I}(J_{EI}r - J_{II})^{n_I})^{-\frac{1}{n_I - 1}}$ (Eq. 3.4), with $(J_{EI}r - J_{II}) > 0$. Substituting ν_I in Eq. 2.2, we get

$$a_E^{-\frac{1}{n_E}} \nu_E^{\frac{1}{n_E}} + (\frac{J_{IE}}{r} - J_{EE})\nu_E = \frac{1}{r}(a_I n_I (rJ_{EI} - J_{II}))^{-\frac{1}{n_I - 1}}(1 - \frac{1}{n_I}) - b_E + \frac{b_I}{r}. \tag{3.5}$$

The solution of Eq. 3.5 corresponds to the maximal *E* firing rate in the supersaturating activity regime. To increase the *E* firing rate peak, we can tune r, J_{IE}, and J_{EI}. Specifically, by decreasing r while adjusting J_{IE} and J_{EI} such that the terms J_{IE}/r and $J_{EI}r$ remain constant, we can specifically raise the right-hand side of the equation as $1/r$ increases. If $J_{IE}/r - J_{EE}$ is positive, the left-hand side of the equation is a monotonically increasing function of ν_E, and increasing the right side therefore leads to increases of μ_E to satisfy the equality.

This method assumes that the left hand side of Eq. 3.5 is an increasing function of ν_E (i.e. $J_{IE}/r - J_{EE} > 0$). In that case, the peak of the excitatory activity can be increased indefinitely (Fig. 3.3.B). On the other hand, if $J_{IE}/r - J_{EE}$ is negative, the peak of supersaturation is bounded. This occurs for supersaturating networks for which $\det J$ is negative, as this leads to $J_{IE}/r < J_{EE}$. The peak of supersaturation is therefore bounded, and decreasing r can lead to an unstable situation where no steady state exist (see section 3.4). On the contrary, in the situation where $\det J$ is positive and $r > \frac{J_{IE}}{J_{EE}}$, the peak can be modulated until the point where the peak splits into two branches and any firing rate can be stabilized (see Fig. 3.5.B).

3.3. Bistability

One of the most prominent experimentally recorded neural activity features *in vivo* is the network's ability to switch between higher and lower firing levels. One example is spontaneously alternating intervals of tonic firing and silence observed across different cortical areas [84]. Another example is the sustained firing rate in the prefrontal cortex after stimulus withdrawal during decision-making tasks, which is hypothesized to represent short-term memory [58, 7]. The coexistence of multiple activity states can be explained theoretically by the fact that the network equations can admit multiple solutions for a given set of parameters. In particular, if two stable states coexist, the system is bistable and can sustain its activity in either of the two states for the same level of external input. In the situation where a bistable network can sustain its high activity level in the absence of feedforward input, it is said to have persistent activity. If multiple stable states coexist in a network model, a sufficiently large perturbation can drive network activity away from its current state towards another attractor.

Bistability and persistent activity have been shown to be realizable in balanced networks, using additional sources of nonlinearity such as synaptic plasticity [118] or complex synaptic weight distributions [91]. In the SSN, bistable networks can be obtained without the need of additional ingredients [95], as the neuronal activation function already is a source of nonlinearity.

Unlike supersaturation or the ISN, bistability cannot be delimited by a simple tractable condition on network parameters. It is characterized by the fact that the characteristic function $\mathcal{F}$ (Eq. 2.4) admits multiple zeros which fulfill Eq. 2.7. The search for bistability cannot be performed analytically from $\mathcal{F}$ without numerical methods. We can use the conditions presented in [95] as a starting point to guide the search for bistability in biologically realistic spiking networks. The multistability conditions derived in [95] are necessary but not sufficient to the existence of multistability, and are derived under restrictive assumptions on the power law parameters ($a_E = a_I = 1$, $b_E = b_I = 0$, $n_E = n_I \in \mathbb{N}$). Nonetheless, they provide a helpful starting point from which the parameters can be fine-tuned. One such example is shown in Fig. 3.4.

The LIF network simulation confirms the SSN-predicted bistability (Fig. 3.4.A): the network can sustain either low activity or high activity for external inputs in the $2-4\,\mathrm{mV/s}$ range. Although the SSN rate model is deterministic, the spiking network simulation is not. Due to the

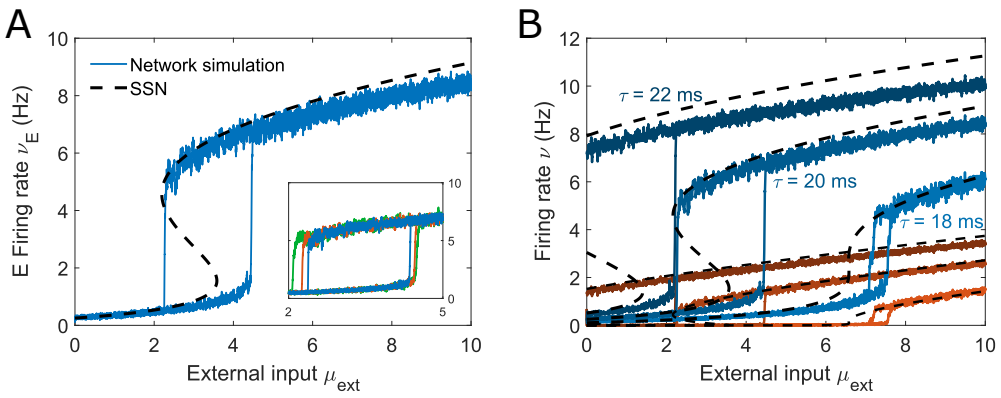


Figure 3.4.: **The SSN-predicted bistability is confirmed in LIF spiking network. A:** Bistability is characterized by the coexistence of two stable steady states for the same level of external input μ_{ext}. Here, the bistability window predicted by the SSN is the range of external inputs between 2 and $3.5\,\mathrm{mV/s}$. The rates of the LIF simulation are in agreement and bistability is confirmed too. The exact location of the bistability window differs slightly between the SSN and the spiking simulation. This can be caused by inaccuracy in the power law approximation of the LIF activation function as the edges of the window are sensitive to changes in parameters (Fig. A.7). Furthermore, the bistability window varies across realizations of the spiking network simulation (see inset) due to the stochastic nature of neuronal spiking which can cause spontaneous transitions if the system is near its edge. **B:** The location and width of the bistability window can be controlled through the membrane time constant τ_E, which characterizes the excitability of the excitatory population. For sufficiently slow τ_E, the network supports persistent activity, where it maintains elevated firing rates in the absence of external input. In the SSN, a change of τ_E translates to changes in the a, b, and n parameters (Fig. 2.2).

stochastic nature of the neuronal activity, fluctuations in firing can cause spontaneous transitions between steady states (shown in Fig. 3.4.A, inset). The width of the bistability window predicted in the mean-field rate model does not account for the fluctuation-driven switches which lead to different switching points in every realization of the spiking network simulation. Such spontaneous transitions between the up and down states have not been reported in the bistable balanced network with short-term plasticity [118]. This is because the fluctuations driving spontaneous transitions are finite-size effects [159], and the switching probability decreases with network size.

It must be noted that LIF simulations of bistable networks can easily lead to synchronization, and result in activity regimes where whole populations spike together to generate population spikes [18, 24] (see Fig. A.8). This violates the assumption of asynchrony and irregularity in the mean-field approach. In order to avoid such effects, synaptic delays and exponential-decay synapses are used to distribute the synaptic input in time (section A.1). These synaptic characteristics only affect the temporal properties of synaptic transmission and therefore do not affect the mean-field steady state.

Interestingly, increasing the excitatory membrane time constant τ_E broadens the window of bistability (Fig. 3.4.B), making bistability more robust to spontaneous fluctuations and easier to locate in phase space. As τ_E increases, the bistability window shifts towards lower feedforward input which can ultimately lead the bistability window to intersect the vertical $\mu_{ext} = 0$ axis, allowing for persistent activity.

3.3.1. Condition on the multiplicity of solutions

As shown in [95], the two-dimensional SSN equation can be rewritten as a single characteristic function, where the steady states of the system correspond to zeros of $\mathcal{F}$ (Eq. 2.4). The number of zero crossings of the characteristic function corresponds to the number of fixed points of the system. Since $\mathcal{F}$ is a continuous function, its number of zero crossings only changes when two solutions merge into one or when one solution splits into two. This corresponds to the situation when extrema of $\mathcal{F}$ fall on zero:

$$\begin{cases} \mathcal{F}(\mu_E) = 0 \\ \mathcal{F}'(\mu_E) = 0. \end{cases} \tag{3.6}$$

Here, $\mathcal{F}'$ denotes the derivative of $\mathcal{F}$ with respect to μ_E. This condition corresponds to changes in the number of solutions. This approach can be used to delimit the range of bistability or absence of solutions (see section 3.4). All parameter sets contained in the same domain delimited by these boundaries have the same number of solutions. The number of such steady states can be obtained by determining the number of zero crossings of $\mathcal{F}$.

Interestingly, since the condition $\mathcal{F}'(\mu_E) = 0$ is equivalent to $|T\mathcal{M}| = 0$ it matches the onset of instability (Eq. 2.9). This shows that the boundary at which the number of solution changes are edges of instability where stable and unstable fixed points are merging or splitting. Similarly, the condition $\mathcal{F}'(\mu_E) = 0$ can be interpreted as points where the firing rate vs external input curve, $\nu\,(\mu_{ext})$, has a vertical tangent since $\mathcal{F} = 0$ and $\mathcal{F} + \mathcal{F}'\delta\mu_E = 0$ (ν_E and $\nu_E + d\nu_E$ are solutions).

It is also worth noting that the ISN is intrinsically linked to the emergence of multiple solutions since $|T\mathcal{M}| = 0$ requires that $f'_E J_{EE} > 1$ (Eq. 2.9). Interestingly, the inhibitory population cannot support multiple solutions in the absence of excitatory activity ($f'_E = 0$ leads to $|T\mathcal{M}| > 0$, see Eq. 2.9), whereas the *E* population alone can (see Fig. 3.1).

The network mechanisms leading to bistability can be understood from the effect of the ISN (Fig. 3.5.A). The *E* activity grows with the external input until the network reaches the point where recurrent excitatory feedback can support itself without relying on the feedforward input to grow. At this point, the curve bends (Eq. 3.6) and the unstable branch progresses with $\frac{d\nu_E}{d\mu_{ext}} < 0$. Eventually, if $\det J > 0$, the cross-population interaction will lead to sufficiently strong recurrent inhibition to cause another bend as further growth will require stronger external input. After this point, the upper branch of the bistability can supersaturate if the recurrent inhibition is sufficiently strong (see section 3.2).

Incidentally, this mechanism where the excitatory activity builds up until recurrent inhibition catches up to stabilize it at a higher level explains how the excitatory membrane time constant τ_E affects the width of the bistability window. Increasing τ_E increases the excitability of the *E* population, which translates into lower values of the power law parameter b_E (see Fig. 2.2). The *E* population starts firing for lower levels of external input and has a larger head start before the recurrent inhibition can catch up, leading to a more developed bistability window.

Finally, there also exists bistable systems for which the unstable branch never merges with the upper stable branch (see Fig. 3.5.B which does not have a critical point IV, as in Fig. 3.5.A). In that case, once the network is in the up state it cannot be pushed to collapse towards the down

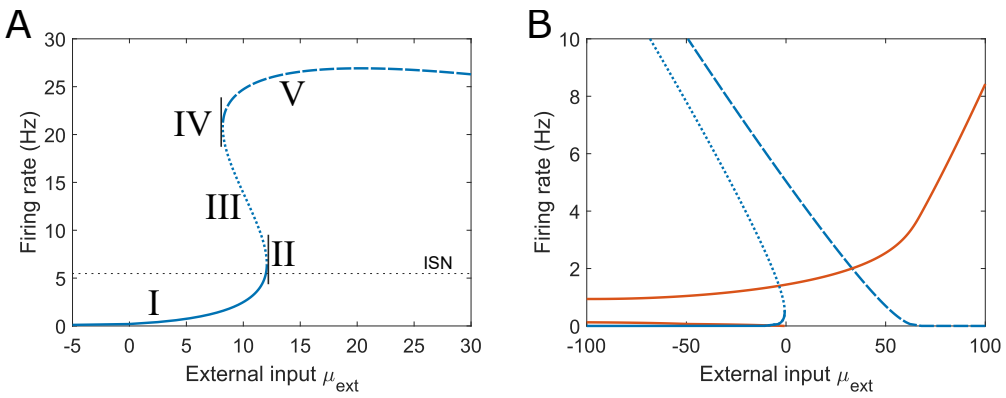


Figure 3.5.: **Breakdown of a bistability curve. A:** Bistability can be broken down into 5 regimes. (I) For low *E* rates below the ISN, the firing rates increase with the external input. (II) The slope of the *E* curve grows until it reaches a point where its tangent is vertical. This happens when $|T\mathcal{M}|$ is 0, which requires that the system is beyond the ISN threshold. Beyond that point, the recurrent inhibition can no longer stabilize the network. (III) The unstable part of the bistability grows for decreasing inputs. (IV) Due to the supralinear growth of the inhibitory population, the *I* activity could reach a point where it can stabilize the system and the curve shows another bend characterized by a vertical tangent. (V) The upper branch is stabilized by inhibition and can eventually supersaturate if $r > \frac{J_{II}}{J_{EI}}$. **B:** In cases where $\frac{J_{IE}}{J_{II}}$ with $\det J > 0$, the unstable branch and the upper branch are separate, so no point IV exists as in panel A (see Eq. 3.8).

state by tuning down its input. Instead, it would require a sudden perturbation to push it out of equilibrium so its activity crosses the unstable separatrix. In such cases, the stable branch keeps rising as the external input diminishes since recurrent excitation must compensate for the feedforward suppression. If the upper branch never collapses, we have $\lim_{\mu_{ext}\to-\infty} \frac{d\nu_E}{d\mu_{ext}} < 0$ for the up branch and $\lim_{\mu_{ext}\to-\infty} \nu_E = 0$ for the down branch. In order for the up branch to have a non-zero fixed point in the limit of infinite external suppression, the $\mathcal{F}$ function (Eq. 2.4) must have two zeros with $\mu_E > b_E$, corresponding to the unstable branch and the up branch. With $\det J > 0$, these two crossings are separated by a maximum μ_E^* of $\mathcal{F}$ (section 3.4). The two fixed points are guaranteed to exist as long as the maximum $\mathcal{F}(\mu_E^*)$ remains above 0. The maximum μ_E^* must satisfy $\mathcal{F}'(\mu_E^*) = 0$:

$$J_{EE} f_E'^* = \left(J_{II} + f_E'^* \det J\right) f_I'^* + 1.$$

In the limit of $\mu_{ext} \to -\infty$, with $f_E'^* \gg 1$, the maximum of $\mathcal{F}$ becomes a simple condition on $f_I'^*$

$$f_I'^* = \frac{J_{EE}}{\det J}.$$

Since the limit of $f_I'^*$, and therefore the limit of $\mu_I(\mu_E^*)$ remain constant as μ_{ext} tends to $-\infty$, the limit of $f_E'^*$ is given by Eq. 2.5

$$f_E'^* = \frac{J_{II} - rJ_{EI}}{\det J}\mu_{ext}. \tag{3.7}$$

By replacing Eq. 3.7 in $\mathcal{F}$, and with $f_I'^*$ constant, we get

$$\lim_{\mu_{ext}\to-\infty} \mathcal{F}\left(\mu_E^*\right) = \frac{\mu_{ext}J_{EI}}{\det J}\left(J_{IE} - rJ_{EE}\right). \tag{3.8}$$

For networks with $\det J > 0$ satisfying the supersaturation condition, the local maximum of $\mathcal{F}$ can be above 0 for arbitrarily strong feedforward suppression. Meanwhile, $\mathcal{F}$ tends to $-\infty$ in the limit of large μ_E. Therefore, the $\mathcal{F}$ function has at least one solution $\mathcal{F}(\mu_E) = 0$ such that $b_E < \mu_E^* < \mu_E$ and $\mathcal{F}'(\mu_E) < 0$. For systems with positive $\det J$, the system is guaranteed to be bistable for sufficiently strong negative feedforward input if $r > \frac{J_{IE}}{J_{EE}}$. While the static stability condition is satisfied (Eq. 2.7 or $|T\mathcal{M}| > 0$), the stability condition linked to the population time constants (Eq. 2.10) will eventually break as excitatory firing rates keep growing. In the case where $\det J < 0$, the highest zero crossing of $\mathcal{F}$ is unstable (Eq. 2.7, section 3.4) so this analysis cannot guarantee bistability in the limit of $\mu_{ext} \to -\infty$.

3.4. Structural instability

The previous section addressed the cases where a system can admit more than one fixed point, and in particular multiple stable steady states. Inversely, there exist cases where a network does not admit a single steady state for a given input level. This phenomenon was shown in the case of the 1D excitatory network (Fig. 3.1), where the *E* firing rate grows with increasing external input until it reaches the ISN threshold (Eq. 3.2), after which the firing rate increases further for decreasing external inputs. For any external input beyond the ISN threshold, the network no longer has any fixed point. Beyond that level, the feedback excitation will always push the activity higher, so no steady state can be reached. While recurrent inhibition has been shown to stabilize high activity and prevent such runaway dynamics, even 2D networks can enter conditions where no steady state exists. One such example is shown in Fig. 3.6, where the *E* and *I* firing rates cannot settle on a steady state for a range of external input and instead tend

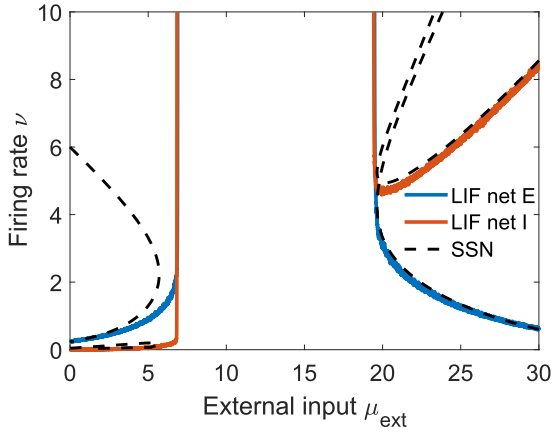


Figure 3.6.: **Structurally unstable networks do not support a finite activity level for a range of external inputs.** The LIF simulation and the SSN both predict the same fixed points outside of the structural instability window. Within this range, the SSN does not have a mathematical solution, and the LIF neurons fire at the maximum rate allowed by the simulation time step. Once the network enters that regime, it cannot be recovered as it escapes the control of the feedforward input.

to ∞.

Each steady state of the system corresponds to zero crossings of the characteristic $\mathcal{F}$ function (Eq. 2.4). Therefore, to study the cases where no steady state exists, one must study how many zero crossings the $\mathcal{F}$ function has, and when it does not cross zero at all.

Parity of solutions

The number of zero crossings of the $\mathcal{F}$ function can be studied through its limits. Assuming that the FI curves are supralinear ($n_E > 1$ and $n_I > 1$), we get:

$$\lim_{\mu_E \to -\infty} \mathcal{F} = \infty$$
$$\lim_{\mu_E \to +\infty} \mathcal{F} = \lim_{\mu_E \to +\infty} a_E J_{EE} {\mu_E}^{n_E} - a_I J_{EI} \left(\frac{a_E \det J}{J_{EI}} \mu_E^{n_E} \right)_+^{n_I}.$$

If $\det J > 0$, the second limit tends to $-\infty$. Therefore, the function has at least one zero and the $\mathcal{F}$ function for positive determinants has an odd number of solutions (Mean-value theorem) [95]. On the other hand, if the determinant is negative, the second term in the $\mu_E \to \infty$ limit is zero because of the $(.)_+$ operator and the limit tends to ∞. In that case, there is no guarantee that the system has any fixed solution and the number of solutions is even [4, 95]. Multiple roots (corresponding to Eq. 3.6) are counted separately in this calculation.

Since networks with positive $\det J$ have an odd number of fixed points, such networks will always have at least one steady state. In the case where the highest steady state is in the ISN, it can be unstable if the ratio τ_{P_E}/τ_{P_I} is low (Eq. 2.10), but will otherwise be stable (Eq. 2.7). In the case where the determinant is negative, the number of solutions to $\mathcal{F}(\mu_E) = 0$ can be 0. The number of solutions would still depend on the external input. For sufficiently strong negative feedforward input, $\nu_E = \nu_I = 0$ is always a solution ($\lim_{\mu_{ext}\to-\infty} \mathcal{F}(\mu_{ext}) = 0$), and it is stable (Fig. 3.7, Eq. 2.7, Eq. 2.10).

Similarly, in the limit of strong excitatory feedforward input, for networks with $\det J < 0$, we can distinguish two cases:

- If the network does not support supersaturation ($r < \frac{J_{II}}{J_{EI}}$), no solution exists in the limit $\mu_{ext} \to \infty$. The function $\mu_I(\mu_E) = \frac{a_E \det J}{J_{EI}} (\mu_E - b_E)_+^{n_E} + \frac{J_{II}}{J_{EI}}\mu_E + \mu_{ext}\left(r - \frac{J_{II}}{J_{EI}}\right)$ has an upper bound and is always lower than b_I for sufficiently high external input μ_{ext}. In that case, the inhibitory population cannot stabilize the excitatory population. If μ_E passes the ISN threshold, in absence of recurrent inhibition it will lead to the absence of fixed point. For sufficiently high external input, all networks which do not support supersaturation and have a negative J determinant will reach a regime where no fixed point exists (see Fig. 3.7.D-E).

- If the network supports supersaturation ($r > \frac{J_{II}}{J_{EI}}$), the network admits a stable solution where the *E* population is silenced in the limit $\mu_{ext} \to \infty$. The characteristic function evaluated at the firing threshold of the *E* population yields $\mathcal{F}(b_E) = \mu_{ext} - b_E - J_{EI} a_I \left(\mu_{ext}(r - \frac{J_{II}}{J_{EI}}) + b_E \frac{J_{II}}{J_{EI}} - b_I\right)_+^{n_I}$, which is negative for sufficiently high external input μ_{ext}. At the same time, for a fixed value of μ_{ext}, we have $\lim_{\mu_E\to-\infty} \mathcal{F}(\mu_E) = \infty$. Therefore, for sufficiently high external input μ_{ext}, there will always be a zero crossing such that $\mu_E < b_E$, which corresponds to a situation where the *E* population is silenced by recurrent inhibition ($J_{EI}\nu_I > \mu_{ext}$). With $\det J < 0$, there must be another unstable solution with $\nu_E > 0$ so the number of solutions is even. For networks which have a negative J determinant and support supersaturation, there is a guarantee that a silent state can be reached for sufficiently high input. Even if there is a window of moderate external inputs for which the network does not have a solution, this window is finite Fig. 3.7.F.

The absence of a fixed point for a network is pathological since it can lead to an irrecoverable transition. Once the network enters the domain where it is structurally unstable and does not

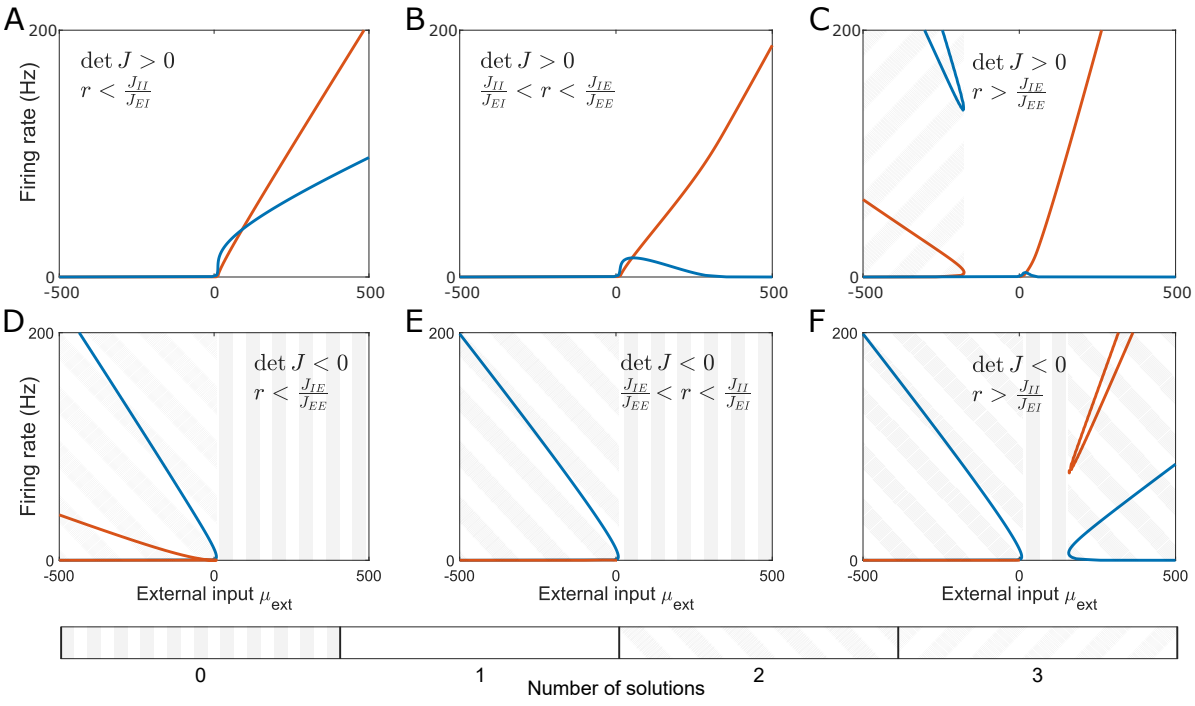


Figure 3.7.: **The 6 different activity regimes of the SSN in the limit of large feedforward input.** A B and C correspond to networks with positive $\det J$, with increasing external input ratios (r) from left to right. D, E, and F correspond to networks with a negative $\det J$. **A:** For low ratios of external input r, the network is silent for strong feedforward inhibitory input and has growing activity for strong excitatory feedforward input. **B:** For values of r that support supersaturation, the *E* activity eventually is silenced for strong excitatory feedforward input. **C:** For values of r larger than $\frac{J_{IE}}{J_{EE}}$, the network supports bistability in the limit of strong feedforward inhibition (see Eq. 3.8), where the low state corresponds to both populations being silent. **D:** For values of r lower than $\frac{J_{IE}}{J_{EE}}$, the system does not support any finite activity for strong feedforward excitation but has two steady states for strong feedforward inhibition: a stable state where both populations are silent and an unstable one where both populations are active. **E:** For values of r larger than $\frac{J_{IE}}{J_{EE}}$, the *I* population is silent in both steady states for strong negative μ_{ext}. **F:** If r supports supersaturation, the network admits two solutions for strong excitatory feedforward input: a stable state where *E* is silent while *I* is active and an unstable state where both populations are active.

have a fixed point, the firing rates will grow without limit. Even if the external input is then changed back to a condition that supports a steady state solution, $\det J < 0$ implies that there is an even number of solutions and that the fixed point with the highest E activity is unstable ($\mathcal{F}' > 0$). This means that regardless of the network regime, there is always an upper separatrix beyond which the activity is out of the basin of attraction of any stable steady state. Because of this, a network that enters the domain of structural instability will undergo runaway dynamics. After this process is initiated, stable activity can only be recovered if the network is moved back quickly to a condition where the state of activity has not yet escaped the basin of attraction of any fixed point.

For networks that have a positive J determinant, structural instability can only occur if all fixed points are in the ISN and population time constants τ_{P_X} break the stability condition (Eq. 2.10). In that case, the positive feedback loop will lead to the same runaway dynamics, but finite activity can always be recovered by modifying μ_{ext} to a level where the highest state activity is below the ISN threshold. This is the case because with $\det J > 0$, the highest zero crossing always satisfies ($\mathcal{F}' < 0$). Therefore, if this fixed point satisfies Eq. 2.10, it is an attractor. Finally, there always exists a level of feedforward input μ_{ext} for which the highest zero crossing of $\mathcal{F}$ corresponds to a silent E population (either $\mu_{ext} \to -\infty$ if $(rJ_{EI} - J_{II})\, J_{EE}/\det J < 1$ -Eq. 3.8-, or $\mu_{ext} \to \infty$ if the network supports supersaturation).

Overall, the absence of fixed points for a given level of feedforward input is a feature of the SSN and is due to the assumption of an unlimited power law activation function. The LIF neurons from which the power law is derived do not satisfy this behavior since the FI curve becomes linear or saturates for high firing rates (Fig. 2.1). In particular, if the FI curve saturates, the firing rates cannot grow indefinitely and will stabilize to a high value near the saturation rate. This will be the case if the neurons have a refractory period. In simulations of LIF neurons, the firing rates are limited by the integration time step dt, which effectively plays the role of a refractory period. That is because neurons cannot spike between them and the firing rates cannot rise beyond dt^{-1}. In this context, it must be noted that the effect of structural instability operates far beyond the domain of validity of the LIF-fitted SSN ($\nu \leq 10\,\mathrm{Hz}$). Nonetheless, the SSN analysis of structural instability can highlight conditions where the firing rates will blow up far beyond the range of experimentally reported firing rates. In biological systems, such self-sustaining hyperactivity is typically linked to pathological conditions such as seizures [85, 61].

3.5. Mapping of regimes

In previous sections, we demonstrated that the SSN framework can be used to locate specific computational regimes such as supersaturation and the Paradoxical Effect in parameter space and that the observations derived from the SSN are confirmed in corresponding spiking neural networks. Since the SSN provides specific conditions for the onset of these operational regimes (Eq. 3.2, Eq. 3.4, Eq. 3.6), they can be accurately located in parameter space. Here, we focus on the 2D space of the feedforward input and the I/E external input ratio (μ_{ext}, r) and illustrate how the SSN can predict the range of parameters in which diverse network properties occur. In particular, we focus on two example connectivity matrices J, one whose determinant is positive and one with a negative $\det J$. We also delimit the input regimes for which the network permits a balanced limit solution. It must be noted that besides characterizing the number of fixed points a system has [4, 95], the sign of $\det J$ also determines whether the system can have a stable balanced state (Eq. A.4) or lacks it. In networks where the sign of $\det J$ is negative, a balanced solution can exist with positive firing rates if $r > \frac{J_{II}}{J_{EI}}$, but it is unstable [147] (section A.3).

Fig. 3.8 shows the map of feedforward inputs and the corresponding computational regime for two examples of the connectivity matrix J. Panel A corresponds to the connectivity parameters from the bistable network shown in Fig. 3.4 with $\det\ J > 0$. The region where supersaturation occurs is located in the top right corner since it is limited to be above a critical value of $r > \frac{J_{II}}{J_{EI}}$ and requires sufficient feedforward input so that the inhibitory activity can reach the supersaturation threshold (Eq. 3.4). Similarly, the region where the ISN occurs is mainly located in the bottom right corner since it requires sufficient feedforward input to reach the required excitatory activity. At the same time, it is pushed down by the supersaturating region, preventing its growth. In particular, for high enough values of the parameter r, the network can never enter the ISN since the recurrent inhibition prevents the excitatory activity from ever reaching the point where the *E* subnetwork would become unstable. The region where bistability is expected matches the results observed in the spiking network simulation in Fig. 3.4 with $r = 1$. Interestingly, its edge is located near the onset of the ISN. This can be explained by the fact that the onset of bistability imposes a stricter condition than the ISN, which takes into account the effect of recurrent inhibition. Since this network reaches the ISN while the inhibitory activity is still low, it nearly immediately becomes bistable (Eq. 2.9 with $f_I' \approx 0$). It must be noted that the up state within the bistable window is necessarily in the ISN. Regarding the domain of validity of

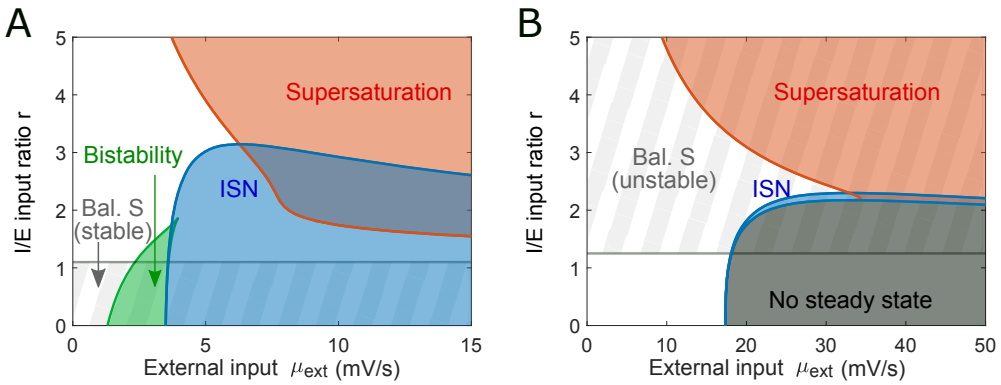


Figure 3.8.: **Mapping the computational states of the SSN model in input space for two representative connectivity regimes.** Using the conditions for the onset of the ISN (Eq. 3.2), the onset of supersaturation (Eq. 3.4), the merge and split of multiple solutions (Eq. 3.6), and the condition for the existence of a Balanced State limit (Eq. A.4) we can determine where the different regimes occur with respect to the two feedforward input parameters μ_{ext} and r. **A:** The J parameters chosen here correspond to the bistable network shown in Fig. 3.4 and represents the situation $\det J > 0$. Multiple operational regimes can coexist since the domain of supersaturation and the ISN overlap. The bistable domain closely borders the ISN domain since the low-activity branch ends soon after the ISN onset, and the high-activity branch is always in the ISN. With $\det J > 0$, the network admits a balanced solution if r is below $\frac{J_{II}}{J_{EI}}$, so it is incompatible with supersaturation. **B:** The J parameters chosen here represent a situation with $\det J < 0$. Here, the ISN domain is limited because recurrent inhibition cannot stabilize high excitatory activity unless the parameter r is high. Additionally, a range of external inputs cannot support a steady state activity, leading to structural instability. With $\det J < 0$, the Balanced State is necessarily unstable and requires $r > \frac{J_{II}}{J_{EI}}$.

an E-I Balanced State, it is defined and is stable for low values of r but does not exist beyond $r > \frac{J_{II}}{J_{EI}}$.

The connectivity parameters in Panel B correspond to a network with $\det J < 0$. While the supersaturation region operates the same as in Panel A, the ISN appears to be much less accessible, as a large part of the bottom right corner is occupied by structural instability, where no steady state exists. Here again, the boundary of the structural instability closely approaches the boundary of the ISN. This is due to the fact that the *E* subnetwork is unstable in both regimes, but in the case of the ISN, it is stabilized through recurrent inhibition. Finally, with $\det J < 0$, the Balanced State is only defined for sufficiently high values of the parameter r, but it is dynamically unstable (section A.3).

3.6. Controllability

One key advantage of the SSN over the mathematically exact Ricciardi self-consistent rate model (Eq. 1.4) is the simplicity of its mathematical formulation. This makes it easier to manipulate, make predictions, or design network parameters to achieve a desired effect. Knowing how the 2D firing rates emerge from recurrent and feedforward connectivity in the SSN allows us to invert this relation and select external inputs such that they lead to the desired *E* and *I* activity trajectory in the spiking network. This is illustrated in Fig. 3.9, where we target a complex 2D trajectory by defining dynamic feedforward inputs μ_{extE} and μ_{extI}. We obtain the feedforward inputs that result in the desired dynamics $\nu_E(t)$ and $\nu_I(t)$ by inverting Eq. 2.2:

$$\mu_{extE}(t) = \left(\frac{\nu_E(t)}{a_E}\right)^{1/n_E} + J_{EI}\nu_I(t) - J_{EE}\nu_E(t) + b_E$$
$$\mu_{extI}(t) = \left(\frac{\nu_I(t)}{a_I}\right)^{1/n_I} + J_{II}\nu_I(t) - J_{IE}\nu_E(t) + b_I.$$

Fig. 3.9 illustrates how the obtained targeting can lead to a high fidelity between the target and simulated trajectory. Notably, the obtained timescale of the autocorrelation function of neuronal activity (as defined in [122]) is around $300\,\mathrm{ms}$, which is in line with experimentally recorded cortical activity [37, 36, 40]. These results indicate that complex dynamic trajectories evolving on biologically realistic timescales can be accurately captured by the SSN steady states (Eq. 2.2). It follows that the mapping between the steady states of the SSN and spiking

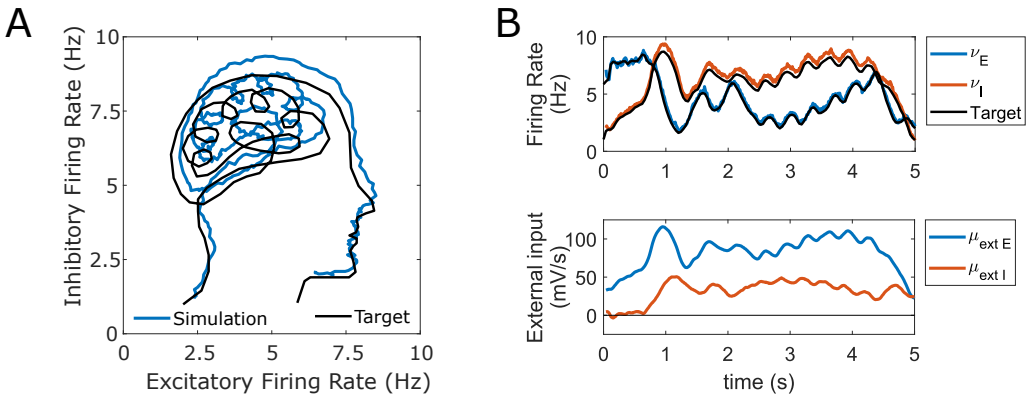


Figure 3.9.: **A LIF spiking network can be designed to follow a user-defined trajectory by defining a feedforward input trajectory using the SSN. A:** The spiking network can accurately follow an intricate arbitrarily defined trajectory in the 2D firing rate space. Given the recurrent connectivity, the SSN equations can be inverted to return the feedforward inputs which will lead to the desired firing rates. The small distortion at high rates is due to the recurrent noise, which is not accounted for in the SSN but raises the FI curve of simulated LIF neurons. **B:** Time trajectory of the 2D simulated firing rates alongside the target (Above) and the corresponding 2D feedforward input trajectory (Below).

neural networks provides a valuable approximation even for spiking network dynamics at slow timescales. Although this approach is guaranteed to provide the values of external inputs for which the target rates ν_E and ν_I are a steady state, if that state is unstable (which does not depend on the feedforward inputs μ_{extE} and μ_{extI}), it will not be accessible through a spiking simulation of LIF neurons. Similarly, if the network is bistable, the LIF simulation can land on a different equilibrium depending on the history of the network.

We used here dynamic feedforward inputs to move along the activity trajectory to reflect the control of upstream brain regions. It is equally possible to use the SSN to tune other parameters, such as the connectivity or power law parameters, to reach any user-defined network state or trajectory.

3.7. Computational implications

The broad range of nonlinear responses accessible to simple E-I networks of LIF neurons can be the foundation for diverse signal processing and computations which neural networks execute. In order to perform complex tasks, systems such as electronic circuits or artificial neural networks rely on the nonlinear behavior of their basic components, such as transistors or rectified linear units [76, 160]. On the other hand, due to their linear response function, E-I balanced networks are very limited in the computations they are capable of carrying out, and extensions

of the model have been proposed to permit richer and higher dimensional dynamics [14, 118].

Here, we illustrate the computational capabilities of networks of LIF neurons by demonstrating that they can robustly perform logical operations. One such example is the XOR operation which is one of the key computing components of logical circuits and is famously challenging to implement in a neural network [117]. We find that a supersaturating network can take advantage of its nonlinear response to approximate a XOR operation on two inputs (Fig. 3.10.C), as the excitatory activity is weak for low inputs (below threshold activation) and high inputs (silencing due to recurrent inhibition) but has a clear peak of activity for moderate input. In order for the network to reach its peak activity, it must receive input from one of its two signal sources, but not both at the same time. Similarly, one could design an approximate NOT logical gate using a supersaturating network where the peak activity would be reached for limited input but would be suppressed for any significant incoming signal.

The sharp transition between low and high activity in bistable networks can also be leveraged to perform computations. Here, we demonstrate how a bistable network can perform the binary AND operation on two incoming signals (Fig. 3.10.D). The network is in the down state for low inputs and can only collapse to the up state if it receives sufficient input, which occurs if both its input sources are active. Similarly, the same mechanism can be used to perform the OR operation if the down-up transition can occur for sufficiently low inputs that either of the two signals could trigger the switch on its own. Another valuable feature of bistable networks, besides their sharp transition between states, is their ability to retain information about prior states. Since a bistable network will remain in the up or down state as long as its input does not escape the bistable window, its activity can serve as binary information storage.

While we only demonstrated how a single E-I network can perform unitary binary operations, more complex computations can build upon this mechanism by combining multiple such E-I networks together. In the same fashion, as electronic circuits can perform a wide range of tasks by associating multiple components with different response properties, multiple E-I networks with diverse recurrent structures can interact to accomplish a richer repertoire of information processing. As an example, the association of supersaturation-based XOR and bistability-based AND operation can lead to the implementation of a bit-wise half adder circuit from LIF neurons Fig. 3.10.B.

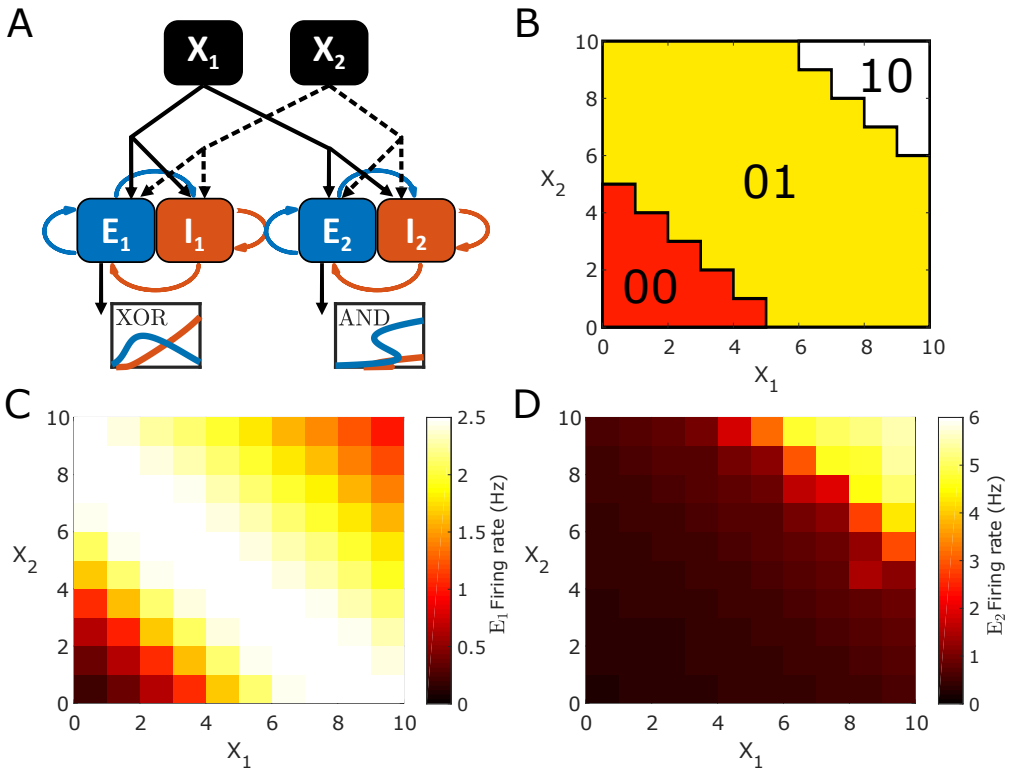


Figure 3.10.: **The nonlinearities in spiking networks of LIF neurons can be used to perform operations, such as bit-wise summation. A:** Schematic representation of the network architecture. Two E-I networks simultaneously process the feedforward input which originates from two independent external populations X_1 and X_2. The firing rate of each *E* population is considered as readout. **B:** If the two subnetworks perform the XOR and AND operations, the circuit is a half-adder and can execute a bit-wise summation of the two inputs. The XOR subnetwork determines the last digit and the AND network determines the first digit. **C:** The XOR operation can be performed using a supersaturating network. The *E* activity has a peak for moderate input level and is low otherwise. **D:** The AND operation can be performed by a bistable network. The *E* activity is low on the down branch and has a sharp transition to the up branch at the end of the bistability window. Here, we assume the system is initialized at $X_1 = X_2 = 0$ so the possible memory implications of bistability are not considered (i.e., remaining within the bistability window in either state between signals).

3.8. Convergence to the Balanced State

We have found that biologically-sized networks can generate diverse nonlinear responses to external input. Meanwhile, the Balanced State limit implies that network response becomes linear as network size approaches infinity. How do networks transition from nonlinear to linear regimes for increasing network size N?

To tackle this question, we rescale the recurrent connections j_{XY} by the factor $1/\sqrt{N}$ as a function of network size N and increase N from $N = 4 \times 10^3$ to 5×10^5 while keeping the connection probabilities fixed, leading to an effective re-scaling of the population-wise J_{XY} by $\sqrt{N}$. This parameter re-scaling follows the convention of the Balanced State theory [185, 186] and allows us to investigate whether these nonlinear spiking networks converge to the expected Balanced State and, if so, when and how. In the Balanced State convention, the feedforward input follows the same rescaling and grows with $\sqrt{N}$. This is because the external input can be seen as originating from another population located outside the network. The three populations (*E*, *I*, and ext) follow the same scaling where the individual synaptic weights decrease with network size, but the effective population-wise connection strength does increase with network size N.

A dynamically stable Balanced State limit can only exist if $\det J$ is positive and the fraction of external input weights r satisfies $r < \frac{J_{II}}{J_{EI}}$ (see Eq. A.4). In this network convergence study, we focus on three spiking networks: one supersaturating network with $\det\ J > 0$ (same connectivity parameters as in Fig. 3.3), one bistable network with $\det J > 0$ (same connectivity parameters as in Fig. 3.4), and a supersaturating network with $\det J < 0$ (same connectivity parameters as in Fig. 3.8.B, r=3). Among these three example networks, one does not admit a Balanced State limit (supersaturation), one supports a stable Balanced State solution (bistable network), and one does support a Balanced State which is dynamically unstable ($\det J < 0$ with $r > \frac{J_{II}}{J_{EI}}$).

We find that in all three cases, the SSN model remains an accurate description of the spiking network mean activity across different network sizes N (Fig. 3.11.A-C, up to 40 000 neurons). Beyond this network size, spiking network simulations become increasingly prone to synchronization, which makes the comparison with the SSN impractical. Nonetheless, we verify that the behavior of the SSN for sizes up to 500 000 neurons still qualitatively and quantitatively matches the predictions obtained using the Φ_{sc} self-consistency solution (Fig. 3.11.D-F, inset).

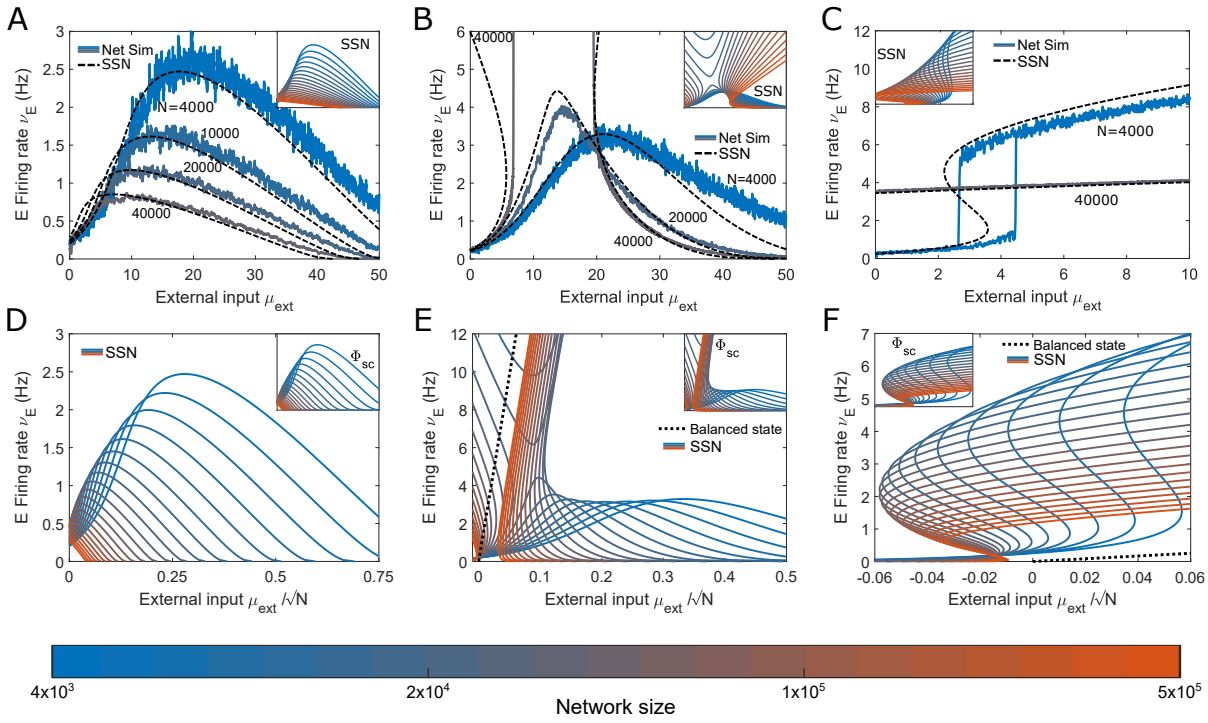


Figure 3.11.: **Convergence to the Balanced State of nonlinear network responses as the network size N increases.** Panels A, B, and C show the LIF spiking simulations of three networks at different sizes N, along with the corresponding SSN prediction. The SSN prediction matches the LIF simulation even for large network size N. Panels D, E, and F show the same three networks as a function of the external input before scaling to highlight a possible convergence to the Balanced State limit. The Balanced State theory assumes that the effective recurrent and feedforward connection strength scale with $\sqrt{N}$. The SSN matches qualitatively and quantitatively the self-consistency equation Φ_{sc} derived from the Ricciardi equation, (Eq. 1.4) (inset). **A, D:** A supersaturating network with $\det J > 0$. This network does not support a balanced solution (Eq. A.4). As N grows, the inhibitory suppression grows stronger and the excitatory activity peak shrinks. In the limit of infinite size, the excitatory population is permanently silenced. **B, E:** A supersaturating network with $\det J < 0$. This network supports a balanced solution, but it is dynamically unstable (Eq. A.4). As N increases, the network becomes structurally unstable since the stable and unstable branches collide (see Panel B inset) and leave a gap where no steady state solution exists. In the limit of infinite size, the unstable solution beyond the gap converges to the balanced limit (dashed line), while the stable solution tends to 0. **C, F:** A non-supersaturating network with $\det J > 0$ supporting bistability. This network supports a stable balanced solution (Eq. A.4). As N increases, the firing rates decrease and the bistability window shifts to lower external inputs, leading to persistent activity. When the external input is considered before scaling (Panel F), the width of the bistability window decreases with N. In the limit of infinite N, the up branch of the bistability will converge to the Balanced State. Still, this convergence can be remarkably slow. Even for unrealistically large network sizes (up to half a million neurons), the firing rates are far from the balanced limit (dashed line).

This verification is important since the SSN is nonlinear in its definition, and we must control that nonlinearities that remain for large network size N are not specific features of the SSN model.

We find that the network response can remain nonlinear even for very large network sizes consisting of up to half a million neurons (see inset). By considering the feedforward input before scaling ($\mu_{ext}/\sqrt{N}$) [185, 186], the network response should converge toward a single linear solution as N grows: the balanced limit (section 1.2.4).

In the case of the supersaturating network (Fig. 3.11.A,D), the balanced limit does not exist, as it would lead to negative *E* firing rates. Therefore, in the limit of large network size, the *E* firing rate tends to zero. It becomes consistently suppressed as it receives significantly more recurrent inhibition than feedfoward input, while the inhibitory population is itself at balance with the external input.

For the network with $\det J < 0$ (Fig. 3.11.B,E), the network exhibits supersaturation for $N = 4000$ (see Fig. 3.8.B). However, as N increases, the network enters a region of structural instability where it has no steady state and the firing rates blow up. The inset shows how this instability is caused by the collision of the two steady state branches, leaving a gap where the firing rates are unbound. For this network, one of the mean-field solutions converges to the balanced limit as N increases (Fig. 3.11.E). However, the Balanced State limit is unstable here, and it only matches with the unstable mean-field solution (high firing rates part of the branch) whereas the stable low activity solution tends to zero.

Finally, for the bistable network (Fig. 3.11.C,F), a Balanced State limit does exist and is stable. However, even though the network response must converge toward it as its size grows, the convergence can be very slow. Here, we find that the network response is still far from converging to the balanced limit and remains distinctly nonlinear, with persistent activity, even for networks of unrealistically large size (up to $N = 5 \times 10^5$).

Overall, our example networks illustrate that for many classes of spiking networks with biologically plausible sizes and connectivity configurations, the activity will escape the predictions of the Balanced State. Depending on the parameters, the Balanced State might not exist or be unstable. Furthermore, even when a stable Balanced State solution does exist, it is only expected in the limit of infinite size N. Convergence to this limit can be slow, meaning that network activity can significantly deviate from this prediction, even for unrealistically large sizes N. Nevertheless, this phenomenon is especially remarkable in this example since its response

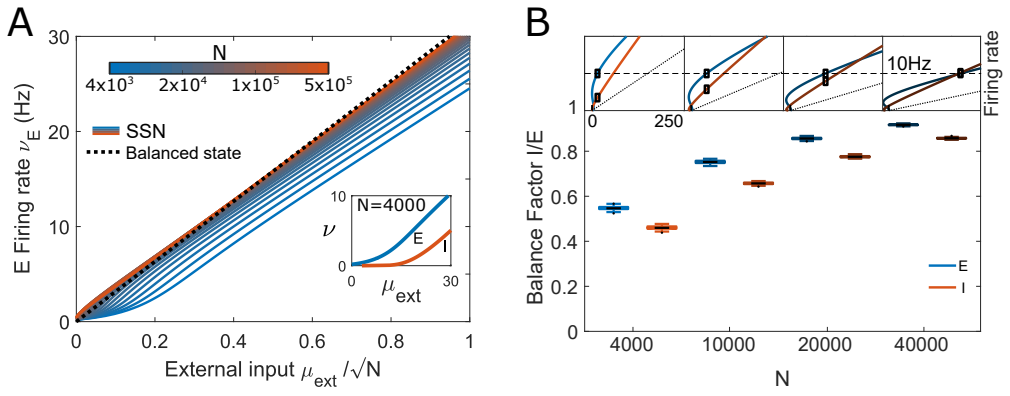


Figure 3.12.: **Quality of the Balanced State convergence depends on the linearity of the network response and the considered balance criterion. A:** Unlike examples shown in Fig. 3.11, networks can converge to the Balanced State for network sizes in the order of 10^4 neurons. The speed of the convergence depends on the connectivity parameters. If a network has a linear response at small sizes (see inset, N = 4000), the convergence to the Balanced State can occur faster. **B:** A high degree of E-I balance does not guarantee that the firing rates conform to the Balanced State. The Balanced Factor (BF) is the ratio of inhibition and excitation received by neurons in a recurrent network (Eq. 3.9). In the Balanced State limit, perfect cancellation is achieved and BF reaches 1. In the bistable network shown in Fig. 3.11.C,F, the BF indeed tends to 1 as N grows, showing that balance is getting tighter. Nonetheless, the network activity is still far from the firing rates predicted in the balanced theory. The BF is measured for four different network sizes at the point where ν_E = 10 Hz (see inset).

curve is highly nonlinear, and the convergence toward the linear balanced limit is therefore particularly slow. Depending on the connectivity parameters, a reasonable convergence to the balanced limit can be achieved for smaller network sizes (Fig. 3.12.A).

Next, we investigate whether a network that is nonlinear and does not conform to the Balanced State prediction can demonstrate other features of E-I balance as N increases. We introduce the balance factor (BF) to measure how close a network is to E-I balance. It is defined for population X receiving positive external input as

$$BF = \mu_{XI} / \left(\mu_{XE} + \mu_{extX}\right). \tag{3.9}$$

If the network operates at balance, the recurrent inhibitory input will cancel out the total excitatory input and lead to a BF of 1. If BF is much lower than 1, then the neurons of population X receive an excess of excitation and the network is unbalanced. It should be clarified that the balance we are considering here must be understood in the sense of *tight balance* [185, 186, 144], meaning that inhibition matches the excitation and leads to near-perfect cancellation. While a partial cancellation is considered a *loose balance* [150], it does not lead to characteristic features

such as predictable firing rates and a linear network response to external input. We measure the BF for different network sizes in a bistable network. The point at which BF is measured for each network size is chosen to be where ν_E reaches $10\,\mathrm{Hz}$. In the case of N=4000 neurons, the values of BF are near 0.5 (Fig. 3.12.B), which is an indicator that the network is operating far from balance as neurons in both populations receive nearly twice as much excitation as inhibition. However, the BF is network size dependent: increasing the size from 4000 to 40000 neurons strongly increased the ratio of inhibition to excitation. Meanwhile, the firing rate activity does not conform to the Balanced State solution (Fig. 3.12.B) even though the network response becomes locally more linear as N increases.

In summary, the observation of significant cancellation of incoming excitatory and inhibitory signals does not guarantee that the Balanced State framework is applicable to predict the firing activity. Even small deviations from tight balance (BF <1) can lead to significant deviations in the resulting network activity regime.

3.9. Analysis of the Balanced State convergence within the SSN framework

The analysis of the characteristic function $\mathcal{F}$ can lead us to determine the conditions under which the SSN can admit a balanced solution. Assuming positive external input and applying the conventional $\sqrt{N}$ scaling, we get:

$$\begin{aligned}\mathcal{F}\left(\mu_E\right) &= J^0_{EE}\sqrt{N} f_E\left(\mu_E\right) - J^0_{EI}\sqrt{N} f_I\left(\mu_I(\mu_E)\right) - \mu_E + \mu^0_{ext}\sqrt{N} \\ \mu_I(\mu_E) &= \left(\det J^0 \sqrt{N} f_E(\mu_E) + J^0_{II}\mu_E + \mu^0_{ext}\sqrt{N}\left(r J^0_{EI} - J^0_{II}\right)\right)/J^0_{EI},\end{aligned} \tag{3.10}$$

where the superscript x^0 indicates the value of the parameter before scaling. In the limit of large network size N, assuming $f_E, f_I > 0$ the solution of $\mathcal{F}$ must satisfy

$$J^0_{EE} f_E - J^0_{EI} f_I + \mu^0_{ext} = 0.$$

If f_I is finite in the $N \to \infty$ limit, the $\mu_I(\mu_E)$ equation leads to

$$\det J^0 f_E = \mu^0_{ext}(J^0_{II} - r J^0_{EI}).$$

Taken together, these conditions lead to the predicted firing rates of the Balanced State limit (Eq. 1.5) [186]. This means that the only steady state solution of the SSN such that ν_E and ν_I are not silent in the limit of $N \to \infty$ is precisely the balanced limit. In the SSN, non-zero firing rates can only be achieved if the input of both populations is above the firing threshold $\mu_E > b_E$ and $\mu_I > b_I$. If $\det J < 0$, this condition imposes $r > \frac{J_{II}}{J_{EI}}$ (Eq. 3.10). This is shown in Fig. 3.11.E, where supersaturating networks with a $\det J < 0$ have a high activity unstable solution. The instability of this balanced state can be recovered from the condition on the slope of $\mathcal{F}'$ (Eq. 2.7) where the term in $\det J$ dominates other terms and prevents a stable balanced state

$$\mathcal{F}' = \sqrt{N} J^0_{EE} f'_E - \sqrt{N} J^0_{II} f'_I - N \det J^0 f'_E f'_I - 1 < 0.$$

If $\det J > 0$, the condition $\mu_E > b_E$ imposes $r < \frac{J_{IE}}{J_{EE}}$. Furthermore, if the network supports supersaturation ($r > \frac{J_{II}}{J_{EI}}$), μ_I grows with both μ_{ext} and μ_E. Therefore, any increase of ν_E as μ_{ext} grows will lead to a supralinear increase of ν_I. Because of this, recurrent inhibition cannot be fine-tuned to balance the excitatory input and such supersaturating networks cannot support a balanced limit. This shows that the conditions for the existence of a stable Balanced State (Eq. A.3) can be recovered from the analysis of the SSN in the limit of strong recurrent connections J and feedforward input μ_{ext}. Moreover, the SSN is bound to converge to the balanced limit when such a stable balanced state exists.

3.9.1. Slope of the firing rate curve

Similarly, the analysis of the slope of the firing rate curve with respect to external inputs shows that compatible networks will converge to a balanced slope for strong feedforward input. This convergence is independent of network size and the conventional rescaling of recurrent connection strengths.

The slope of the firing rate curve can be obtained from linearizing the characteristic function around a fixed point

$$\mathcal{F}(\mu_E, \mu_{ext}) = \mathcal{F}(\mu^0_E, \mu^0_{ext}) + \frac{\partial \mathcal{F}}{\partial \mu_E} d\mu_E + \frac{\partial \mathcal{F}}{\partial \mu_{ext}} d\mu_{ext},$$

where μ^0_E is a fixed point of the network at the external input μ^0_{ext}. We seek a value μ_E that is a fixed point corresponding to the external input μ_{ext}. The characteristic function must be zero at

both points and we get

$$\frac{\partial \mathcal{F}}{\partial \mu_E} d\mu_E = -\frac{\partial \mathcal{F}}{\partial \mu_{ext}} d\mu_{ext}.$$

Using the activation function to convert the input into firing rates ($d\nu_E = f'_E d\mu_E$), the slope is given by

$$\begin{aligned}\frac{d\nu_E}{d\mu_{ext}} &= -f'_E \frac{\frac{\partial \mathcal{F}}{\partial \mu_{ext}}}{\frac{\partial \mathcal{F}}{\partial \mu_E}} \\ &= \frac{f'_E + f'_E f'_I (J_{II} - r J_{EI})}{1 + f'_I J_{II} + f'_E f'_I \det J - f'_E J_{EE}}.\end{aligned}$$

For networks that are compatible with the Balanced State (i.e., non-supersaturating), ν_E and ν_I grow with the external input. Furthermore, in the SSN, f'_E and f'_I grow with the firing rate. Therefore, as μ_{ext} grows, the terms with $f'_E f'_I$ will dominate over the others and the slope will converge towards the balanced slope limit: $(J_{II} - r J_{EI})/\det J$.

This shows that the rich repertoire of network behaviors that deviate from the balanced limit is limited to moderate levels of external inputs. If the external input becomes large, the network activity will either be structurally unstable ($\det J < 0$), silenced ($r J_{EI} < J_{II}$) or linear. This is also illustrated in the convergence to balance as N increases, where network response can remain nonlinear for large network size if we focus on the low external input regime (Fig. 3.11 panels A,B,C compared to D,E,F).

4. Inhomogeneous Networks

In the previous sections, we assumed that the network is composed of two homogeneous populations. This means all neurons within each population have the same properties, receive the same mean external input (μ_{ext}), have the same number of recurrent connections ($C_{XY} = N_Y p_{XY}$) and all synapses between two populations have the same strength (j_{XY}). The noise received by each neuron in a population has the same variance, but it is a private realization: The process of the feedforward white noise is specific to each neuron, and the recurrent noise depends on the specific identity of the presynaptic neurons and their spike train. Due to this assumption, all neurons in a population receive the same mean input but a unique noise pattern. This design ensures that the whole population fires at the same rate while avoiding synchronization.

In this section, we investigate networks that break the homogeneity assumption. We consider inhomogeneous populations such that each neuron can receive different levels of mean input. We distinguish *unstructured networks* in which the properties of the neurons are randomly distributed across the population and *structured networks* in which the intra-population variability is explained by an intrinsic neuronal feature.

4.1. Unstructured networks

In the mean-field approximation, the description of a high-dimensional system is simplified to the average property of its components. While the focus on the collective properties within the system provides a valuable approximation to analyze it, it can be challenging to estimate to which degree of detail the population can be simplified without leading to significant inaccuracy. In previous chapters, the state of the network could be fully described by the average firing rate of each population. In homogeneous networks, the firing properties do not vary across the population. Deviations from the average firing rate are only caused by temporal fluctuations and correlations, which are finite-size effects [25, 28, 159] and are outside the scope of the mean-field framework. In cases where the network is inhomogeneous, the firing rates can vary within a population. In that case, due to nonlinear interactions within the system, the average firing rate can itself depend on higher moments of the firing distribution. We will focus here on cases where the distribution of inputs within a population can be characterized by a Gaussian.

4.1.1. Distribution of firing rates

The input received by a neuron i in an inhomogeneous population is given as

$$I^i(t) = \mu_0 + \varsigma\eta_q^i + \sigma\eta_t(t),$$

Where η_q and η_t are two normally distributed random variables such that

$$\begin{aligned}\langle\eta_t(t)\rangle &= 0\\ \langle\eta_t(t) \times \eta_t(t')\rangle &= \delta\left(t - t'\right)\\ \langle\eta_q\rangle &= 0\\ \left\langle\eta_q^i \times \eta_q^j\right\rangle &= \delta_{i,j}.\end{aligned}$$

The mean input μ_0 is the average across time and over the whole population. The two sources of variability are very different in nature. The quenched noise characterized by its amplitude ς is time-invariant and defines how the baseline input varies from neuron to neuron. The temporal noise characterized by σ is responsible for the fluctuations in input over time to a neuron. Therefore, in the transfer function of the LIF neuron, the mean input is given by

$$\mu^i = \mu_0 + \varsigma\eta_q^i,$$

where σ plays the same role as in the homogeneous network (see Eq. 1.3). Similarly, in the SSN, σ is included in the fitted a, b, and n parameters and the power law applies to the time-invariant input μ^i.

While the input distribution across the population follows a simple Gaussian, the firing rate distribution is itself distorted by two nonlinearities due to the neuronal activation function. The first nonlinearity is the silencing cutoff which occurs for inputs lower than the activation threshold b of the power law (Fig. 4.1.A-B). This effect is especially prominent for low mean input μ_0. The second source of nonlinearity is the supralinear power law which results in an asymmetric skewed transformation (Fig. 4.1.C-D) and is especially marked for large mean input μ_0. Due to the distortion of the firing rate distribution, the population-average firing rate shifts with the quenched variability ς

$$\langle\nu\rangle = E[f\left(\mu^i\right)] \neq f\left(\mu_0\right) = \nu_0.$$

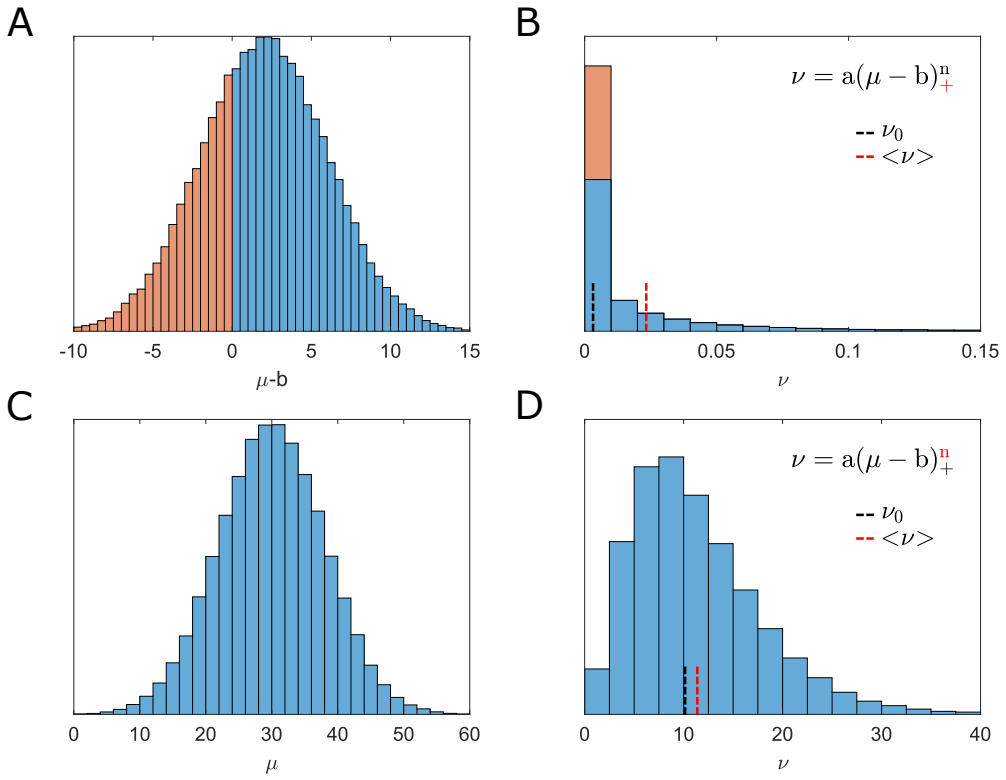


Figure 4.1.: **Distortion of the firing rate distribution due to the nonlinear activation function. A:** For inputs near the firing threshold (b parameter of the fitted power law, a fraction of the population will receive subthreshold input (shown in red) and remain silent. **B:** Distribution of firing rates in a population fed with the input distribution shown in A. The red bar shows the fraction of the population at 0 Hz, which corresponds to the red fraction in A, regardless of the specific value of the subthreshold input they are fed. **C:** Shows another distribution of inputs such that input is higher than threshold for the whole population. **D:** The supralinear power law transformation leads to an asymmetrical distortion of the distribution and a long tail of high firing rate neurons. Due to this asymmetry, the mean firing rate (dashed red line) is higher than the median firing rate (dashed black line), which corresponds to the firing rate of a neuron fed with the mean input of the population (C).

The asymmetric firing rate distribution $\mathcal{P}_\nu$ is derived using the inverse neuronal activation function f^{-1} (Eq. 2.1)

$$\begin{aligned}\mu =& f^{-1}(\nu)\\ f^{-1}(\nu) =& \left(\frac{\nu}{a}\right)^{\frac{1}{n}} + b,\end{aligned}$$

and the probability distribution of the neuronal input within the inhomogeneous population, which we assume to be Gaussian

$$\mathcal{P}_\mu(\mu) = \frac{1}{\varsigma\sqrt{2\pi}} e^{-\frac{(\mu-\mu_0)^2}{2\varsigma^2}}.$$

Importantly, f^{-1} only applies for non-zero firing rates as the input to firing rates mapping is not injective. For $\nu = 0$, any input value such that $\mu < b$ is valid. This leads to the two-part definition of $\mathcal{P}_\nu$.

$$\begin{aligned}\mathcal{P}_\nu d\nu =& \mathcal{P}_\mu(f^{-1}(\nu))\, \frac{df^{-1}}{d\nu}\, d\mu, \quad \nu > 0\\ P(\nu = 0) =& \int_{-\infty}^{b} \mathcal{P}_\mu d\mu = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{b-\mu_0}{\sigma\sqrt{2}}\right)\right).\end{aligned}$$

Which yields [132, 149]

$$\mathcal{P}_\nu = \frac{1}{n a^{\frac{1}{n}} \nu^{\frac{n-1}{n}} \varsigma\sqrt{2\pi}} \exp\left(\frac{-\left(b + \left(\frac{\nu}{a}\right)^{\frac{1}{n}} - \mu_0\right)^2}{2\varsigma^2}\right) \mathcal{H}(\nu) + P(\nu = 0)\, \delta(\nu), \tag{4.1}$$

Where $\mathcal{H}$ is the Heaviside function and δ is the Dirac delta function.

4.1.2. Population average firing rate

If the input to each neuron is independently drawn from the distribution $\mathcal{P}_\mu$, the distribution of the firing rates in the population will follow $\mathcal{P}_\nu$ (Eq. 4.1) and the average firing rate in the

inhomogeneous population will be the expected value of the distribution

$$\begin{aligned}\langle\nu\rangle =& E[\mathcal{P}_\nu] \\ =& \int_0^\infty \nu \frac{1}{n a^{\frac{1}{n}} \nu^{\frac{n-1}{n}} \varsigma\sqrt{2\pi}} e^{\frac{-\left(b+\left(\frac{\nu}{a}\right)^{\frac{1}{n}}-\mu_0\right)^2}{2\varsigma^2}} d\nu \\ =& \int_0^\infty \frac{a z^n}{\varsigma\sqrt{2\pi}} e^{\frac{-(z+b-\mu_0)^2}{2\varsigma^2}} d\nu \\ \equiv& \Psi_n(\mu_0, \varsigma).\end{aligned} \tag{4.2}$$

Using an analytical solver (Wolfram Mathematica), we get the expression Ψ_n for the average firing rate of a population receiving Gaussian-distributed input.

$$\begin{aligned}\Psi_n(\mu_0, \varsigma) = \frac{2^{\frac{n-1}{2}} a \varsigma^{n-1} (\mu_0 - b)}{\sqrt{\pi}} \Gamma\left(1 + \frac{n}{2}\right) {}_1F_1\left(\frac{1-n}{2}, \frac{3}{2}, \frac{-(\mu_0 - b)^2}{2\varsigma^2}\right) + \\ \frac{2^{\frac{n}{2}-1} a \varsigma^n}{\sqrt{\pi}} \Gamma\left(\frac{1+n}{2}\right) {}_1F_1\left(\frac{-n}{2}, \frac{1}{2}, \frac{-(\mu_0 - b)^2}{2\varsigma^2}\right),\end{aligned} \tag{4.3}$$

Where Γ is the gamma function and ${}_1F_1$ is the Kummer hypergeometric function [2]. The function Ψ_n plays the role of the transfer function for inhomogeneous power-law-activated populations, similarly to the Φ (Eq. 1.3) and f (Eq. 2.1) activation functions presented in previous sections.

The effect of the distributed input on the average firing rate can be highlighted by comparing the population average $\langle\nu\rangle$ to the median firing rate, which is the input of a neuron receiving the mean input $\nu_0 = f(\mu_0)$. As expected, the effect of the distributed input grows with the standard deviation of the distribution ς (Fig. 4.2.A). Furthermore, the type of effect depends on the mean input μ_0, as the effect of the network inhomogeneity is mainly a relative change in firing rates for low inputs, and an absolute difference in firing rates at high inputs (Fig. 4.2.B-C).

4.1.3. Self-consistency equation

We have shown how inhomogeneous input leads to a shift in the population mean firing rate. Next, we investigate how this effect can be accounted for in the case of a recurrently connected inhomogeneous network. In this section, we list different sources of inhomogeneity and show how they can be accounted for in a self-consistent system of equations to predict the steady state activity levels in a network.

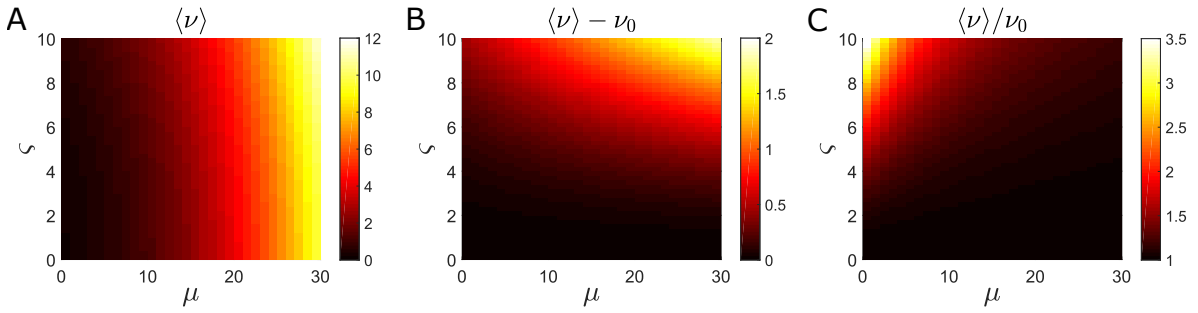


Figure 4.2.: **Effect of the input variability ς on the population average firing rate, for *E* neurons. A:** Heatmap of the population average firing rate $\langle\nu\rangle$, as a function of the population mean input μ and the standard deviation of the input distribution ς. The mean firing rates grow with both μ and ς according to Eq. 4.3. The activation function at the population level maintains a general power law shape across input variability values. **B:** Absolute effect of the input variability. The difference between the mean and median firing rates shows how the mean firing rates are increased due to input inhomogeneity. The difference is highest for high mean input μ. **C:** Relative effect of the input variability. The relative increase in average firing rate caused by ς is highest for low inputs where the firing rate for homogeneous input ν_0 is near 0.

► Inhomogeneous Feedforward input

With a heterogeneous feedforward input, the firing rates in the network will be heterogeneous, which will in turn make the recurrent inputs heterogeneous as well, as different neurons in the network will receive more or less recurrent input depending on the activity of the presynaptic neurons projecting onto them.

Still assuming the distribution of input to neurons is Gaussian, the variance of the firing rate distribution is given by

$$\begin{aligned}\mathrm{Var}[\nu] &= \int_0^\infty (\nu - \langle\nu\rangle)^2 \mathcal{P}_\nu d\nu \\ &= \int_0^\infty \nu^2 \mathcal{P}_\nu d\nu - \langle\nu\rangle^2 \\ &= a\int_0^\infty \frac{az^{2n}}{\varsigma\sqrt{2\pi}} e^{\frac{-(z+b-\mu_0)^2}{2\varsigma^2}} d\nu - \langle\nu\rangle^2 .\end{aligned}$$

The integral in the first term has the same form as in Eq. 4.2 and can be replaced by $\Psi_{2n}(\mu_0, \varsigma)$ (Eq. 4.3). From here, the variability in the recurrent connections can be quantified

$$\begin{aligned}\mathrm{Var}[j\sum_C \nu] &= j^2 C\ \mathrm{Var}[\nu] \\ &= \frac{J^2}{C}\ \mathrm{Var}[\nu],\end{aligned}$$

where j is the strength of a single synapse, C is the number of incoming synaptic connections and J is the effective connection strength at the population level. For a given effective connection strength, the variance decreases with the number of connections as the average firing rate perceived by a neuron through its synapses will converge to $\langle\nu\rangle$. Since the variability in firing rates caused by the feedforward input is itself amplified through the recurrent connections, it must be included in a self-consistent system of equations to follow its effect:

$$\begin{cases} \langle\nu_E\rangle & = \Psi_{En}(\mu_E, \varsigma_E) \\ \mathrm{Var}[\nu_E] & = a_E\Psi_{E2n}(\mu_E, \varsigma_E) - (\Psi_{En}(\mu_E, \varsigma_E))^2 \\ \mu_E & = J_{EE}\langle\nu_E\rangle - J_{EI}\langle\nu_I\rangle + \mu_{ext0} \\ \varsigma_E^2 & = \frac{J_{EE}^2}{C_{EE}}\mathrm{Var}[\nu_E] + \frac{J_{EI}^2}{C_{EI}}\mathrm{Var}[\nu_I] + \varsigma_{extE}^2 \\ \langle\nu_I\rangle & = \Psi_{In}(\mu_I, \varsigma_I) \\ & \ldots \end{cases} \tag{4.4}$$

where Ψ_{Xn} is given by Eq. 4.3, and Ψ_{X2n} has the same form with the power law parameter n_X doubled.

The combined effect of the nonlinear dependence of Ψ on the input variability (Fig. 4.2) and the amplification of this inhomogeneity through recurrent inputs leads to a progressively stronger effect as the feedforward input variability ς_{ext} grows (Fig. A.6).

▶ Inhomogeneous activation function

If the source of heterogeneity originates from the variability in the neuronal properties, it could result in variability in LIF parameters such as the membrane time constant τ (Eq. 1.1) or the input noise σ (see section 2.3.2). Variability on these parameters would lead to variability in the activation function, which would in turn lead to variability in the three power law parameters a, b, and n (see Fig. 2.2.B). While such effects cannot be incorporated into an analytically tractable self-consistent system of equations, an approximation can be made to achieve this easily. If the variability in the neuronal excitability ($\mathrm{Var}[\tau], \mathrm{Var}[\sigma]$) is small, its effect on the activation function can be translated into changes of only one of the three power law parameters. That is, variability in the neuronal activation function within the population is well approximated by

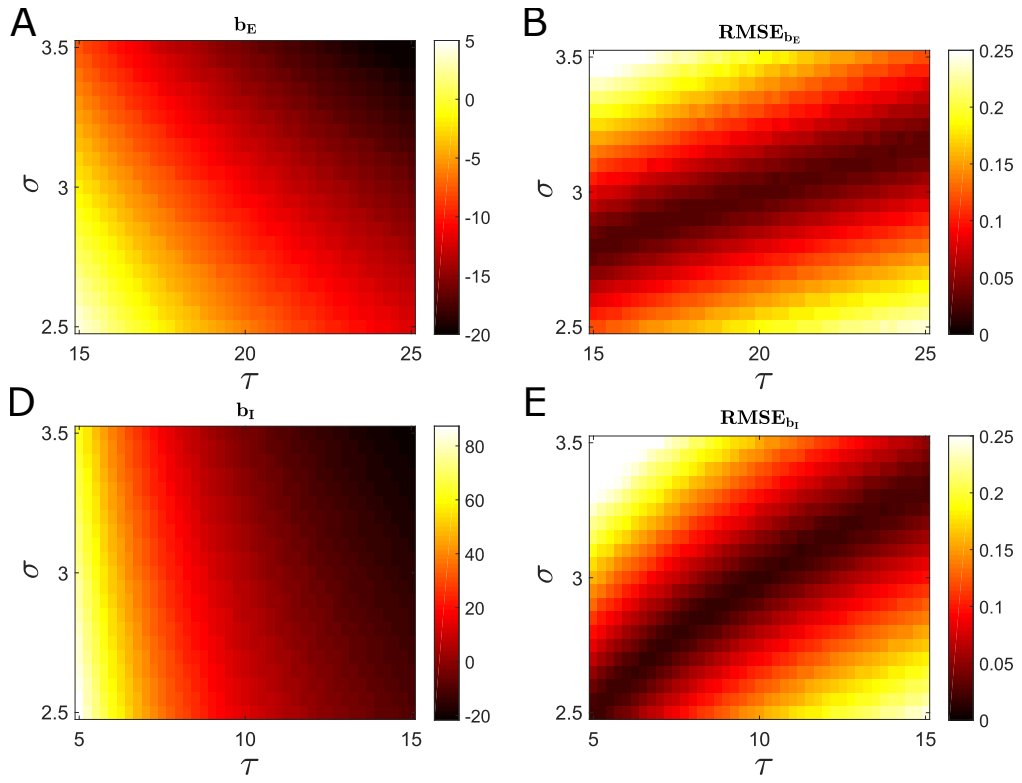


Figure 4.3.: **Small changes in the neuronal properties can be accurately reflected in changes of a single power law parameter: the threshold b. A:** starting from the fitted power law parameters corresponding to the *E* population (Table 2.5, τ_E=20 ms σ_{ext}=3 mV/$\sqrt{\text{s}}$), the power law parameter b can be adjusted to accommodate for deviations in τ or σ. Effectively, the parameters a and n are kept constant and changes in b are equivalent to a horizontal shift of the activation function. **B:** The goodness of fit, as measured by the Root Mean Square Error (RMSE), corresponding to the shifted power law from panel A. The shifted power law is a good approximation of the FI curve if the deviations in τ and σ are small. Interestingly, if variations in σ and τ are positively correlated, neurons can remain in the diagonal domain where the fit quality is maintained and will not cause inaccuracy in the self-consistent rate model. **C:** Same as Panel A, but in the case of the *I* neurons (τ_I=10 ms). **D:** The goodness of fit associated with Panel D shows a similar trend as for *E* neurons.

tuning one of the power law parameters while the other two remain fixed. Incidentally, the one parameter that can best adjust the activation function is the firing threshold input value b (see Fig. 4.3).

Therefore, a population within which the excitability is variable can be modeled as a population in which the power law parameter b is variable. Since this parameter plays the role of a virtual input, the self-consistent system has the same form as Eq. 4.4. In this case, the variability in b caused by the variability in the neuronal properties would be included with the external input variability ς_{ext}.

► Inhomogeneous recurrent connection strengths

The heterogeneity in the network can be caused by heterogeneity in the synaptic connections. In that case, the synapses between two populations are not all identical. Here again, different neurons will receive varying recurrent inputs depending on the strength of the incoming synapses. This will result in a non-uniform firing rate distribution within the population, further contributing to the variability in recurrent input. The recurrent input received by a neuron at the steady state is given by the sum over all its synapses of the product of the synaptic strength with the firing rate of the presynaptic neuron. The firing rate of a postsynaptic neuron is correlated with the firing rate of its presynaptic partners and the strength of its incoming synapses. It is however independent of the strength of its own outgoing synapses. Since j_i and ν_i are independent, we can get:

$$\begin{aligned} \mathrm{Var}[\sum_{C} j_i \nu_i] &= C\ \mathrm{Var}[j\nu] \\ &= C\left(j_0^2\ \mathrm{Var}[\nu] + \langle\nu\rangle^2 \varsigma_j^2 + \mathrm{Var}[\nu]\ \varsigma_j^2\right) \\ &= \frac{J_0^2}{C}\ \mathrm{Var}[\nu] + C\varsigma_j^2\left(\ \langle\nu\rangle^2 + \mathrm{Var}[\nu]\ \right), \end{aligned} \tag{4.5}$$

where j_0 is the average synaptic strength and ς_j^2 is the variance of the synaptic strength distribution in the synaptic population. The inclusion of a distribution of synaptic weights can be supported by the self-consistent inhomogeneous SSN system (Eq. 4.4) by including two additional terms in the input variability ς_X.

► Inhomogeneous number of synaptic connections

Until here, we considered that each neuron in population X receives exactly the same number of connections C from a population of synapses. That is, all neurons have the same in-degree, but the out-degree can be randomly distributed as it does not condition the firing rate of the neuron. A less restrictive approach would be the case of the Erdös-Rényi network, where all pairs of nodes are connected -or not- with the same probability independently of one another [51]. The in-degree of neurons in such a network would not be uniform across the population but would instead follow a binomial distribution. For a presynaptic population of size N with a

connection probability p, the number of connections C follow

$$\begin{aligned} P(C=k) =& \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} \\ E[C] =& pN \\ \mathrm{Var}[C] =& Np\,(1-p)\,. \end{aligned} \tag{4.6}$$

The heterogeneity in the in-degree of neurons will cause a heterogeneity in the recurrent input, similar to the case where the connection strength is non-uniform. The variance in the recurrent input can be obtained using the law of total variance:

$$\begin{aligned} \mathrm{Var}\left[\sum_C j_i \nu_i\right] =& E\left[\mathrm{Var}\left[\sum_C j_i \nu_i \,\middle|\, C\right]\right] + \mathrm{Var}\left[E\left[\sum_C j_i \nu_i \,\middle|\, C\right]\right] \\ =& E\left[C\right] \mathrm{Var}\left[j\nu\right] + E[j\nu]^2 \,\mathrm{Var}\left[C\right]. \end{aligned}$$

Using Eq. 4.5 and Eq. 4.6, we get

$$\mathrm{Var}[\sum_C j_i \nu_i] = Np\left(j_0^2 \mathrm{Var}\left[\nu\right] + \varsigma_j^2 \left\langle \nu \right\rangle^2 + \mathrm{Var}\left[\nu\right] \varsigma_j^2 + (1-p) j_0^2 \left\langle \nu \right\rangle^2\right), \tag{4.7}$$

which can be inserted in the self-consistent system (Eq. 4.4).

The effect of variability associated with an Erdös-Rényi connection scheme predicted by the rate model accurately matches the results of LIF simulations. Nonetheless, the difference with the homogeneous case is minute (Fig. 4.4). Although the effect size of this variability will vary with the size and connectivity of networks, this result suggests that the analysis of the homogeneous case can usually suffice to study the activity regimes of neural networks.

4.1.4. Mathematical tractability

Through the characterization of the mean input and firing rates, the quenched variability in input and firing rates, and how these quantities interact with one another, we can predict the state of the network using a self-consistent system of equations derived from the SSN (Eq. 4.4). However, unlike in the case of homogeneous networks where the SSN has the benefit of being tractable and easy to manipulate thanks to its simple mathematical formulation, the expansion to inhomogeneous networks loses this quality. This added mathematical complexity is due to the fact that the state of each population is characterized by two instead of one variable, increasing

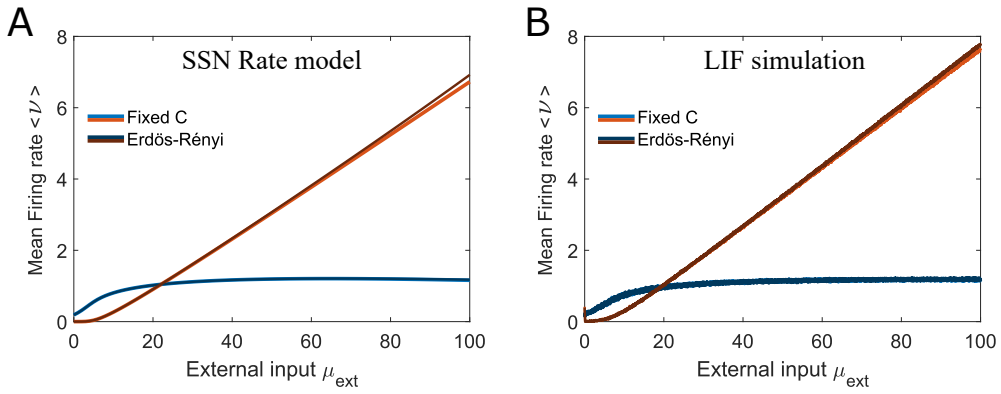


Figure 4.4.: **The variability caused by a Erdös-Rényi connectivity scheme leads to negligible effects. A:** The SSN firing rate profile corresponding to Fig. 2.3 is compared between the homogeneous case where all neurons receive the same fixed number of synaptic connections (bright color), to the Erdös-Rényi situation where the in-degree of neurons is binomially distributed (dark color). Using the self-consistent rate model (Eq. 4.4 with Eq. 4.7), the difference between the two scenarios is imperceptible for the *E* profile, and is in the order of $0.2\,\mathrm{Hz}$ for *I* at high input. **B:** Same comparison as in Panel A, performed in LIF spiking simulation. The same observation can be made, as the *E* profile is unchanged and the *I* rates increase only marginally. The rate-model prediction is confirmed in the spiking simulation and the quenched variability due to the connectivity scheme is negligible as its effect is smaller than the difference between the spiking network and its SSN approximation.

the dimensionality of the analysis. Furthermore, the activation function Ψ (Eq. 4.3) cannot be manipulated as easily as the power law function f of the homogeneous case (Eq. 2.1). However, the model can be simplified using the same strategy we used to derive the SSN (Eq. 2.2) from the self-consistent Ricciardi equations (Eq. 1.4), by approximating Ψ with a power law with 3 parameters, similarly to Eq. 2.1.

$$\Psi \approx a_\Psi \left(\langle\mu\rangle - b_\Psi\right)_+^{n_\Psi}, \tag{4.8}$$

where the three power law parameters are fitted based on Eq. 4.3. Since the input variability will itself depend on the state of the network, the dependency of Eq. 4.8 on ς must be made explicit. Fortunately, the power law can again save the day as the dependence of a_Ψ, b_Ψ and n_Ψ on ς can be well approximated with a power law function. That is, the deviation from the SSN power law can itself be modeled with a power law. For the average firing rate $\langle\nu_X\rangle$ of the *E* and

I populations, we have:

$$\begin{aligned}
\ln\left(a_{\Psi E}\right) &= \ln\left(a_E\right) - \left(\frac{\varsigma}{8.815}\right)^{1.843} \\
b_{\Psi E} &= b_E - \left(\frac{\varsigma}{3.367}\right)^{1.938} \\
n_{\Psi E} &= n_E + \left(\frac{\varsigma}{23.2}\right)^{1.775} \\
\ln\left(a_{\Psi I}\right) &= \ln\left(a_I\right) - \left(\frac{\varsigma}{5.456}\right)^{1.967} \\
b_{\Psi I} &= b_I - \left(\frac{\varsigma}{2.808}\right)^{2.074} \\
n_{\Psi I} &= n_I + \left(\frac{\varsigma}{13.05}\right)^{1.845}.
\end{aligned}$$

The same approximation can be applied for the function Ψ_{2n} used to quantify the firing rate variability $\mathrm{Var}[\nu_X]$:

$$\begin{aligned}
\ln\left(a_{\Psi E^2}\right) &= \ln\left(a_E\right) - \left(\frac{\varsigma}{4.591}\right)^{1.930} \\
b_{\Psi E^2} &= b_E - \left(\frac{\varsigma}{2.528}\right)^{2.031} \\
n_{\Psi E^2} &= 2n_E + \left(\frac{\varsigma}{11.44}\right)^{1.830} \\
\ln\left(a_{\Psi I^2}\right) &= \ln\left(a_I\right) - \left(\frac{\varsigma}{3.47}\right)^{1.943} \\
b_{\Psi I^2} &= b_I - \left(\frac{\varsigma}{2.398}\right)^{2.107} \\
n_{\Psi I^2} &= 2n_I + \left(\frac{\varsigma}{8.143}\right)^{1.785}.
\end{aligned}$$

Using this approximation, the mathematical formulation of the self-consistent system of equations can be simplified to facilitate both the numerical resolution and the analytical study. Inevitably, this simplification comes at the cost of a reduced accuracy for high values of ς.

4.1.5. Limitations

► **Range of validity of the activation function**

The population-average firing rate is derived from the expected value over the pdf $\mathcal{P}_\nu$. Because of this, the accuracy of the prediction will be conditioned by the quality of the power law approximation over a larger domain than the range of considered average firing rates $\langle\nu\rangle$. For instance, if the mean firing rate of the population is at $10\,\mathrm{Hz}$ with a variance $\mathrm{Var}[\nu]$ of $5\,\mathrm{Hz}$,

47% of neurons will fire at a rate higher than 10 Hz. The accuracy of the power law approximation beyond 10 Hz would come into play in the determination of the mean activity. Differences between the power law and LIF activation function at high rates can then result in a misestimation of the population-average firing rates. In other words, the range of validity of the power law activation function applies to all neurons in the population and not only to their average.

► **Recurrent temporal noise**

In the case where the synaptic connection strength is non-uniform, the effect of the recurrent input on the temporal noise σ becomes significant if ς_j is high. If unaccounted for, this effect can lead to significant discrepancies between the spiking network simulation and its rate model prediction (Fig. 4.5). The temporal noise is given by

$$\sigma^2 = \sigma_{ext}^2 + \sum_{C_E} j_E^2 \nu_E - \sum_{C_I} j_I^2 \nu_I$$
$$E[\sigma^2] =< \sigma_{ext}^2 > + N_E p_E \left(j_{E0}^2 + \varsigma_{jE}^2\right) \langle \nu_E \rangle + N_I p_I \left(j_{I0}^2 + \varsigma_{jI}^2\right) \langle \nu_I \rangle.$$

In cases where high variability in recurrent weights is considered, the contribution of recurrent noise can no longer be neglected and the power law fit must be appropriately adjusted to account for it. Besides the shift in recurrent temporal noise, the inclusion of a high variance in the recurrent connection strengths will affect the statistical properties of the input which will further deviate from the assumed white noise. As the heterogeneity in the synaptic weights ς_j increases, the input dynamics will be dominated by rare events and neuronal spiking will be dictated by few exceptionally strong synapses. As ς increases, the white noise assumption underlying the diffusion approximation and the Ricciardi Φ (Eq. 1.3) activation function of LIF neurons becomes invalid.

Furthermore, the effect of the network inhomogeneity on the recurrent noise will affect different neurons differently, leading to further heterogeneity in the firing properties within a population. The variance of the recurrent noise is given by

$$\mathrm{Var}[\sum_C j^2 \nu] = E[C] \left(\left(j_0^2 + \varsigma_j^2\right)^2 \left(2\langle\nu\rangle^2 + 3\mathrm{Var}[\nu]\right) - 2j_0^4 \left(\langle\nu\rangle^2 + \mathrm{Var}[\nu]\right) \right)$$
$$+ \mathrm{Var}[C] \langle\nu\rangle^2 \left(j_0^2 + \varsigma_j^2\right)^2 .$$

The variability in the recurrent noise within a population leads to heterogeneity in the activa-

tion functions of the population. This effect can be included in the self-consistent system of equations as a heterogeneous external input (subsubsection 4.1.3).

► **Effect size**

Ultimately, we find that the effect of inhomogeneity within neural populations is imperceptible or negligible unless we consider high variability. The effect of variability in the synaptic connection strength only becomes significant when its standard deviation is of the same order as the mean ($\frac{s_J}{J_0} \approx 1$). With such high variability, 16% of the synapses would break Dale's law, which states that biological neurons cannot have simultaneously excitatory and inhibitory synapses [48, 44]. Similarly, the effect of a fully random connectivity scheme, as in the case of the Erdös-Rényi network, is negligible (Fig. 4.4). From this, it seems that biologically realistic degrees of variability within neural populations are generally inconsequential and do not pose a significant limitation to the generalization of the results obtained under the assumption of a homogeneous network in the previous chapters. Nonetheless, if a network possesses a strong source of variability or if multiple concurrent mechanisms collectively lead to a consequent level of heterogeneity, such effects can be studied in a closed-form rate model based on the SSN.

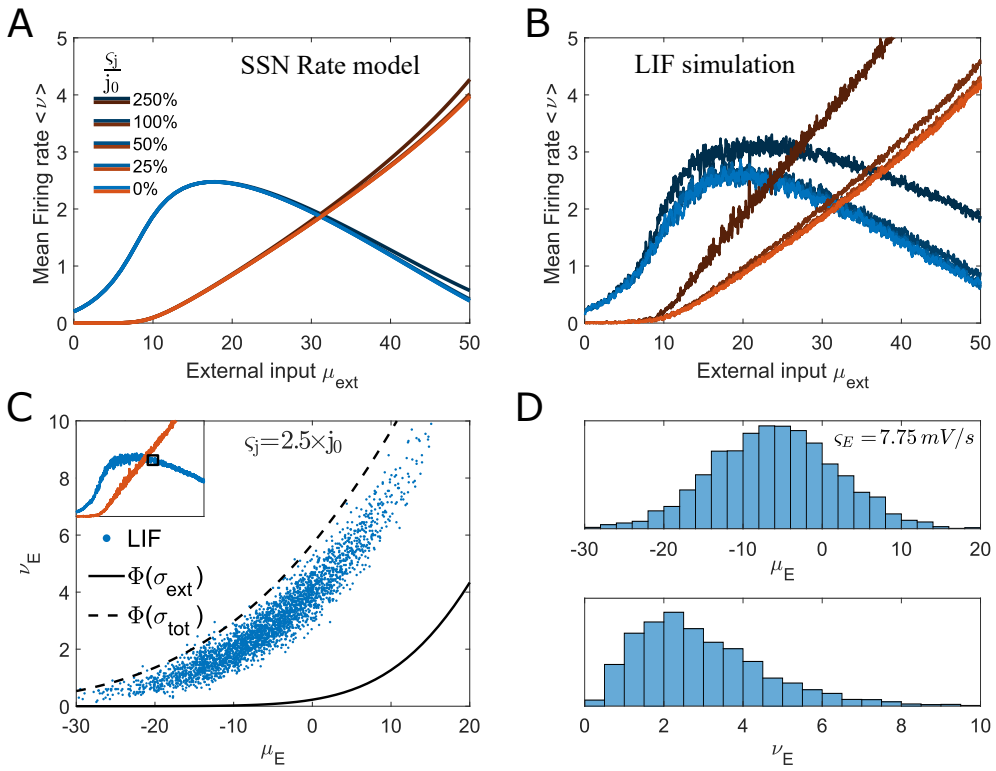


Figure 4.5.: **For high variability in the recurrent connection strength, the SSN rate model fails to predict the activity of spiking simulations. A:** Prediction of the rate model for increasing levels of variability in the connection strengths (from light to dark colors, increasing variability.). The effect of the inhomogeneity is imperceptible until $\varsigma_j > j_0$. **B:** In the spiking simulation, the effect of the variability is much stronger and deviates significantly from the rate model prediction beyond $\varsigma_j > 0.5 \times j_0$. The parameters are the same as in Panel A. **C:** The firing rates and inputs of the *E* neurons are recorded at a fixed point of simulation where the rates mismatch the prediction (see inset). The observed FI curve of the LIF neurons follows a power law that strongly deviates from the expected FI curve (solid line) assumed in the SSN. The deviation can be partially explained by the contribution of the recurrent input to the temporal noise (dashed line) but the rates are then overestimated. **D:** The total input of *E* neurons recorded in C follows a Gaussian distribution with a large standard deviation (Top). This results in a skewed distribution of firing rates with high variance (bottom).

4.2. Structured networks

We previously analyzed unstructured inhomogeneous networks where the difference between neurons within a population cannot be explained by underlying variables. Because of this, the properties of individual neurons are treated as random variables and the mean-field theory is built on a statistical analysis of the population properties. In the case of structured networks, the neuronal populations are organized in a feature space, and the features of a neuron determine its feedforward input and recurrent connectivity. This organization can be topographic and directly translate to the spatial organization of circuits, as is the case in numerous brain regions such as the somatotopic map of the somatosensory cortex [68], the tonotopic map in the auditory cortex [153] or the retinotopic map of the visual system [192]. Alternatively, in situations where the cortical map is not spatially organized, such as the salt-and-pepper organization of the orientation tuning map in mouse visual cortex [83, 88], the feature space is defined by the receptive field of neurons instead of their topographic location. In that framework, neurons that are nearby in feature space receive feedforward input of similar amplitude since the neurons are selective to similar stimuli. Furthermore, the connection probability between two neurons can also depend on their distance in feature space as neurons with similar receptive fields have been shown to preferentially connect together [94] (see Fig. 4.6).

Since the total input of neurons depends on their position in feature space θ, the firing rate profile of the population varies along the feature dimensions of that space. In a 1-D feature space, given the feedforward input as a function of position $M_{ext}(\theta)$ and the pair-wise connection probability as a function of the distance between neurons in feature space $P(\theta - \theta')$, the input to a neuron in population X at position θ is given by

$$\mu_X(\theta) = M_{ext}(\theta) + \int_{\theta_{min}}^{\theta_{max}} \left(j_{XE}\eta_E P_{XE}(\theta - \theta')\,\nu_E(\theta') - j_{XI}\eta_I P_{XI}(\theta - \theta')\,\nu_I(\theta') \right) d\theta'.$$

Where η is the linear density of neurons along the dimension θ. The geometric properties of the feature space depend on the nature of the considered feature: Periodic boundary conditions in the case of orientation preference, infinite for tonotopy, etc. The mean-field analysis of neural populations organized in feature space is conceptually equivalent to the analysis of an infinite system of populations, where each population is selective for an infinitesimal section of feature space [150, 147]. A local network comprises all neurons whose preferred feature is in the range

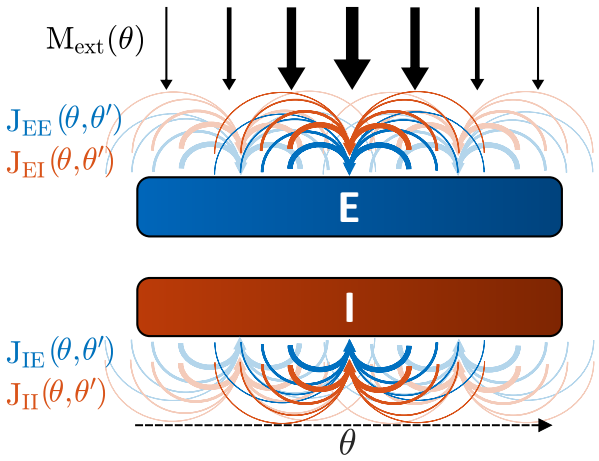


Figure 4.6.: **Schematic diagram of a structured E-I network in feature space θ.** In structured networks, the feedforward input of each neuron is determined by its receptive field, which translates into its location in the feature space θ. The recurrent connection probability is also a function of the position in feature space of the pre- and postsynaptic neurons.

$[\theta,\ \theta + \Delta\theta]$ where the neural properties can be considered homogeneous, and the mean-field assumption considers that the number of neurons is large enough to ignore finite-size effects and focus on collective effects such as the population mean firing rate.

Using the activation function of the neuronal populations, we obtain the spatially structured rate model

$$\begin{cases} \nu_E(\theta) = & f_E\Big(\left(J_{EE} * \nu_E\right)(\theta) - \left(J_{EI} * \nu_I\right)(\theta) + M_{ext}(\theta) \Big) \\ \nu_I(\theta) = & f_I\Big(\left(J_{IE} * \nu_E\right)(\theta) - \left(J_{II} * \nu_I\right)(\theta) + rM_{ext}(\theta) \Big), \end{cases} \tag{4.9}$$

Where $*$ is the convolution operation, and the spatial recurrent connection functions J_{XY} are given by $j_{XY}\eta_Y P_{XY}(\Delta\theta)$.

4.2.1. SSN implementation in structured networks

Using the LIF-adjusted SSN activation function, the firing rate profile in feature space is given by

$$\begin{cases} \nu_E(\theta) = & a_E \left(J_{EE} * \nu_E - J_{EI} * \nu_I + M_{ext} - b_E\right)_+^{n_E} \\ \nu_I(\theta) = & a_I \left(J_{IE} * \nu_E - J_{II} * \nu_I + rM_{ext} - b_I\right)_+^{n_I}. \end{cases} \tag{4.10}$$

The firing rate profiles of the two populations are solutions to the coupled Fredholm integral equation [10]. The power law transformation makes the equation especially challenging to solve analytically. In order to approach a solution numerically, a weighted residual method can be used. The connection probability is assumed to be a Gaussian function of feature distance between the pair of neurons, of the form

$$P_{XY}(\theta,\theta') = p_{XY}\, e^{\frac{-(\theta-\theta')^2}{2\lambda_{XY}^2}}. \tag{4.11}$$

Here, p is the peak probability, corresponding to the connection probability of neurons with the same feature preference, and λ_{XY} is the lengthscale of the connections in feature space. Similarly, we assume the external input is a Gaussian function of feature, centered at the stimulus feature θ_0.

$$M_{ext}(\theta) = \mu_{ext}\, e^{\frac{-(\theta-\theta_0)^2}{2\lambda_{ext}^2}}, \tag{4.12}$$

where λ_{ext} is a measure of the selectivity in the receptive field of neurons. Finally, both firing rate profiles are approached as a weighted sum of Gaussian curves of various lengthscales centered around θ_0,

$$\nu = \sum_i K_i e^{-\frac{(\theta-\theta_0)^2}{2\lambda_i^2}}.$$

The recurrent convolution becomes

$$\begin{aligned} J_{XY} * \nu_Y(\theta) =& j_{XY} p_{XY} \eta_Y \sum_i K_{Yi} \int_{-\infty}^{\infty} e^{-\frac{(\theta-\theta')^2}{2\lambda_{XY}^2}} e^{-\frac{(\theta'-\theta 0)^2}{2\lambda_i^2}} d\theta' \\ =& j_{XY} p_{XY} \eta_Y \sum_i K_{Yi} \sqrt{\frac{2\pi\lambda_{XY}^2\lambda_i^2}{\lambda_{XY}^2+\lambda_i^2}} \quad e^{-\frac{(\theta-\theta_0)^2}{2(\lambda_{XY}^2+\lambda_i^2)}}. \end{aligned}$$

Using this decomposition of the firing rate profile, the spatially structured SSN equation (Eq. 4.10) can be solved through optimization of the coefficients K_{Ei}, K_{Ii}.

While this approach offers a numerical framework to solve the firing profiles of spatially structured neural networks (Fig. 4.7), it remains intractable and unsuited for analytical manipulation. Despite the simple mathematical formulation of the SSN, the inclusion of its power law in the integral equation prevents an analytical resolution. In order to achieve an analytical resolution of spatially distributed networks, we must rely on even simpler rate models, such as the Balanced State model.

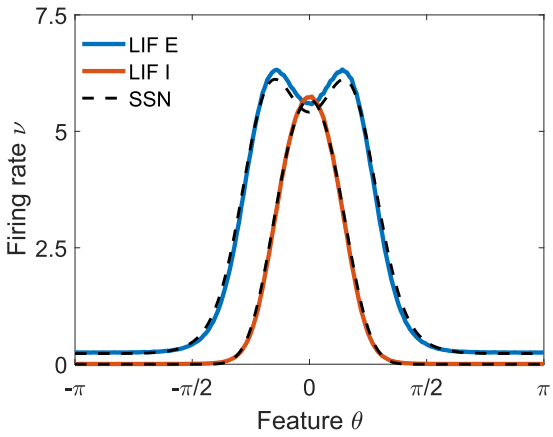


Figure 4.7.: **Firing rate profile in a structured E-I network with periodic boundaries.** The prediction of the SSN accurately matches the results of the spiking network simulation. The E profile has a minimum in 0, where the external input is strongest because this network supports supersaturation ($r \int J_{EI} > \int J_{II}$).

4.2.2. Balanced State analysis of structured networks

This subsection is adapted from elements in the methods and results section of the article Bernáez Timón Laura, Ekelmans Pierre, Konrad Sara, Nold Andreas and Tchumatchenko Tatjana. "Synaptic plasticity controls the emergence of population-wide invariant representations in balanced network models". *Physical Review Research* 4.1 (2022): 013162. https://doi.org/10.1103/PhysRevResearch.4.013162

Unlike other rate models which rely on the activation function to determine the firing rate of neurons (Eq. 4.9), the Balanced State prediction stems from the assumption that excitatory and inhibitory inputs cancel out ($\mu_E, \mu_I = \mathcal{O}(\frac{1}{\sqrt{N}}) \quad \text{as } N \to \infty$) [186]. The profile equation becomes

$$\begin{cases} (J_{EE} * \nu_E)(\theta) - (J_{EI} * \nu_I)(\theta) &= -M_{ext}(\theta) \\ (J_{IE} * \nu_E)(\theta) - (J_{II} * \nu_I)(\theta) &= -rM_{ext}(\theta)\,. \end{cases} \tag{4.13}$$

The convolution can be converted into a simple multiplication by transforming the equation

in the Fourier space [147]

$$\begin{cases} \hat{J}_{EE}\hat{\nu}_E(\omega) - \hat{J}_{EI}\hat{\nu}_I(\omega) &= -\hat{M}_{ext}(\omega) \\ \hat{J}_{IE}\hat{\nu}_E(\omega) - \hat{J}_{II}\hat{\nu}_I(\omega) &= -r\hat{M}_{ext}(\omega), \end{cases} \tag{4.14}$$

where ω is the frequency decomposition of the feature space θ. In order for the network to be perfectly balanced, the balance condition must be fulfilled at all frequencies ω. The Fourier coefficients of $\hat{\nu}_E$ and $\hat{\nu}_I$ at all frequencies are then obtained from inverting the weight matrix in frequency domain (Eq. 4.14).

For Gaussian functions J_{XY} and M_{ext} in feature space (Eq. 4.11, Eq. 4.12), we get

$$\begin{aligned} \hat{J}_{XY}(\omega) &= j_{XY}p_{XY}\eta_Y\lambda_{XY}\sqrt{2\pi}e^{-\frac{1}{2}\omega^2\lambda_{XY}^2} \\ \hat{M}_{ext}(\omega) &= \mu_{ext}\sqrt{2\pi}e^{-\frac{1}{2}\omega^2\lambda_{ext}^2}. \end{aligned}$$

We assume that the lengthscale of recurrent connections is characterized by the projection distance of the presynaptic neuron ($\lambda_{EX} = \lambda_{IX} \equiv \lambda_X$) and get

$$\begin{aligned} \hat{\nu}_E(\omega) &= \frac{rj_{EI}p_{EI} - j_{II}p_{II}}{j_{EE}p_{EE}j_{II}p_{II} - j_{EI}p_{EI}j_{IE}p_{IE}} \frac{\lambda_{ext}\mu_{ext}}{\lambda_E\eta_E} e^{-\frac{1}{2}\omega^2\left(\lambda_{ext}^2 - \lambda_E^2\right)} \\ \hat{\nu}_I(\omega) &= \frac{rj_{EE}p_{EE} - j_{IE}p_{IE}}{j_{EE}p_{EE}j_{II}p_{II} - j_{EI}p_{EI}j_{IE}p_{IE}} \frac{\lambda_{ext}\mu_{ext}}{\lambda_I\eta_I} e^{-\frac{1}{2}\omega^2\left(\lambda_{ext}^2 - \lambda_I^2\right)}. \end{aligned}$$

Or in feature space

$$\begin{aligned} \nu_E(\theta) &= \left(\frac{rj_{EI}p_{EI} - j_{II}p_{II}}{j_{EE}p_{EE}j_{II}p_{II} - j_{EI}p_{EI}j_{IE}p_{IE}}\right) \frac{\lambda_{ext}\mu_{ext}}{\lambda_E\eta_E\sqrt{2\pi\left(\lambda_{ext}^2 - \lambda_E^2\right)}} e^{-\frac{\theta^2}{2\left(\lambda_{ext}^2 - \lambda_E^2\right)}} \\ \nu_I(\theta) &= \left(\frac{rj_{EE}p_{EE} - j_{IE}p_{IE}}{j_{EE}p_{EE}j_{II}p_{II} - j_{EI}p_{EI}j_{IE}p_{IE}}\right) \frac{\lambda_{ext}\mu_{ext}}{\lambda_I\eta_I\sqrt{2\pi\left(\lambda_{ext}^2 - \lambda_I^2\right)}} e^{-\frac{\theta^2}{2\left(\lambda_{ext}^2 - \lambda_I^2\right)}}. \end{aligned} \tag{4.15}$$

From Eq. 4.15, it appears that besides the usual conditions for existence and stability of the Balanced State (Eq. A.4), the spatially structured balanced network imposes $\lambda_{ext} > \max(\lambda_E, \lambda_I)$ [147]. The intuitive interpretation of this condition is that the recurrent input profile must exactly compensate the externally imposed feedforward input profile for balance to be achieved. The recurrent connections therefore need to be finer to prevent any overcorrection.

Even in the context of the Balanced State, simplifying assumptions are required to obtain an

explicit expression for the firing rate profiles. If the condition ($\lambda_{EX} = \lambda_{IX}$, $X = E, I$) is not fulfilled, the Gaussian kernels cannot be factorized and the inverse Fourier transform cannot be defined in general.

Remarkably, the shape of the *E* firing profile is independent of the *I* connectivity lengthscale, and vice versa. This again stems from the E-I balance condition. The recurrent input profile must exactly cancel the feedforward input profile, which is achieved if both the recurrent *E* and the recurrent *I* inputs have the same shape as the feedforward profile. Since the recurrent *E* input profile is a convolution of the ν_E profile with a kernel of lengthscale λ_E, it is independent of λ_I.

4.2.3. Effect of synaptic plasticity on the firing profile shape

In previous sections, we considered that the connection strength of individual synapses was fixed. Here we consider how the network firing rate profile is transformed through the effect of synaptic plasticity, where the strength of individual recurrent synaptic connections is modulated by neuronal activity.

We consider a short-term plasticity (STP) model derived by Mongillo et al. [118], which is itself an extension of the classic Tsodyks-Markram STP model [179]. The effective synaptic strength results from the interplay of the depression and facilitation mechanisms. Depression is caused by the depletion of neurotransmitters from previously emitted spikes. The neurotransmitter pool replenishes over time but will be low for high presynaptic firing rates. Facilitation is caused by the retention of bound calcium in the presynaptic terminal, which is necessary for the exocytosis of synaptic vesicles. The emission of a presynaptic spike promotes the binding of calcium, which unbinds over time. The fraction of bound calcium will be high for high presynaptic firing rates. The two mechanisms determine the probability of neurotransmitter release w. In the mean-field at steady state, the release probability is a polynomial fraction in the presynaptic firing rate ν [177]

$$w(\nu) = \frac{U^2 D\nu^2 + UD\left(D + F\left(1+U\right)\right)\nu + UFD\left(F+D\right)}{U^2\nu^3 + U\left(F+D\right)\left(1+U\right)\nu^2 + \left(U\left(F+D\right)^2 + FD\right)\nu + FD\left(F+D\right)}, \quad (4.16)$$

where U is the spike-induced calcium increase, F is the calcium unbinding rate and D is the neurotransmitter replenishing rate.

Since the STP factor w operates as a tuning coefficient of the synaptic strength j, the profile

of the recurrent input from population X to population Y is given by

$$\mu_{recXY}(\theta) = J^0_{XY} * w_Y \nu_Y, \tag{4.17}$$

where J^0_{XY} is the recurrent connection kernel for the baseline synaptic strength, corresponding to a release probability of 1. We assume that the plasticity factor is independent of the postsynaptic neuron ($w_E = w_{EE} = w_{IE}$ and $w_I = w_{EI} = w_{II}$). We define

$$\aleph_X(\nu_X) = w_X(\nu_X) \times \nu_X,$$

and we can modify Eq. 4.13 according to Eq. 4.17. The exact same derivation then leads to

$$\begin{aligned}
\aleph_E(\theta) &= \left(\frac{r j_{EI} p_{EI} - j_{II} p_{II}}{j_{EE} p_{EE} j_{II} p_{II} - j_{EI} p_{EI} j_{IE} p_{IE}}\right) \frac{\lambda_{ext}\mu_{ext}}{\lambda_E \varrho_E \sqrt{2\pi\left(\lambda^2_{ext} - \lambda^2_E\right)}} e^{-\frac{\theta^2}{2\left(\lambda^2_{ext} - \lambda^2_E\right)}} \\
\aleph_I(\theta) &= \left(\frac{r j_{EE} p_{EE} - j_{IE} p_{IE}}{j_{EE} p_{EE} j_{II} p_{II} - j_{EI} p_{EI} j_{IE} p_{IE}}\right) \frac{\lambda_{ext}\mu_{ext}}{\lambda_I \eta_I \sqrt{2\pi\left(\lambda^2_{ext} - \lambda^2_I\right)}} e^{-\frac{\theta^2}{2\left(\lambda^2_{ext} - \lambda^2_I\right)}}.
\end{aligned} \tag{4.18}$$

The firing rate profile can be recovered by extracting the ν_X profile that generated the predicted $\aleph_X$ profile using the inverse $\aleph_X$ function $\nu_X = \aleph_X^{-1}(w_X \nu_X)$.

If $\aleph$ is not injective, multiple ν_E profiles are compatible with a tight balanced condition, which can be the basis for bistability at the spatial profile level. In the case of the STP rule (Eq. 4.16), this cannot occur as the function $\aleph$ is monotonic ($\frac{d\aleph}{d\nu} > 0$) and therefore injective. The inclusion of STP in the Balanced State can still lead to bistability as shown in [118] if the STP factors can depend on the postsynaptic neuron type (i.e., $w_{EE} \neq w_{IE}$). But then again, an explicit expression for the firing rate profile cannot be derived in general.

If $\aleph$ is not surjective, $\aleph^{-1}$ will not exist for high values of $w \times \nu$. This means that no firing rate profile can generate the recurrent input profile required for tight input balance. In that case, a balanced limit only exists for low external inputs μ_{ext} such that the $\aleph$ profile remains in the range where it is invertible. This can occur with the STP plasticity rule (Eq. 4.16) since depression imposes an upper bound, $\aleph \leq D$.

The application of the $\aleph^{-1}$ transformation to the balanced profile can lead to distortions of the spatial firing rate profile. Whether the geometric properties of the firing rate profiles vary with the strength of the feedforward stimulation μ_{ext} depends on the function $\aleph^{-1}$. The profile of $w_X \nu_X$ in Eq. 4.18 can be seen as the product of two factors, one which accounts for the

input strength and the exponential factor which accounts for the dependence on feature space ($\aleph(\nu) = \chi_\mu \times \chi_\theta$). The shape of the profile will be preserved across changes in χ_μ if the effect of the input strength is only a multiplicative rescaling of an invariant profile function. That is if the function $\aleph^{-1}$ is scale-invariant:

$$\aleph^{-1}(\chi_\mu \chi_\theta) = \frac{\aleph^{-1}(\chi_\mu)\,\aleph^{-1}(\chi_\theta)}{\aleph^{-1}(1)}. \tag{4.19}$$

Through function composition, we can rewrite Eq. 4.19 to fall back on the linearity condition

$$\ln \aleph^{-1} e^{\ln \chi_\mu + \ln \chi_\theta} = \ln \aleph^{-1} e^{\ln \chi_\mu} + \ln \aleph^{-1} e^{\ln \chi_\theta} - \ln \aleph^{-1} e^{\ln 1}.$$

Since $\ln(\aleph^{-1}(\exp))$ is a linear function of $\ln(x)$, it follows a linear equation of the form

$$\ln \aleph^{-1} e^{\ln x} = k \ln x + x_0.$$

Overall, it results that transformations $\aleph^{-1}$ which can support scale invariance and therefore maintain the spatial profile independently of the feedforward input strength must follow the power law

$$\aleph^{-1}(x) = x_0 x^k. \tag{4.20}$$

This effect can be seen in Fig. 4.8, where the response of the same network can be either distorted with increasing feedforward input (panel A) or scale-invariant (panel B), depending on whether the plasticity rule leads to a homogeneous $\aleph^{-1}$ function. Interestingly, the condition Eq. 4.20 can be worked back to determine that it requires that the plasticity rule is a power law as well

$$w_X = x_0^{-1/k} \nu_X^{1/k-1}.$$

While this is the strict requirement in the mathematical sense, it can be sufficient to impose that $\aleph^{-1}$ is well approximated with a power law to achieve reasonable scale invariance. Therefore, approximate scale invariance can also be achieved with plasticity rules of diverse forms, such as the polynomial fraction of the STP rule (Eq. 4.16), as in Fig. 4.8.B. For scale-invariant plasticity rules, the exponent k determines whether the profile is compressed or stretched. The spatial

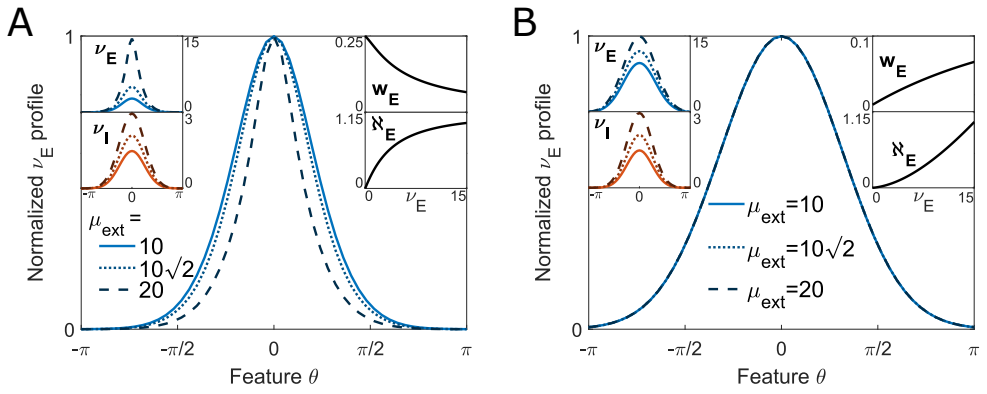


Figure 4.8.: **In balanced network, the width of the activity profile is independent of external input if the synaptic plasticity rule is approximately power law.** In both panels, the connectivity is identical and only E synapses are plastic ($w_I = 1$). The two panels differ in the STP parameters of w_E (Eq. 4.16). **A:** As the external input grows, the width of the firing rate profile shrinks. At the same time, the peak amplitude for both populations grows with μ_{ext} to maintain E-I balance (left insets). The plasticity rule is dominated by depression as w_E decreases with ν_E (top right inset). For high rates, the function $\aleph$ saturates (bottom right inset). Because of the saturating behavior of $\aleph$, which is incompatible with a power law approximation, the underlying firing rate profile does not scale homogeneously with external input and distortions occur. STP parameters: U=0.25, F=2, D=1.15. **B:** Here, the amplitude of the feedforward input only affects the amplitude of the response curve (right inset), but the width of the spatial profile is invariant with input strength. The plasticity rule is dominated by facilitation and w_E grows with ν_E (top right inset). The function $\aleph$ is supralinear and can be well approximated with a power law (bottom right inset). STP parameters: U=0.01, F=1.82, D=50.

profile has the form

$$\aleph_X^{-1}(\chi_\theta) = x_0 e^{-\theta^2 \left(2\frac{\left(\lambda_{ext}^2 - \lambda_X^2\right)}{k}\right)^{-1}},$$

so the lengthscale of the spatial profile is $\frac{\lambda_{ext}^2 - \lambda_X^2}{k}$. The profile is stretched for values of k lower than 1, which correspond to facilitating plasticity, and it is compressed for $k > 1$ or depressing synapses.

The property of scale invariance has been observed in multiple sensory cortical networks and can be the basis for reliable sensory perception [31, 35, 22], as it ensures that the intensity of a sensory stimulus does not interfere with its position in feature space. For instance, the primary visual cortex is sensitive to the contrast of the visual stimulus as well as its orientation [30]. Then, contrast invariance allows for separate encoding of the two stimulus properties, as contrast modulates the peak V1 activity whereas the orientation is encoded in the selectivity of neurons that respond to the stimulus. Without contrast-invariance, the selectivity of the network would change with contrast and could lead to blurry or less accurate perception.

5. Conclusion

5.1. Main results

In this work, I have shown how the SSN rate model can be modified to accurately match the predictions from spiking network simulations while maintaining a simple and tractable mathematical formulation. This rate model is based on the observation that the FI curve of LIF neurons fed with white noise closely follows a power law function in the low firing rate range. Based on the observation that biological neurons mainly operate in this low-rate regime, it follows that the power law approximation is suitable for studying biologically realistic neural networks.

In order to ground the model in biological realism, the power law for each population is based on experimentally reported parameters. Similarly, the network connectivity can be derived from available databases to recreate a realistic model of neural circuitry. Following this, I have shown that the activity regime in layer 2/3 of the mouse primary visual cortex predicted by the SSN rate model matches closely with spiking LIF simulations. Notably, both frameworks capture the nonlinear relation between the firing rate of this network and its incoming feedforward input.

Because of the intrinsically supralinear response of LIF neurons, reflected in the built-in power law of the SSN, simple recurrent E-I networks are capable of various nonlinear input-output transformations. Thanks to the mathematical tractability of the SSN, it can be used to predict the occurrence of these regimes. In particular, they can be precisely mapped in parameter space using exact mathematical conditions. I have illustrated four cases: The ISN, which stabilizes high excitation and supports the Paradoxical Effect, occurs when the E rates rise beyond a threshold. Supersaturation, which results in a decrease of E activity for increasing feedforward input, occurs when the I rates cross another threshold. Bistability, which supports the coexistence of two distinct stable states for the same level of input, occurs when the characteristic function $\mathcal{F}$ has more than two zeros. Finally, Structural Instability, which does not support any fixed point and results in runaway dynamics, occurs if the characteristic function has no zero. These various regimes are accessible to biologically-sized E-I networks without synaptic plasticity by tweaking the effective connection weights J_{XY}. Still, these weights were maintained within the range spanned by the experimentally reported mouse V1 connectivity to limit deviations from biological realism.

The ability of local E-I networks to carry out diverse nonlinear transformations on inputs can have profound functional implications. Since complex information processing must rely on the execution of underlying nonlinear operations, the various activity patterns accessible to local E-I networks can form the necessary basis for higher-order computations. As a proof of concept, I demonstrated how the association of supersaturating and bistable E-I networks can perform a binary summation on two signals.

While the SSN can generate various nonlinear transformations in networks of biologically realistic size, the Balanced State theory predicts that all regimes become linear as the network size tends to infinity. In the case where the network supersaturates, the balanced limit is unstable or simply does not exist. Then, I showed that the excitatory population gets increasingly suppressed as the network grows until it is permanently silenced. Additionally, if the network connectivity has a negative determinant, it will become structurally unstable and diverge to unlimited firing rates as size increases. Alternatively, if the connectivity satisfies the condition for a stable balanced limit, it will converge toward it. Still, I showed that this convergence can be quite slow, and network activity regimes can remain distinctly nonlinear at unrealistically large sizes. This illustrates that the predictions of the Balanced State framework are only valid at the theoretical limit and are not guaranteed to provide relevant insights for finite networks, no matter how large they are.

Finally, I demonstrated how the domain of validity of the SSN can be extended beyond homogeneous networks within which all neurons of a population share the same properties and receive the same input. In the case of unstructured networks, the variability cannot be explained by an underlying structure. The SSN can then be extended into a self-consistent system of equations by accounting for two additional variables: input variance and firing rate variance. This approach provides an exact framework to study the effect of inhomogeneities, but it comes at the cost of a significant increase in the complexity of the model. Nonetheless, it leads to valuable insights into the robustness of SSN-derived results with respect to network heterogeneity. In particular, I showed that moderate levels of variability, such as the use of an Erdös-Rényi connection scheme, only have a marginal effect on the resulting activity. In the case of structured networks, the underlying organization in feature space explains the heterogeneity. These systems are analogous to a continuum of E-I networks, and the self-referential nature of their equations requires them to be analyzed using numerical methods. Only the simplest rate models, such as the Balanced State, can be used to study structured networks analytically. Even then, additional assumptions on parameters must be made to reach an explicit solution.

5.2. Perspective

The matter of generating nonlinear transformations in neural network models has been an important point of focus in theoretical neuroscience. Because of the limitations imposed by the assured linearity of the classical Balanced State framework, multiple extensions of the model have been put forward to allow for richer dynamics. One potential source of nonlinearity is synaptic plasticity. By modulating synaptic strengths as a function of the firing rate, the Balanced State can be modified to support such complex regimes as bistability [118]. Another source of nonlinearity that has been studied in the Balanced State is the structure of the connectivity. Using a high-variance distribution of synaptic strengths instead of a homogeneous connectivity, the network activity is supported by a fraction of the neuronal population only and scales sublinearly with input [91]. Similarly, a complex connectivity structure can lead to the selective suppression of subsets of the neuronal populations. In this case, the Semi-Balanced framework considers that all neurons are either E-I balanced or silenced [13]. Following this, the global network response becomes a piece-wise linear manifold whose support varies with the specific subsets of active neurons. In this work, I restricted the analysis to nonlinearities emerging from the neuronal FI curve alone. Still, these diverse other sources of nonlinearity could be included into a more elaborate model, resulting in an even richer repertoire of network activity regimes. Likewise, a more detailed description of the neuronal FI curve could be introduced into the model to capture the effect of the activation function beyond the initial supralinear phase. In particular, the saturation that occurs for high firing rates because of the neuronal refractory period is another source of nonlinearity that has been shown to enable diverse network regimes [155].

One key aspect of LIF spiking simulations that could not be adequately captured by the LIF-adjusted SSN is its temporal dynamics. Although I showed that the steady state SSN equation suffices to predict the slow time evolution of spiking neural networks driven by external changes, it cannot be applied to model the fast dynamics that are intrinsic to these systems. In the classical SSN [150], the time evolution of the system is characterized by first-order ODEs and depends on one fixed time constant τ_{P_X} for each population. Yet, establishing a mathematical link between the value of these time constants and the parameters of a LIF neural network remains a formidable endeavor. The temporal properties of the network result from the complex interplay of the LIF membrane time constant, the dynamics of synaptic transmission, and the

state of the network. The exact transfer function governing the dynamics of neuronal populations has been derived using the LIF Fokker-Planck equations [156, 25, 26, 106]. Still, this transfer function only applies in the frequency domain, in the context of infinitesimal oscillations around an equilibrium point at which the system is linearized. Future research could set out to provide an applicable formulation of the LIF temporal transfer function, enabling rate models to track firing rate trajectories in time. Using this, the LIF-fitted SSN model presented here could be extended to capture the dynamical properties of spiking networks, which would open the way to study various features of these systems. It would allow for a complete characterization of the steady state stability conditions, which depend on the population time constants. It would also be needed to derive a mathematical description of mechanisms driving finite-size effects, such as spontaneous firing rate fluctuations. Finally, dynamic network regimes, such as the slow oscillations observed in the classic SSN [95], could be located in the parameter space of LIF networks and linked to biologically realistic implementations.

The approach presented here relies on the mathematical tractability of the SSN to allow for a thorough analysis of its potential features. Despite the sheer simplicity of homogeneous 2-population LIF networks, such systems can support a rich repertoire of transformations. Evidently, including more complex features into the network would result in even richer dynamics and a more realistic depiction of biological systems. Still, this would come at the cost of preventing such in-depth mathematical analysis.

Following this, the choice of studying rate models is motivated by their much lower dimensionality compared to the spiking network alternatives. Yet again, this framework presents limitations since diverse phenomena fall outside the scope of a firing-rate-centered description. Mechanisms that rely on the precise timing of spike emission are erased in the conversion from spiking networks to rate models. Despite the fact that such mechanisms as spike-timing dependent plasticity, dendritic computations, or neuronal avalanches could play a significant role in the function of neural circuits, they are inaccessible to the analysis of rate models. Fundamentally, the choice of a rate-based description of neural network activity exposes an underlying tacit assumption: that the information flow of neural communication is carried by firing rates rather than the precise timing of spike trains. Through the lens of spike-rate neural coding, substantial simplifications can be operated, resulting in a conveniently low-dimensional system. Still, while this approach is necessary to distill complex neural activity into more manageable quantities, it could overlook relevant features of neural network function [99, 187, 23].

Finally, the SSN-based approach I presented in this work is a promising tool for scaling up the analysis of neural networks. The SSN analysis can resolve the functional characteristics of neural systems at the level of medium-sized local circuits by reducing them to phenomenological input-output transformations. Following this, the analysis of large-sized neural networks can be simplified by converting them into interconnected systems of such well-characterized local elements. The issue of adopting an appropriate framework with respect to the scale of the system being investigated is becoming increasingly pertinent as ever-expanding amounts of detailed experimental measurements are made available [120, 108]. Despite this wealth of data, a complete mechanistic description of neural circuit operation, including algorithmic implementation and structural organization, has only ever been achieved for the simplest systems consisting of a handful of neurons [109, 191]. Meanwhile, at the other end of the spectrum, the Balanced State theoretical framework only applies to networks of infinite size and can result in unpredictable inaccuracy when used to study biologically-sized networks. Overall, I propose that the low-dimensional analysis of local networks can constitute a valuable intermediate step on the way to elucidating the mechanisms underlying the emergence of cognitive functions.

Bibliography

[1] L. F. Abbott. Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain research bulletin*, 50(5-6):303–304, 1999.

[2] M. Abramowitz, I. A. Stegun, and R. H. Romer. Handbook of mathematical functions with formulas, graphs, and mathematical tables, 1988.

[3] R. K. Adair. Noise and stochastic resonance in voltage-gated ion channels. *Proceedings of the National Academy of Sciences*, 100(21):12099–12104, 2003.

[4] Y. Ahmadian, D. B. Rubin, and K. D. Miller. Analysis of the Stabilized Supralinear Network. *Neural Computation*, 25(8):1994–2037, aug 2013. ISSN 0899-7667.

[5] A. K. Alijani and M. J. Richardson. Rate response of neurons subject to fast or frozen noise: from stochastic and homogeneous to deterministic and heterogeneous populations. *Physical Review E*, 84(1):011919, 2011.

[6] Allen Institute for Brain Science. Synaptic physiology coarse matrix dataset, 2019. Available from: https://brain-map.org/explore/connectivity/synaptic-physiology.

[7] 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.

[8] D. J. Amit and M. Tsodyks. Quantitative study of attractor neural network retrieving at low spike rates. i. substrate-spikes, rates and neuronal gain. *Network: Computation in neural systems*, 2(3):259, 1991.

[9] D. Arendt. Elementary nervous systems. *Philosophical Transactions of the Royal Society B*, 376(1821):20200347, 2021.

[10] G. Arfken and J. Romain. Mathematical methods for physicists. *Physics Today*, 20(5): 79–79, 1967.

[11] A. Arkhipov, N. W. Gouwens, Y. N. Billeh, S. Gratiy, R. Iyer, Z. Wei, Z. Xu, R. Abbasi-Asl, J. Berg, M. Buice, et al. Visual physiology of the layer 4 cortical circuit in silico. *PLoS Computational Biology*, 14(11):e1006535, 2018.

[12] B. V. Atallah, W. Bruns, M. Carandini, and M. Scanziani. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron*, 73(1):159–170, 2012.

[13] C. Baker, E. Froudarakis, D. Yatsenko, A. S. Tolias, and R. Rosenbaum. Inference of synaptic connectivity and external variability in neural microcircuits. *Journal of Computational Neuroscience*, 48(2):123–147, 2020. ISSN 15736873. doi: 10.1007/s10827-020-00739-4.

[14] C. Baker, V. Zhu, and R. Rosenbaum. Nonlinear stimulus representations in neural circuits with approximate excitatory-inhibitory balance. *PLOS Computational Biology*, 16 (9):1–30, 09 2020.

[15] S. J. Barnes, C. E. Cheetham, Y. Liu, S. H. Bennett, G. Albieri, A. A. Jorstad, G. W. Knott, and G. T. Finnerty. Delayed and temporally imprecise neurotransmission in reorganizing cortical microcircuits. *Journal of Neuroscience*, 35(24):9024–9037, 2015.

[16] S. J. Barnes, R. P. Sammons, R. I. Jacobsen, J. Mackie, G. B. Keller, and T. Keck. Subnetwork-specific homeostatic plasticity in mouse visual cortex in vivo. *Neuron*, 86 (5):1290–1303, 2015.

[17] R. Basu, R. Gebauer, T. Herfurth, S. Kolb, Z. Golipour, T. Tchumatchenko, and H. T. Ito. The orbitofrontal cortex maps future navigational goals. *Nature*, 599(7885):449–452, 2021.

[18] S. Becker, A. Nold, and T. Tchumatchenko. Modulation of working memory duration by synaptic and astrocytic mechanisms. *PLoS Computational Biology*, 18(10):e1010543, 2022.

[19] L. Bernáez Timón, P. Ekelmans, N. Kraynyukova, T. Rose, L. Busse, and T. Tchumatchenko. How to incorporate biological insights into network models and why it matters. *The Journal of Physiology*, 601(15):3037–3053, 2023.

[20] Y. N. Billeh, B. Cai, S. L. Gratiy, K. Dai, R. Iyer, N. W. Gouwens, R. Abbasi-Asl, X. Jia, J. H. Siegle, S. R. Olsen, C. Koch, S. Mihalas, and A. Arkhipov. Systematic Integration of Structural and Functional Data into Multi-scale Models of Mouse Primary Visual Cortex. *Neuron*, 106(3):388–403.e18, May 2020. ISSN 0896-6273. doi: 10.1016/j.neuron.2020.01.040.

[21] H. Blockus and F. Polleux. Developmental mechanisms underlying circuit wiring: Novel insights and challenges ahead. *Current opinion in neurobiology*, 66:205–211, 2021.

[22] K. A. Bolding and K. M. Franks. Recurrent cortical circuits implement concentration-invariant odor coding. *Science*, 361(6407):eaat6904, 2018.

[23] R. Brette. Philosophy of the spike: rate-based vs. spike-based theories of the brain. *Frontiers in systems neuroscience*, 9:151, 2015.

[24] N. Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience*, 8(3):183–208, 2000.

[25] N. Brunel and V. Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural computation*, 11(7):1621–1671, 1999.

[26] 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, 2001.

[27] M. A. Buice and C. C. Chow. Dynamic finite size effects in spiking neural networks. *PLoS computational biology*, 9(1):e1002872, 2013.

[28] M. A. Buice, J. D. Cowan, and C. C. Chow. Systematic fluctuation expansion for neural network activity equations. *Neural computation*, 22(2):377–426, 2010.

[29] A. N. Burkitt. A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological cybernetics*, 95(1):1–19, 2006.

[30] L. Busse. The mouse visual system and visual perception. In *Handbook of Behavioral Neuroscience*, volume 27, pages 53–68. Elsevier, 2018.

[31] L. Busse, A. R. Wade, and M. Carandini. Representation of concurrent stimuli by population activity in visual cortex. *Neuron*, 64(6):931–942, 2009.

[32] R. Capocelli and L. Ricciardi. Diffusion approximation and first passage time problem for a model neuron. *Kybernetik*, 8(6):214–223, 1971.

[33] F. Capone, M. Paolucci, F. Assenza, N. Brunelli, L. Ricci, L. Florio, and V. Di Lazzaro. Canonical cortical circuits: current evidence and theoretical implications. *Neuroscience and Neuroeconomics*, pages 1–8, 2016.

[34] M. Carandini and D. J. Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, 2012.

[35] M. Carandini and F. Sengpiel. Contrast invariance of functional maps in cat primary visual cortex. *Journal of vision*, 4(3):1–1, 2004.

[36] S. E. Cavanagh, J. D. Wallis, S. W. Kennerley, and L. T. Hunt. Autocorrelation structure at rest predicts value correlates of single neurons during reward-guided choice. *elife*, 5: e18937, 2016.

[37] S. E. Cavanagh, L. T. Hunt, and S. W. Kennerley. A diversity of intrinsic timescales underlie neural computations. *Frontiers in Neural Circuits*, 14:81, 2020.

[38] S. Chakraborty, R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao, et al. Interpretability of deep learning models: A survey of results. *IEEE*, pages 1–6, 2017.

[39] L. Chao-Yi and O. Creutzfeldt. The representation of contrast and other stimulus parameters by single neurons in area 17 of the cat. *Pflügers Archiv*, 401:304–314, 1984.

[40] G. Chen and P. Gong. Computing by modulating spontaneous cortical activity patterns as a mechanism of active visual processing. *Nature communications*, 10(1):1–15, 2019.

[41] B. W. Connors and M. J. Gutnick. Intrinsic firing patterns of diverse neocortical neurons. *Trends in neurosciences*, 13(3):99–104, 1990.

[42] L. Cossell, M. F. Iacaruso, D. R. Muir, R. Houlton, E. N. Sader, H. Ko, S. B. Hofer, and T. D. Mrsic-Flogel. Functional organization of excitatory synaptic strength in

primary visual cortex. *Nature*, 518(7539):399–403, 2015. ISSN 14764687. doi: 10.1038/nature14182.

[43] K. K. Cover and B. N. Mathur. Axo-axonic synapses: Diversity in neural circuit function. *Journal of Comparative Neurology*, 529(9):2391–2401, 2021.

[44] H. Dale. Pharmacology and nerve-endings, 1935.

[45] C. P. de Kock, J. Pie, A. W. Pieneman, R. A. Mease, A. Bast, J. M. Guest, M. Oberlaender, H. D. Mansvelder, and B. Sakmann. High-frequency burst spiking in layer 5 thick-tufted pyramids of rat primary somatosensory cortex encodes exploratory touch. *Communications biology*, 4(1):1–14, 2021.

[46] J. DeFelipe, L. Alonso-Nanclares, and J. I. Arellano. Microstructure of the neocortex: comparative aspects. *Journal of neurocytology*, 31:299–316, 2002.

[47] A. Destexhe, M. Rudolph, J. M. Fellous, and T. J. Sejnowski. Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons. *Neuroscience*, 107(1): 13–24, 2001.

[48] J. C. Eccles, P. Fatt, and K. Koketsu. Cholinergic and inhibitory synapses in a pathway from motor-axon collaterals to motoneurones. *The Journal of physiology*, 126(3):524, 1954.

[49] M. Eisenstein et al. Artificial intelligence powers protein-folding predictions. *Nature*, 599(7886):706–708, 2021.

[50] P. Ekelmans, N. Kraynyukova, and T. Tchumatchenko. Targeting operational regimes of interest in recurrent neural networks. *PLOS Computational Biology*, 19(5):e1011097, 2023.

[51] P. Erdös and A. Rényi. On random graphs i. *Publ. math. debrecen*, 6(290-297):18, 1959.

[52] G. B. Ermentrout and N. Kopell. Parabolic bursting in an excitable system coupled with a slow oscillation. *SIAM journal on applied mathematics*, 46(2):233–253, 1986.

[53] 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. ISSN 08966273. doi: 10.1016/j.neuron.2007.02.029.

[54] R. FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical journal*, 1(6):445–466, 1961.

[55] N. Fourcaud and N. Brunel. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural computation*, 14(9):2057–2110, 2002.

[56] 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.

[57] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

[58] J. M. Fuster and G. E. Alexander. Neuron activity related to short-term memory. *Science*, 173(3997):652–654, 1971.

[59] L. J. Gentet, Y. Kremer, H. Taniguchi, Z. J. Huang, J. F. Staiger, and C. C. Petersen. Unique functional properties of somatostatin-expressing gabaergic neurons in mouse barrel cortex. *Nature neuroscience*, 15(4):607–612, 2012.

[60] W. Gerstner and W. M. Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002.

[61] C. N. Giachello and R. A. Baines. Inappropriate neural activity during a sensitive period in embryogenesis results in persistent seizure-like behavior. *Current Biology*, 25(22): 2964–2968, 2015.

[62] A. Goulas, F. Damicelli, and C. C. Hilgetag. Bio-instantiated recurrent neural networks: Integrating neurobiology-based network topology in artificial networks. *Neural Networks*, 142:608–618, 2021.

[63] J. W. Graham, M. E. Gevaert, T. K. Berger, G. A. A. Kahou, C. A. Sanchez, M. W. Reimann, H. Markram, E. Muller, S. Ramaswamy, M. Abdellah, et al. Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2), 2015.

[64] J. M. Guest, A. Bast, R. T. Narayanan, and M. Oberlaender. Thalamus gates active dendritic computations in cortex during sensory processing. *bioRxiv*, 2021.

[65] M. Gur and D. M. Snodderly. Physiological differences between neurons in layer 2 and layer 3 of primary visual cortex (v1) of alert macaque monkeys. *The Journal of physiology*, 586(9):2293–2306, 2008.

[66] R. L. Hahnloser. On the piecewise analysis of networks of linear threshold neurons. *Neural Networks*, 11(4):691–697, 1998.

[67] 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.

[68] S. Harding-Forrester and D. E. Feldman. Somatosensory maps. *Handbook of clinical neurology*, 151:73–102, 2018.

[69] K. B. Hengen, M. E. Lambo, S. D. Van Hooser, D. B. Katz, and G. G. Turrigiano. Firing rate homeostasis in visual cortex of freely behaving rodents. *Neuron*, 80(2):335–342, 2013.

[70] A. L. Hodgkin and A. F. Huxley. The components of membrane conductance in the giant axon of loligo. *The Journal of physiology*, 116(4):473, 1952.

[71] A. L. Hodgkin and A. F. Huxley. Currents carried by sodium and potassium ions through the membrane of the giant axon of loligo. *The Journal of physiology*, 116(4):449, 1952.

[72] A. L. Hodgkin and A. F. Huxley. The dual effect of membrane potential on sodium conductance in the giant axon of loligo. *The Journal of physiology*, 116(4):497, 1952.

[73] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4): 500–544, 1952.

[74] A. L. Hodgkin, A. F. Huxley, and B. Katz. Measurement of current-voltage relations in the membrane of the giant axon of loligo. *The Journal of physiology*, 116(4):424, 1952.

[75] S. B. Hofer, H. Ko, B. Pichler, J. Vogelstein, H. Ros, H. Zeng, E. Lein, N. A. Lesica, and T. D. Mrsic-Flogel. Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neuroscience*, 14(8):1045–52, 2011. doi: 10.1038/nn.2876.

[76] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[77] T. Hosoya. The basic repeating modules of the cerebral cortical circuit. *Proceedings of the Japan Academy, Series B*, 95(7):303–311, 2019.

[78] T. Hromádka, M. R. DeWeese, and A. M. Zador. Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS biology*, 6(1):e16, 2008.

[79] Y. Hu, J. Trousdale, K. Josić, and E. Shea-Brown. Motif statistics and spike correlations in neuronal networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2013 (03):P03012, 2013.

[80] D. Hubel, T. Wiesel, and S. LeVay. Functional architecture of area 17 in normal and monocularly deprived macaque monkeys. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 40, pages 581–589. Cold Spring Harbor Laboratory Press, 1976.

[81] E. M. Izhikevich. Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6):1569–1572, 2003.

[82] E. M. Izhikevich. Which model to use for cortical spiking neurons? *IEEE transactions on neural networks*, 15(5):1063–1070, 2004.

[83] J. Jang, M. Song, and S.-B. Paik. Retino-cortical mapping ratio predicts columnar and salt-and-pepper organization in mammalian visual cortex. *Cell reports*, 30(10):3270–3279, 2020.

[84] D. Jercog, A. Roxin, P. Bartho, A. Luczak, A. Compte, and J. de la Rocha. Up-down cortical dynamics reflect state transitions in a bistable network. *Elife*, 6:e22425, 2017.

[85] P. Jiruska, J. Csicsvari, A. D. Powell, J. E. Fox, W.-C. Chang, M. Vreugdenhil, X. Li, M. Palus, A. F. Bujan, R. W. Dearden, et al. High-frequency network activity, global increase in neuronal activity, and synchrony expansion precede epileptic seizures in vitro. *Journal of Neuroscience*, 30(16):5690–5701, 2010.

[86] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. Siegelbaum, A. J. Hudspeth, S. Mack, et al. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

[87] K. Kar, J. Kubilius, K. Schmidt, E. B. Issa, and J. J. DiCarlo. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature neuroscience*, 22(6):974–983, 2019.

[88] M. Kaschube, M. Schnabel, S. Löwel, D. M. Coppola, L. E. White, and F. Wolf. Universality in the evolution of orientation columns in the visual cortex. *science*, 330(6007): 1113–1116, 2010.

[89] B. Katz and R. Miledi. The measurement of synaptic delay, and the time course of acetylcholine release at the neuromuscular junction. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 161(985):483–495, 1965.

[90] D. Keller, C. Erö, and H. Markram. Cell densities in the mouse brain: a systematic review. *Frontiers in neuroanatomy*, 12:83, 2018.

[91] R. Khajeh, F. Fumarola, and L. Abbott. Sparse balance: Excitatory-inhibitory networks with small bias currents and broadly distributed synaptic weights. *PLOS Computational Biology*, 18(2):e1008836, 2022.

[92] A. G. Khan, J. Poort, A. Chadwick, A. Blot, M. Sahani, T. D. Mrsic-Flogel, and S. B. Hofer. Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nature Neuroscience*, 21(6):851–859, 2018. ISSN 15461726. doi: 10.1038/s41593-018-0143-z.

[93] B. W. Knight. Dynamics of encoding in a population of neurons. *The Journal of general physiology*, 59(6):734–766, 1972.

[94] H. Ko, S. B. Hofer, B. Pichler, K. A. Buchanan, P. J. Sjöström, and T. D. Mrsic-Flogel. Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473 (7345):87–91, 2011. ISSN 00280836. doi: 10.1038/nature09880.

[95] N. Kraynyukova and T. Tchumatchenko. Stabilized supralinear network can give rise to bistable, oscillatory, and persistent activity. *Proceedings of the National Academy of Sciences*, 115(13):3464–3469, 2018.

[96] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25: 1097–1105, 2012.

[97] L. Lapicque. Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Gen.*, 9:620–635, 1907.

[98] G. Laurent. Private communication, March. 2021.

[99] V. Lawhern, A. A. Nikonov, W. Wu, and R. J. Contreras. Spike rate and spike timing contributions to coding taste quality information in rat periphery. *Frontiers in Integrative Neuroscience*, 5:18, 2011.

[100] T. Ledgeway, Z. Chang'an, A. P. Johnson, Y. Song, and C. L. Baker. The direction-selective contrast response of area 18 neurons is different for first-and second-order motion. *Visual neuroscience*, 22(1):87–99, 2005.

[101] S. Lefort, C. Tomm, J.-C. F. Sarria, and C. C. Petersen. The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex. *Neuron*, 61(2):301–316, 2009.

[102] D. S. Lemons and P. Langevin. *An Introduction to Stochastic Processes in Physics*. JHU Press, 2002.

[103] L.-y. Li, Y.-t. Li, M. Zhou, H. W. Tao, and L. I. Zhang. Intracortical multiplication of thalamocortical signals in mouse auditory cortex. *Nature Neuroscience*, 16(9):1179–1181, Sept. 2013. ISSN 1546-1726. doi: 10.1038/nn.3493.

[104] Y.-t. Li, L. A. Ibrahim, B.-h. Liu, L. I. Zhang, and H. W. Tao. Linear transformation of thalamocortical input by intracortical excitation. *Nature neuroscience*, 16(9):1324–1330, 2013.

[105] A. D. Lien and M. Scanziani. Tuned thalamic excitation is amplified by visual cortical circuits. *Nature Neuroscience*, 16(9):1315–1323, 2013.

[106] B. Lindner and L. Schimansky-Geier. Transmission of noise coded versus additive signals through a neuronal ensemble. *Physical Review Letters*, 86(14):2934, 2001.

[107] M. London and M. Häusser. Dendritic computation. *Annu. Rev. Neurosci.*, 28:503–532, 2005.

[108] S. Loomba, J. Straehle, V. Gangadharan, N. Heike, A. Khalifa, A. Motta, N. Ju, M. Sievers, J. Gempt, H. S. Meyer, et al. Connectomic comparison of mouse and human cortex. *Science*, 377(6602):eabo0924, 2022.

[109] E. Marder. Neuromodulation of neuronal circuits: back to the future. *Neuron*, 76(1): 1–11, 2012.

[110] T. W. Margrie, M. Brecht, and B. Sakmann. In vivo, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflügers Archiv*, 444(4):491–498, 2002.

[111] H. Markram, E. Muller, S. Ramaswamy, M. W. Reimann, M. Abdellah, C. A. Sanchez, A. Ailamaki, L. Alonso-Nanclares, N. Antille, S. Arsever, G. A. A. Kahou, T. K. Berger, A. Bilgili, N. Buncic, A. Chalimourda, et al. Reconstruction and Simulation of Neocortical Microcircuitry. *Cell*, 163(2):456–492, Oct. 2015. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2015.09.029.

[112] R. H. Masland. Neuronal cell types. *Current Biology*, 14(13):R497–R500, 2004.

[113] S. D. Meriney and E. Fanselow. *Synaptic transmission*. Academic Press, 2019.

[114] H. S. Meyer, V. C. Wimmer, M. Oberlaender, C. P. De Kock, B. Sakmann, and M. Helmstaedter. Number and laminar distribution of neurons in a thalamocortical projection column of rat vibrissal cortex. *Cerebral cortex*, 20(10):2277–2286, 2010.

[115] K. D. Miller and A. Palmigiano. Generalized paradoxical effects in excitatory/inhibitory networks. *bioRxiv*, 2020.

[116] M. Minsky and S. Papert. Perceptron: an introduction to computational geometry, 1969.

[117] M. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, MA, USA, 1969.

[118] G. Mongillo, D. Hansel, and C. Van Vreeswijk. Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. *Physical review letters*, 108(15):158101, 2012.

[119] E. Montbrió, D. Pazó, and A. Roxin. Macroscopic description for networks of spiking neurons. *Physical Review X*, 5(2):021028, 2015.

[120] A. Motta, M. Berning, K. M. Boergens, B. Staffler, M. Beining, S. Loomba, P. Hennig, H. Wissler, and M. Helmstaedter. Dense connectomic reconstruction in layer 4 of the somatosensory cortex. *Science*, 366(6469), 2019.

[121] V. B. Mountcastle. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of neurophysiology*, 20(4):408–434, 1957.

[122] J. D. Murray, A. Bernacchia, D. J. Freedman, R. Romo, J. D. Wallis, X. Cai, C. Padoa-Schioppa, T. Pasternak, H. Seo, D. Lee, et al. A hierarchy of intrinsic timescales across primate cortex. *Nature neuroscience*, 17(12):1661–1663, 2014.

[123] M. Oberlaender, C. P. de Kock, R. M. Bruno, A. Ramirez, H. S. Meyer, V. J. Dercksen, M. Helmstaedter, and B. Sakmann. Cell type–specific three-dimensional structure of thalamocortical circuits in a column of rat vibrissal cortex. *Cerebral cortex*, 22(10): 2375–2391, 2012.

[124] K. Obermayer and G. G. Blasdel. Geometry of orientation and ocular dominance columns in monkey striate cortex. *Journal of Neuroscience*, 13(10):4114–4129, 1993.

[125] D. H. O'Connor, S. P. Peron, D. Huber, and K. Svoboda. Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron*, 67(6):1048–1061, 2010.

[126] T. Ohshiro, D. E. Angelaki, and G. C. DeAngelis. A normalization model of multisensory integration. *Nature neuroscience*, 14(6):775–782, 2011.

[127] M. Okun and I. Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature neuroscience*, 11(5):535–537, 2008.

[128] J. O'Kusky and M. Colonnier. A laminar analysis of the number of neurons, glia, and synapses in the visual cortex (area 17) of adult macaque monkeys. *Journal of Comparative Neurology*, 210(3):278–290, 1982.

[129] R. OpenAI. Gpt-4 technical report (arxiv: 2303.08774), 2023.

[130] M. Opper and D. Saad. *Advanced mean field methods: Theory and practice*. MIT press, 2001.

[131] H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron*, 62(4):578–592, 2009.

[132] A. Palmigiano, F. Fumarola, D. P. Mossing, N. Kraynyukova, H. Adesnik, and K. D. Miller. Common rules underlying optogenetic and behavioral modulation of responses in multi-cell-type v1 circuits. *bioRxiv*, 2023. doi: 10.1101/2020.11.11.378729.

[133] J. W. Peirce. The potential importance of saturating and supersaturating contrast response functions in visual cortex. *Journal of vision*, 7(6):13–13, 2007.

[134] E. Persi, D. Hansel, L. Nowak, P. Barone, and C. Van Vreeswijk. Power-law input-output transfer functions explain the contrast-response and tuning properties of neurons in visual cortex. *PLoS Comput Biol*, 7(2):e1001078, 2011.

[135] C. C. Petersen. Whole-cell recording of neuronal membrane potential during behavior. *Neuron*, 95(6):1266–1281, 2017.

[136] K.-E. Petousakis, A. A. Apostolopoulou, and P. Poirazi. The impact of hodgkin–huxley models on dendritic research. *The Journal of Physiology*, 601(15):3091–3102, 2023.

[137] C. K. Pfeffer, M. Xue, M. He, Z. J. Huang, and M. Scanziani. Inhibition of inhibition in visual cortex: The logic of connections between molecularly distinct interneurons. *Nature Neuroscience*, 16(8):1068–1076, 2013. ISSN 10976256. doi: 10.1038/nn.3446.

[138] D. M. Piscopo, R. N. El-Danaf, A. D. Huberman, and C. M. Niell. Diverse visual features encoded in mouse lateral geniculate nucleus. *Journal of Neuroscience*, 33(11):4642–4656, 2013.

[139] N. J. Priebe and D. Ferster. Inhibition, spike threshold, and stimulus selectivity in primary visual cortex. *Neuron*, 57(4):482–497, 2008.

[140] N. J. Priebe, F. Mechler, M. Carandini, and D. Ferster. The contribution of spike threshold to the dichotomy of cortical simple and complex cells. *Nature Neuroscience*, 7(10):1113–1122, 2004.

[141] D. Purves and S. M. Williams. *Neuroscience*. Sinauer Associates, 2001.

[142] S. Ramaswamy, J.-D. Courcol, M. Abdellah, S. R. Adaszewski, N. Antille, S. Arsever, G. Atenekeng, A. Bilgili, Y. Brukau, A. Chalimourda, et al. The neocortical microcircuit collaboration portal: a resource for rat somatosensory cortex. *Frontiers in neural circuits*, 9:44, 2015.

[143] E. A. Rancz, K. M. Franks, M. K. Schwarz, B. Pichler, A. T. Schaefer, and T. W. Margrie. Transfection via whole-cell recording in vivo: bridging single-cell physiology, genetics and connectomics. *Nature neuroscience*, 14(4):527–532, 2011.

[144] 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.

[145] L. M. Ricciardi. The first passage time problem. In *Diffusion Processes and Related Topics in Biology*, pages 61–72. Springer, 1977.

[146] D. Ringach and R. Shapley. Reverse correlation in neurophysiology. *Cognitive Science*, 28(2):147–166, 2004.

[147] R. Rosenbaum and B. Doiron. Balanced networks of spiking neurons with spatially dependent recurrent connections. *Physical Review X*, 4(2):021039, 2014.

[148] G. Roth. Convergent evolution of complex brains and high intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1684):20150049, 2015.

[149] 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.

[150] D. B. Rubin, S. D. Van Hooser, and K. D. Miller. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 85 (2):402–417, 2015.

[151] P. Rupprecht and R. W. Friedrich. Precise synaptic balance in the zebrafish homolog of olfactory cortex. *Neuron*, 100(3):669–683, 2018.

[152] S. Sadeh and C. Clopath. Inhibitory stabilization and cortical computation. *Nature Reviews Neuroscience*, 22(1):21–37, Jan. 2021. ISSN 1471-0048. doi: 10.1038/s41583-020-00390-z.

[153] M. Saenz and D. R. Langers. Tonotopic mapping of human auditory cortex. *Hearing research*, 307:42–52, 2014.

[154] A. Sanzeni, B. Akitake, H. C. Goldbach, C. E. Leedy, N. Brunel, and M. H. Histed. Inhibition stabilization is a widespread property of cortical networks. *eLife*, 9:e54875, June 2020. ISSN 2050-084X. doi: 10.7554/eLife.54875.

[155] A. Sanzeni, M. H. Histed, and N. Brunel. Response nonlinearities in networks of spiking neurons. *PLoS computational biology*, 16(9):e1008165, 2020.

[156] J. Schuecker, M. Diesmann, and M. Helias. Modulated escape from a metastable state driven by colored noise. *Physical Review E*, 92(5):052119, 2015.

[157] D. T. Schultz, S. H. Haddock, J. V. Bredeson, R. E. Green, O. Simakov, and D. S. Rokhsar. Ancient gene linkages support ctenophores as sister to other animals. *Nature*, pages 1–8, 2023.

[158] A. Schüz and G. Palm. Density of neurons and synapses in the cerebral cortex of the mouse. *Journal of Comparative Neurology*, 286(4):442–455, 1989.

[159] T. Schwalger, M. Deger, and W. Gerstner. Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size. *PLoS computational biology*, 13(4):e1005507, 2017.

[160] A. Sedra, K. Smith, T. C. Carusone, and V. Gaudet. Microelectronic circuits 8th edition. *Chapter*, 14:1235–1236, 2020.

[161] S. C. Seeman, L. Campagnola, P. A. Davoudian, A. Hoggarth, T. A. Hage, A. Bosma-Moody, C. A. Baker, J. H. Lee, S. Mihalas, C. Teeter, et al. Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex. *Elife*, 7:e37349, 2018.

[162] I. Segev. Dendritic processing the handbook of brain theory and neural networks ed ma arbib, 1995.

[163] M. N. Shadlen and W. T. Newsome. Noise, neural codes and cortical organization. *Current opinion in neurobiology*, 4(4):569–579, 1994.

[164] Y. Shu, A. Hasenstaub, and D. A. McCormick. Turning on and off recurrent balanced cortical activity. *Nature*, 423(6937):288–293, 2003.

[165] A. J. Siegert. On the first passage time probability problem. *Physical Review*, 81(4):617, 1951.

[166] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

[167] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

[168] 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.

[169] V. S. Sohal and J. L. Rubenstein. Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Molecular psychiatry*, 24(9):1248–1257, 2019.

[170] O. Sporns and R. Kötter. Motifs in brain networks. *PLoS biology*, 2(11):e369, 2004.

[171] M. Stobb, J. M. Peterson, B. Mazzag, and E. Gahtan. Graph theoretical model of a sensorimotor connectome in zebrafish. *PLoS One*, 7(5):e37292, 2012.

[172] A. Stuchlik. Dynamic learning and memory, synaptic plasticity and neurogenesis: an update. *Frontiers in behavioral neuroscience*, 8:106, 2014.

[173] L. E. Suárez, B. A. Richards, G. Lajoie, and B. Misic. Learning function from structure in neuromorphic networks. *Nature Machine Intelligence*, 3(9):771–786, 2021.

[174] D. Sussillo, M. M. Churchland, M. T. Kaufman, and K. V. Shenoy. A neural network that finds a naturalistic solution for the production of muscle activity. *Nature neuroscience*, 18(7):1025–1033, 2015.

[175] A. Y. Tan, B. D. Brown, B. Scholl, D. Mohanty, and N. J. Priebe. Orientation selectivity of synaptic input to neurons in mouse and cat primary visual cortex. *Journal of Neuroscience*, 31(34):12339–12350, 2011.

[176] T. Tchumatchenko, A. Malyshev, T. Geisel, M. Volgushev, and F. Wolf. Correlations and synchrony in threshold neuron models. *Phys. Rev. Lett.*, 104:058102, Feb 2010.

[177] L. B. Timón, P. Ekelmans, S. Konrad, A. Nold, and T. Tchumatchenko. Synaptic plasticity controls the emergence of population-wide invariant representations in balanced network models. *Physical Review Research*, 4(1):013162, 2022.

[178] M. A. Tosches. Developmental and genetic mechanisms of neural circuit evolution. *Developmental biology*, 431(1):16–25, 2017.

[179] M. V. Tsodyks and H. Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the national academy of sciences*, 94(2):719–723, 1997.

[180] M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski, and B. L. McNaughton. Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus*, 6(3):271–280, 1996.

[181] M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski, and B. L. McNaughton. Paradoxical effects of external modulation of inhibitory interneurons. *Journal of neuroscience*, 17 (11):4382–4388, 1997.

[182] A. M. Turner and W. T. Greenough. Differential rearing effects on rat visual cortex synapses. i. synaptic and neuronal density and synapses per neuron. *Brain research*, 329 (1-2):195–203, 1985.

[183] C. W. Tyler and P. A. Apkarian. Effects of contrast, orientation and binocularity in the pattern evoked potential. *Vision research*, 25(6):755–766, 1985.

[184] S. D. Van Hooser, J. A. F. Heimel, S. Chung, S. B. Nelson, and L. J. Toth. Orientation selectivity without orientation maps in visual cortex of a highly visual mammal. *Journal of Neuroscience*, 25(1):19–28, 2005.

[185] C. van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.

[186] C. van Vreeswijk and H. Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural computation*, 10(6):1321–71, 1998.

[187] R. VanRullen, R. Guyonneau, and S. J. Thorpe. Spike times make sense. *Trends in neurosciences*, 28(1):1–4, 2005.

[188] H. R. Wilson and J. D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1–24, 1972.

[189] T. A. Woolsey and H. Van der Loos. The structural organization of layer iv in the somatosensory region (si) of mouse cerebral cortex: the description of a cortical field composed of discrete cytoarchitectonic units. *Brain research*, 17(2):205–242, 1970.

[190] Y. K. Wu and F. Zenke. Nonlinear transient amplification in recurrent neural networks with short-term plasticity. *Elife*, 10:e71263, 2021.

[191] J. Yoshino, R. K. Morikawa, E. Hasegawa, and K. Emoto. Neural circuitry that evokes escape behavior upon activation of nociceptive sensory neurons in drosophila larvae. *Current Biology*, 27(16):2499–2504, 2017.

[192] J. Zhuang, L. Ng, D. Williams, M. Valley, Y. Li, M. Garrett, and J. Waters. An extended retinotopic map of mouse cortex. *elife*, 6:e18372, 2017.

List of Publications

The work I did during my doctoral studies resulted in the following publications:

- Bernáez Timón, L., Ekelmans, P., Konrad, S., Nold, A., and Tchumatchenko, T.
  Synaptic plasticity controls the emergence of population-wide invariant representations in balanced network models.
  Physical Review Research 4, no. 1 (2022): 013162. [177]
  https://doi.org/10.1103/PhysRevResearch.4.013162

- Ekelmans, P., Bernáez Timón, L., Kraynyukova, N., Rose, T., Busse, L., and Tchumatchenko, T.
  How to incorporate biological insights into network models and why it matters.
  The Journal of Physiology, 601(15) (2023): 3037-3053. [19]
  https://doi.org/10.1113/JP282755

- Ekelmans, P., Kraynyukova, N., and Tchumatchenko, T.
  Targeting operational regimes of interest in recurrent neural networks.
  PLOS Computational Biology 19, no. 5 (2023): e1011097. [50]
  https://doi.org/10.1371/journal.pcbi.1011097

A. Appendix A

A.1. Implementation of the spiking network simulator

All LIF spiking network simulations are performed using the spiking network simulator developed collaboratively within the Tchumatchenko lab. The code and practical information can be found on the following GitHub repository:
https://github.com/CompNeuroTchuGroup/TchuSNNet/tree/Ekelmans_et_al_2023
Overall, the simulator considers N_p populations, each consisting of N_X neurons. Each neuron is treated as a separate object with its own incoming and outgoing synapses. The state of each neuron (i.e., its membrane potential V) is tracked throughout the simulation and updated according to the LIF rule (Eq. 1.1). The feedforward input is a Gaussian pseudo-randomly generated number, with a different seed for each cell and each time step. The recurrent input results from the spiking events of LIF neurons and the network connectivity which is randomly established at the initialization stage of the simulation. The simulation follows a forward Euler integration scheme with a fixed time step dt (usually $5 \times 10^{-5}\,\mathrm{s}$). Over the course of a simulation, multiple measurements of the network state can be recorded. In this work, we primarily focus on the average population firing rate which is the total number of emitted spike in a population over a time bin, normalized by the population size and the bin size. The time bin is usually set at $100\,\mathrm{ms}$, although its size should not affect the recorded mean firing rate since the networks are studied at the steady state.

In the default network implementation, synaptic transmissions occur over a single time step:

$$I_{XY}(t) = j_{XY}\,\delta(t - t_s), \tag{A.1}$$

where t_s is the spike time of the presynaptic neuron, and δ is the Dirac delta function.

In some network configurations, this type of synaptic transmission promotes synchronization at the neuronal population level, resulting in population spikes [18, 24] (see Fig. A.8). Such phenomena can only arise when the assumption of asynchrony and irregularity is broken and are therefore outside the scope of the mean-field approach. In order to avoid this synchronization

in such cases, a different type of synaptic transmission can be used:

$$I_{XY}(t) = \frac{j_{XY}}{\tau_s} e^{-\frac{t-(t_s+D)}{\tau_s}} \mathcal{H}(t - t_s - D), \tag{A.2}$$

where τ_s is a synaptic transmission time constant, D is the synaptic delay and $\mathcal{H}$ is the Heaviside function. This type of synapse implementation prevents synchronization as the effect of each spike is more distributed in time and each synaptic connection has a different delay D.

This exponentially decaying synaptic type was employed in the spiking network simulation of bistable or structurally unstable networks, and Dirac-delta synapses were used in all other cases. The exponential synapse parameters are presented in Table A.1. It should be noted that the synaptic delay values we use are higher than experimentally reported [141, 15, 89], but this jitter only serves to prevent population synchrony and does not affect the steady state network activity. On the other hand, the use of exponentially decaying synaptic transmission can lead to changes in the properties of the recurrent noise [5, 55], which would affect the mean-field steady state. Nevertheless, the contribution of recurrent noise is considered negligible compared to external noise (section 2.3.3) and is not accounted for in the SSN framework.

τ_S	D	**Affected Synapses**
10	0 - 100	EE, II

Table A.1.: Synaptic parameters used to prevent synchronization, following Eq. A.2. The exponential synapse time constant τ_S and the synaptic delay D are given in ms. The synaptic delay for each synapse is drawn randomly from a uniform distribution in the given range. The E→I and I→E synapses are not affected and use Dirac-delta synapses without delay (Eq. A.1). This applies to LIF simulations shown in Fig. 3.4,Fig. 3.6, Fig. 3.10.D, Fig. 3.11.B-C, Fig. 3.12.B and Fig. A.7.B.

A.2. Continuation method

The steady state firing rates of neural networks are given by the solutions of self-consistency equations (i.e., Eq. 2.2 or Eq. 1.4). In general, such systems of equations are too complex to solve analytically and require numerical approaches to solve them, such as the *fsolve* Matlab function. When the aim is to determine the geometric locus of solutions with one control parameter freely varying (e.g., the steady state activity of the network as a function of the feedforward input μ_{ext}), the equation solver will require additional guidance.

The equation solver can only converge to a single solution even if the system of equations admits multiple solutions. Furthermore, the algorithm can get trapped near a local minimum and fail to reach the solution. We use the continuation method to guide the nonlinear system solver from one solution to the next in order to facilitate the search and ensure that the algorithm browses along the geometrical locus of solutions and does not return the same solution multiple times.

At the initial point, the state variables Y are determined using the solver for the control parameter X set at its initial point X_0, using an initial guess $\tilde{Y_0}$:

$$Y_0 = \beth\left(\tilde{Y_0}, X_0\right) \quad | \quad X = X_0 \,,$$

where $\beth$ is the optimization function. The next point is obtained using a user-defined vector v_0, which is presumed to point towards the next point, leading to the next guess $\tilde{X_1}, \tilde{Y_1}$. The search is then restricted to be on the hyperplane orthogonal to the vector v_0:

$$\begin{cases} (\tilde{X_1}, \tilde{Y_1}) = & (X_0, Y_0) + v_0 \\ (X_1, Y_1) = & \beth\left(\tilde{X_1}, \tilde{Y_1}\right) \quad | \quad \left((X_1, Y_1) - (\tilde{X_1}, \tilde{Y_1})\right) \cdot v_0 = 0, \end{cases}$$

where the scalar product $(\cdot)$ imposes that the deviation from the guessed solution is orthogonal to the vector v_0. Beyond this, all future points are obtained by extrapolating from the position of the two previous points so the new vector v_i leads to the next guess.

$$\begin{cases} v_i & = (X_i, Y_i) - (X_{i-1}, Y_{i-1}) \\ (\tilde{X}_{i+1}, \tilde{Y}_{i+1}) & = (Y_i, X_i) + \Delta \left\| v_i \right\|^{-1} v_i \\ (X_{i+1}, Y_{i+1}) & = \beth\left(\tilde{X}_{i+1}, \tilde{Y}_{i+1}\right) \quad | \quad \left((X_{i+1}, Y_{i+1}) - (\tilde{X}_{i+1}, \tilde{Y}_{i+1})\right) \cdot v_i = 0, \end{cases}$$

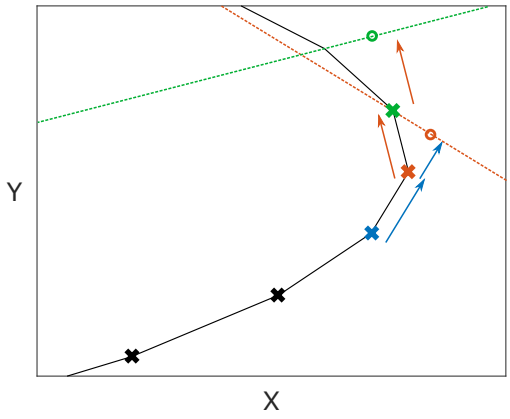


Figure A.1.: **2-D Schematic representation of the continuation method.** The expected position of the (i+1)$^{\text{th}}$ point (red circle) is extrapolated from the location of the points i and (i-1) (blue and red crosses) in the direction of the vector v_i (blue arrow). The (i+1)$^{\text{th}}$ point is then found by solving the objective function with the constraint that the solution (green cross) is located on the hyperplane containing the initial guess (red circle) that is orthogonal to the vector v_i (red line).

where the vector v_i is normalized to control the step size Δ between successive points. In particular, Δ can be adjusted to be smaller near bending points in the curve to limit the error. The degree of bending in the curve can be measured by the scalar product of successive vectors $(v_{i-1} \cdot v_i)$. A schematic representation of the method is shown in Fig. A.1.

The continuation method is particularly valuable in the case of bistability, so all coexisting steady states can be recovered. In particular, it can indiscriminately find the stable and unstable fixed points. This cannot be achieved with the more straightforward approach of integrating the differential equations in time (e.g., Eq. 1.8) to converge towards the steady state, as unstable fixed points are repellent. One drawback of the continuation method is that it can only follow the branch of solutions on which it was initiated. In situations where multiple separate branches exist (Fig. 3.7.C,F), the algorithm must be initialized multiple times to cover them all.

In the case of the 2-D SSN, the steady state solutions as a function of the external inputs can simply be obtained using methods for implicit functions, such as the Matlab function *fimplicit*, on the characteristic function $\mathcal{F}$ (Eq. 2.6). However, for more complex systems such as the 3-D SSN or the Ricciardi self-consistent equations (Eq. 1.4), more robust methods like the continuation method are required.

A.3. Balanced State validity

Within the Balanced State framework, we assume that the dynamics for the change in firing rate of each population is a function of the excess input it receives.

$$\frac{d\nu_E}{dt} = f(\mu_E)$$
$$\frac{d\nu_I}{dt} = g(\mu_I).$$

The firing rate of a population is at the steady state when its total input is balanced ($f(0) = g(0) = 0$). By linearising around a steady state, we get

$$\begin{bmatrix} \dot{\delta\nu_E} \\ \dot{\delta\nu_I} \end{bmatrix} = \begin{bmatrix} f'(\mu_E^0) & 0 \\ 0 & g'(\mu_I^0) \end{bmatrix} \begin{bmatrix} J_{EE} & -J_{EI} \\ J_{IE} & -J_{II} \end{bmatrix} \begin{bmatrix} \delta\nu_E \\ \delta\nu_I \end{bmatrix}.$$

Where $f'(\mu_E^0)$ and $g'(\mu_I^0)$ are positive (excess input drives the firing rate up). The state (ν_E^0, ν_I^0) is stable if the two eigenvalues of the Jacobian matrix

$$\begin{bmatrix} f'J_{EE} & -f'J_{EI} \\ g'J_{IE} & -g'J_{II} \end{bmatrix}$$

have negative real parts.

The eigenvalues λ_1 and λ_2 are roots of the polynomial

$$\lambda^2 - \lambda(f'J_{EE} - g'J_{II}) + f'g'(J_{EI}J_{IE} - J_{EE}J_{II}),$$

or

$$\lambda_1, \lambda_2 = \left(\frac{f'J_{EE} - g'J_{II}}{2}\right)\left(1 \pm \sqrt{1 - \frac{4f'g' \det J}{(f'J_{EE} - g'J_{II})^2}}\right).$$

The steady state is stable $\iff$

$$\begin{cases} f'J_{EE} - g'J_{II} < 0 \\ \det J > 0\,. \end{cases} \tag{A.3}$$

The first condition requires that the response of the inhibitory population (g') is fast and strong enough to prevent a runaway excitatory feedback loop. We do not focus on this condition here because it depends on the dynamic properties of the network (f and g functions), which are beyond the scope of this work. However, the second condition constrains the connectivity matrix such that $J_{EI}J_{IE} - J_{EE}J_{II} > 0$ [147].

The condition on the existence of a non-negative Balanced State (Eq. 1.7) can be combined with the stability condition on connectivity (Eq. A.3) to delineate the parameter range where a Balanced State limit exists and is stable [186, 147]:

$$\begin{cases} 0 < r < \frac{J_{II}}{J_{EI}} \\ \det J > 0 \\ f' J_{EE} - g' J_{II} < 0 \, . \end{cases} \tag{A.4}$$

A.4. Derivation of the Φ function

Here, I provide the main steps of the mathematical derivation of the theoretical activation function of LIF neurons fed with white noise [145, 32, 24].

Given a white noise input, the membrane potential of the LIF neuron follows a Ornstein-Uhlenback process, with a drift caused by the mean input μ, according to the master equation

$$dV = \mu dt - \frac{V}{\tau} + \sigma\sqrt{dt}\mathcal{N}_t^{t+dt}(0,1),$$

where $\mathcal{N}_t^{t+dt}(0,1)$ is a normally distributed random variable for the interval $[t, t+dt]$. The resting potential V_0 is set at 0, without loss of generality. The explicit dependence of the random term with $\sqrt{dt}$ is meant to ensure that the cumulative variance over a period T caused by the random term is independent of the size of the time step dt used to resolve the process:

$$\begin{aligned}\mathrm{Var}[\int dV] &= \mathrm{Var}[\sum_{i=1}^{n=T/dt} \sigma\sqrt{dt}\,\mathcal{N}_{t_i}^{t_i+dt}(0,1)] \\ &= \sigma^2 dt\ \mathrm{Var}[\sum_{i=1}^{n=T/dt} \mathcal{N}(0,1)] \\ &= \sigma^2 T.\end{aligned}$$

The master equation can be converted into the Fokker Planck equation over the probability density function (pdf) of the membrane potential $\mathcal{P}(V)$ [102]

$$\frac{\partial \mathcal{P}}{\partial t} = \frac{\partial\left(\left(\frac{V}{\tau} - \mu\right)\mathcal{P}\right)}{\partial V} + \frac{\sigma^2}{2}\frac{\partial^2 \mathcal{P}}{\partial V^2}. \tag{A.5}$$

At the steady state, the $\mathcal{P}$ function does not vary with time and the left side of Eq. A.5 can be set to 0. This results in an ordinary homogeneous second-order linear differential equation, to which the solution follows the general form

$$\mathcal{P}(V) = e^{-\left(\frac{V-\mu\tau}{\sigma\sqrt{\tau}}\right)^2}\left(K_1 + K_2\,\mathrm{erfi}\left(\frac{V-\mu\tau}{\sigma\sqrt{\tau}}\right)\right),$$

where erfi is the imaginary error function ($\mathrm{erfi}(x) = 2\sqrt{\pi}^{-1}\int_0^x e^{V^2}dV$) and K_1 and K_2 are

integration constants. The values of the two constants are determined by imposing the characteristic boundary conditions of LIF neurons. Since neurons that reach the spiking threshold Θ are automatically set back at the reset potential V_R, these two points act as sink and source for the probability density function $\mathcal{P}(V)$. Because of this, the function is not differentiable in V_R and it must be defined separately on two domains, with the integration constants K_1^-, K_2^- for $V < V_R$ and K_1^+, K_2^+ for $V_R < V < \Theta$. The flux φ of the pdf is given by

$$\begin{aligned}\frac{\partial \varphi}{\partial V} &= -\frac{\partial \mathcal{P}}{\partial t} \\ \varphi &= \mathcal{P}\left(\mu - \frac{V}{\tau}\right) - \frac{\sigma^2}{2}\frac{\partial \mathcal{P}}{\partial V}.\end{aligned}$$

The boundary conditions of the LIF neuron are then

- $\varphi(\Theta) = \nu$; By definition, the flux that reaches the firing threshold is the firing rate.
- $\mathcal{P}(\Theta) = 0$; The membrane potential cannot exceed the firing threshold, and as soon as it reaches it, the neuron fires and is reset.
- $\varphi(V_R^+) - \varphi(V_R^-) = \nu$; The flux that is sunk at Θ is reinjected at V_R, leading to a discontinuity in the flux profile.
- $\mathcal{P}(V_R^+) = \mathcal{P}(V_R^-)$; The probability density function is continuous at the junction of its two subdomains. Because of the Gaussian white noise, the pdf cannot be discontinuous.
- $\int_{-\infty}^{\Theta} \mathcal{P} dV = 1$; $\mathcal{P}$ is a probability density function and its integral must be 1.

Using the first four boundary conditions, we can determine the four constants

$$\begin{aligned}K_2^+ =& -\frac{\nu\sqrt{\pi\tau}}{\sigma} \\ K_1^+ =& -K_2^+ \operatorname{erfi}\left(\frac{\Theta - \mu\tau}{\sigma\sqrt{\tau}}\right) \\ K_2^- =& 0 \\ K_1^- =& K_1^+ + \left(K_2^+ - K_2^-\right)\operatorname{erfi}\left(\frac{V_R - \mu\tau}{\sigma\sqrt{\tau}}\right),\end{aligned}$$

which leads to the profile

$$\mathcal{P}(V) = \frac{\nu\sqrt{\pi\tau}}{\sigma} e^{-\left(\frac{V-\mu\tau}{\sigma\sqrt{\tau}}\right)^2}\left(\operatorname{erfi}\left(\frac{\Theta - \mu\tau}{\sigma\sqrt{\tau}}\right) - \operatorname{erfi}\left(\frac{\max(V, V_R) - \mu\tau}{\sigma\sqrt{\tau}}\right)\right). \tag{A.6}$$

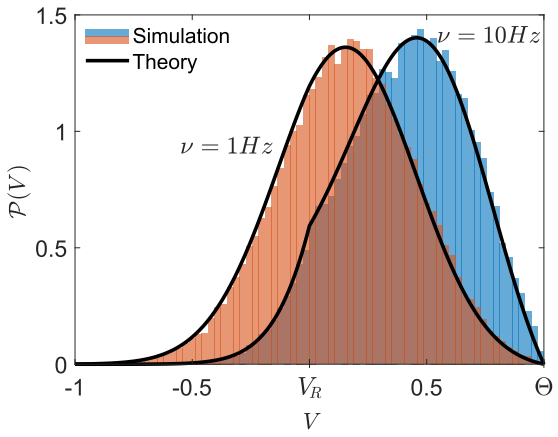


Figure A.2.: **Steady state Probability density function of the membrane potential V of a LIF neuron.** The theoretical distribution (Eq. A.6) matches the distribution obtained through simulation at both low and high firing rates. For high firing rates, the cusp at $V = V_R$ is visible, where the distribution is continuous but non-differentiable.

This profile matches the distribution of membrane potentials recorded in simulations of LIF neurons (Fig. A.2). Finally, the last boundary condition is used to determine the value of the steady state firing rate such that the integral of the pdf is 1. With the change of variable $z = \frac{V-\mu\tau}{\sigma\sqrt{\tau}}$, this results in the Ricciardi Φ equation (Eq. 1.3) [145, 32, 24]:

$$\begin{aligned}\int_{-\infty}^{\infty} \mathcal{P}(V)dV =& 1\\ \nu\tau\sqrt{\pi}\left(\frac{\sqrt{\pi}}{2}\left(\operatorname{erfi}\left(z_\Theta\right) - \operatorname{erfi}\left(z_R\right)\right) + \int_{z_R}^{z_\Theta} e^{z^2}\operatorname{erf} z dz\right) =& 1\\ \tau\sqrt{\pi}\int_{z_R}^{z_\Theta} e^{z^2}\left(1+\operatorname{erf} z\right)dz =& \nu^{-1}.\end{aligned}$$

A.5. LIF approximation at high rate

Although the activation function of LIF neurons given by the Ricciardi equation (Eq. 1.3) is distinctly nonlinear for low input, the high-input/high-rates regime approaches a linear limit (see Fig. 2.1, phase III). In the limit of strong input $\lim_{\mu\to\infty}$, the boundaries $\frac{\Theta-\mu\tau}{\sigma\sqrt{\tau}}$ and $\frac{V_R-\mu\tau}{\sigma\sqrt{\tau}}$ of the integral in Eq. 1.3 tend to $-\infty$. In that limit, using L'Hôpital's rule, we can replace the integrand with a simpler expression,

$$\lim_{z\to-\infty} e^{z^2}(1+\operatorname{erf}(z)) = \frac{-1}{z\sqrt{\pi}}.$$

For high inputs, the Φ equation becomes then

$$\begin{aligned}\nu &= \left(\tau\sqrt{\pi}\int_{\frac{V_R-\mu\tau}{\sigma\sqrt{\tau}}}^{\frac{\Theta-\mu\tau}{\sigma\sqrt{\tau}}} \frac{-1}{z\sqrt{\pi}dz}\right)^{-1} \\ &= \left(\tau\ln\left(1+\frac{\Theta-V_R}{\mu\tau-\Theta}\right)\right)^{-1}. \end{aligned} \tag{A.7}$$

Incidentally, this limit is also the analytical solution of the LIF equation for constant noiseless input (Eq. 1.2).

Although this expression still shows a highly nonlinear relation between input and firing rate, it is approximately linear in the limit of large input μ. Using the Taylor expansion of the $\ln(1+x)$ term in Eq. A.7 we get

$$\nu = \frac{\mu}{\Theta-V_R} - \frac{\Theta+V_R}{2\tau(\Theta-V_R)} + \mathcal{O}\left(\frac{1}{\mu}\right).$$

This shows that the activation function of LIF neurons becomes linear in the high-input/high-rates regime. In that case, the SSN approximation does not hold, and network activity cannot achieve complex operational regimes permitted by the increasing f'_E and f'_I functions (Eq. 3.3). In that limit, the populations are either silenced or in the linear balanced regime (Fig. 3.7).

Nonetheless, if the neuron has a refractory period Δt_{ref}, the FI curve will saturate (see Fig. 2.1, phase IV) and will approach the limit $\nu = \Delta t_{ref}^{-1}$ as

$$\nu = \frac{2\tau\mu-\Theta-V_R}{2\tau(\Theta-V_R)+(2\tau\mu-\Theta-V_R)\Delta t_{ref}}.$$

A.6. Edge cases of the characteristic function

In the process of deriving the characteristic $\mathcal{F}$ function (Eq. 2.4), two assumptions are made on the connectivity:

- The steady state condition of one population carries information about the state of the other population: $J_{EI} \neq 0$.
- The J matrix is invertible : $\det J \neq 0$

Here, I demonstrate that the steady states of the 2D network are always given by zero-crossings of the characteristic function, even in the edge cases where these working assumptions are broken.

► **$J_{EI} = 0$**

By substituting $J_{EI} = 0$ in Eq. 2.3, the first line becomes

$$J_{EE} f_E - \mu_E + \mu_{ext} = 0,$$

which is exactly the $\mathcal{F}$ function for $J_{EI} = 0$. The steady state of the *E* population is then independent of the state of the *I* population.

► **$\det J = 0$**

The case where the J matrix is not invertible corresponds to a situation where the connection weights to the *E* and *I* populations are linearly dependent. So the J matrix can be rewritten as

$$J = \begin{bmatrix} J_E & -J_I \\ \kappa J_E & -\kappa J_I \end{bmatrix},$$

where $\kappa = \frac{J_{IE}}{J_{EE}} = \frac{J_{II}}{J_{EI}}$ is the proportionality factor. The equation Eq. 2.3 becomes

$$\begin{cases} \mu_E = (J_E \nu_E - J_I \nu_I) + \mu_{ext} \\ \mu_I = \kappa\,(J_E \nu_E - J_I \nu_I) + r\mu_{ext}. \end{cases}$$

The recurrent input term in the two equations can be equated

$$\begin{cases} J_E f_E(\mu_E) - J_I f_I(\mu_I) - \mu_E + \mu_{ext} = 0 \\ \mu_I = \kappa \mu_E + \mu_X (r - \kappa), \end{cases}$$

which is identical to the characteristic function $\mathcal{F}$ (Eq. 2.4) with $\det J = 0$ and $\kappa = \frac{J_{II}}{J_{EI}}$.

Overall, the zero-crossings of the characteristic function $\mathcal{F}$ can be used to identify steady states of the rate model for any connectivity matrix J.

A.7. Experimental evidence for biologically plausible neural network size

In the primary visual cortex of mice, it is particularly challenging to define a local network size since the cortical map is unstructured, meaning that neurons that share the same receptive field do not co-localize [184]. However, this does not mean that the concept of homogeneous network cannot be applied to mouse V1, as neurons that share the same receptive field preferentially connect together [94]. Since local networks within mouse V1 cannot be delimited based on spatial anatomy, we refer to other brain regions with anatomically defined networks to determine a reference point.

In mouse somatosensory cortex, distinct neuroanatomical structures known as barrels each receive the sensory input from a corresponding whisker [189]. These structures are perfect candidates to define the typical scale of a homogeneous network. Their diameter ranges from 100 to 400 μm, with a thickness in layer 4 of 100 μm [189]. Using the neuronal density observed in [120], the number of neurons in these structures ranges from 140 to 2200. In the rat barrel cortex, the number of neurons in each layer of multiple projection columns has specifically been counted [114]. It is on the order of $N = 4000$ in layer 4 and $N = 6000$ in layer 2/3.

In primates, the primary visual cortex of macaques has been studied extensively. Unlike rodents, the cortical organization of macaque V1 shows a columnar structure, both for eye dominance and orientation preference [80]. Within a range of orientation preference of 10°, such columns are slab-shaped, with a size of $30\,\mu\mathrm{m}$ by 0.5 to $1\,\mathrm{mm}$ [124]. With the thickness of layer 2 and layer 3 being respectively $225\,\mu\mathrm{m}$ and $310\,\mu\mathrm{m}$ [65], and a cell density of $1.3 \times 10^5\,\mathrm{neurons/mm^3}$ [128], we can deduce that the number of neurons in layer 2/3 in such columns is in the range of 1000 to 2000.

In summary, it appears that across species and cortical regions, we can define functional networks with sizes ranging from hundreds to a few thousands of neurons. In this work, we choose a network size of 4000 because it corresponds to the value reported in [114], which is the only study where the neurons in a cortical column were directly counted. In this context, we can use anatomical analyses of neurons in the same brain region to determine the fraction of excitatory and inhibitory neurons in a network [142, 111], which leads to a E/I ratio of 3.5 in layer 2/3. This corresponds to 3110 *E* neurons and 890 *I* neurons, which we round to 3000 and 1000 respectively.

Finally, we can verify that the considered network size (i.e., 4000 neurons) is plausible for mouse V1 since we could not use a spatial anatomy feature to define a network. We use the reported probability profile (p) for E→E connections as a function of distance for mouse V1 layer 2/3, Fig 4B in [161] (near 20% for nearby pairs, going down to 0% for neuron pairs 150μm apart). Following this observation, the number of excitatory synapses to an excitatory neuron can be estimated through a 3D spatial integration of this connection probability profile: $C_{EE} = \int\int\int p_{EE}(\sqrt{x^2+y^2+z^2})\eta_E dxdydz$, where η_E is the density of excitatory neurons. Over an infinite 3D space, we obtain the total number of E→E connections: $C_{EE} = 8.7\times10^5\eta_E$. Using a neuronal density of 1.64×10^{-4} neurons/μm^3 [90], and assuming a E/I ratio of 3.5, we obtain C_{EE} ≈111 connections. The result of this rough calculation is in the same order of magnitude as the 195 connections we obtain with 4000 neurons and a probability of connection of 6.5% (as described in [6]), which suggests that this network size is a valid approximation for mouse V1 as well. It should be noted that the network we define does not constitute a single block of cortex due to the salt-and-pepper organization of this brain region, but consists instead of distant neurons which receive the same external input and are homogeneously connected within the network.

A.8. Disinhibition

We demonstrated in section 3.2 how the Paradoxical Effect is a signature of the ISN in E-I networks. It should be noted however that in networks comprising more than two populations, other mechanisms can be responsible for a decrease of overall inhibitory activity upon stimulation of the I population. This occurs in the case of disinhibition, where one inhibitory population is suppressed by another [152]. This mechanism has been observed in cortical circuits where different inhibitory cell types have preferential targeting. The Pyramidal neurons are primarily inhibited by PV neurons, whereas PV neurons are themselves inhibited by other interneuron types such as VIP or SST [137]. In this context, the stimulation of SST interneurons can lead to the suppression of the PV neurons and, in turn, the increase of pyramidal activity. An experimenter unaware of the structured interactions between inhibitory subtypes can erroneously interpret the decrease of overall I activity after stimulation of the SST neurons as the Paradoxical Effect and conclude that the network is in the ISN.

Nevertheless, when considering the different interneuron types as separate populations, it becomes clearer that this phenomenon is not a marker of the ISN. While we can observe $\frac{d(\nu_{PV+SST})}{d(\mu_{ext\ PV+SST})} < 0$, we cannot have $\frac{d(\nu_{PV})}{d(\mu_{ext\ PV})} < 0$ or $\frac{d(\nu_{SST})}{d(\mu_{ext\ SST})} < 0$ without being in the inhibition stabilized regime, $f'_E J_{EE} > 1$. Furthermore, disinhibition cases can lead to an increased excitatory activity, which is incompatible with the Paradoxical Effect since the paradoxical decrease of I activity is due to the reduced recurrent excitation (Fig. A.3).
In the considered 3-population system, we get:

$$\begin{bmatrix} \delta\nu_E \\ \delta\nu_{I_1} \\ \delta\nu_{I_2} \end{bmatrix} = \begin{bmatrix} 1 - f'_E J_{EE} & f'_E J_{EI_1} & f'_E J_{EI_2} \\ -f'_{I_1} J_{I_1E} & 1 + f'_{I_1} J_{I_1I_1} & f'_{I_1} J_{I_1I_2} \\ -f'_{I_2} J_{I_2E} & f'_{I_2} J_{I_2I_1} & 1 + f'_{I_2} J_{I_2I_2} \end{bmatrix}^{-1} \begin{bmatrix} f'_E \delta\mu_{extE} \\ f'_{I_1} \delta\mu_{extI_1} \\ f'_{I_2} \delta\mu_{extI_2} \end{bmatrix}.$$

Here again, The determinant $|-T\mathcal{M}|$ is positive if the fixed point is stable [115]. The paradoxical effect requirement on the inhibitory subpopulation reduces to

$$\begin{aligned} \frac{d\nu_{I_1}}{d\mu_{extI_1}} &< 0 \\ (1 - f'_E J_{EE})\left(1 + f'_{I_2} J_{I_2I_2}\right) &< -f'_E f'_{I_2} J_{EI_2} J_{I_2E} \\ f'_E J_{EE} &> 1 + \frac{f'_E f'_{I_2} J_{EI_2} J_{I_2E}}{1 + f'_{I_2} J_{I_2I_2}}. \end{aligned}$$

As I_1 and I_2 play strictly symmetrical roles in this system, the results are equivalent when probing for the paradoxical effect on the population I_2. This condition is more restrictive than

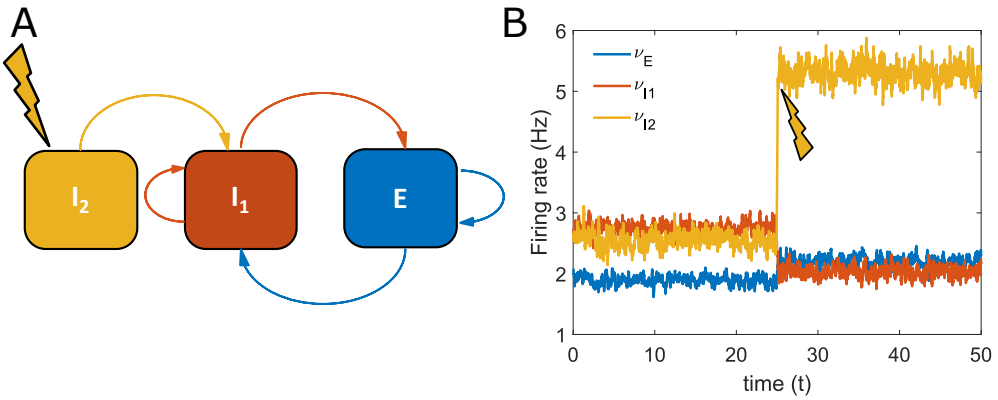


Figure A.3.: **Disinhibition can lead to an apparent Paradoxical Effect without the network having to be in the ISN. A:** Schematic representation of a 3-population network which supports disinhibition. The population I_2 mainly projects to I_1. The effect of its stimulation (represented by the lightning symbol) is a feedforward inhibition of the population I_1. **B:** The stimulation of the population I_2 leads to a decrease of activity of I_1. If the inhibitory population is considered as a whole (I=I_1+I_2), this phenomenon can be mistaken for a paradoxical effect. The fact that the rates of the E populations are driven up by the stimulation illustrates that the suppression of I_1 is not caused by a reduction in recurrent input and is therefore not a signature of the ISN. Figure adapted from [152].

the inhibition stabilization condition since the second term of the inequality is always positive with I_1 and I_2 being inhibitory neurons. In the situation where multiple interneuron populations are considered, the Paradoxical Effect is sufficient but not necessary evidence of the ISN state of the system.

On the other hand, if the two subpopulations I_1 and I_2 are considered together as a single population I, such that $\nu_I = q\nu_{I_1} + (1-q)\,\nu_{I_2}$ where q and $(1-q)$ are the fractions of neurons in each subpopulation. The experiment where a decrease of overall inhibitory activity is observed after stimulation of one subpopulation becomes

$$\begin{aligned} \frac{d\nu_I}{d\mu_{extI_1}} &< 0 \\ q\frac{d\nu_{I_1}}{d\mu_{extI_1}} + (1-q)\frac{d\nu_{I_2}}{d\mu_{extI_1}} &< 0 \\ f'_E J_{EE} &> 1 + f'_E f'_{I_2} J_{I_2E} \frac{qJ_{EI_2} - (1-q)J_{EI_1}}{q\left(1 + f'_{I_2} J_{I_2I_2}\right) - (1-q) f'_{I_2} J_{I_2I_1}}. \end{aligned}$$

Here, the second term can be negative, so the ISN is not a necessary condition for the decrease of overall I activity after stimulation of one inhibitory subpopulation. For instance, if we assume that the I_1 population plays the role of the SST neurons, its strong suppression of the PV subpopulation but not of Pyramidal neurons will lead to a negative second term. This mechanism underlies the disinhibition phenomenon.

A.9. Effect of $\mathbf{J_{EE}}$ on the occurrence of the ISN

Networks are inhibition-stabilized for fixed points that satisfy Eq. 3.2. Therefore, it seems that any network can be in the ISN state, granted that J_{EE} is sufficiently strong. Here, I investigate whether there is any counterexample to this. Are there network configurations that cannot enter the ISN, no matter how high J_{EE} is?

From the SSN equation, by imposing that both firing rates are non-zero, the $(.)_+$ operator can be removed and the external input can be isolated to obtain

$$\begin{cases} \left(\frac{\nu_E}{a_E}\right)^{\frac{1}{n_E}} & -\left(J_{EE}\nu_E - J_{EI}\nu_I - b_E\right) = \mu_{ext} \\ \left(\frac{\nu_I}{a_I}\right)^{\frac{1}{\nu_I}} & -\left(J_{IE}\nu_E - J_{II}\nu_I - b_I\right) = r\mu_{ext}. \end{cases}$$

These equations can be combined to get the crossing of two functions ζ_E and ζ_I

$$\begin{aligned} \zeta_E(\nu_E) &= r\left(\frac{\nu_E}{a_E}\right)^{\frac{1}{n_E}} + \left(J_{IE} - rJ_{EE}\right)\nu_E + rb_E - b_I \\ \zeta_I(\nu_I) &= \left(\frac{\nu_I}{a_I}\right)^{\frac{1}{n_I}} + \left(J_{II} - rJ_{EI}\right)\nu_I. \end{aligned} \tag{A.8}$$

If one population has zero firing rate, it results in an inequality imposing that its total input is lower than its firing threshold b_X.

A steady state in the SSN is therefore one of four situations:

- $\zeta_E(\nu_E) = \zeta_I(\nu_I)$, with $\nu_E, \nu_I > 0$
- $\zeta_I(\nu_I) < \zeta_E(0)$, with $\nu_E = 0$
- $\zeta_E(\nu_E) < \zeta_I(0)$, with $\nu_I = 0$
- $\nu_E = \nu_I = 0$

Since J_{EE} only affects ζ_E, let us first examine how it behaves and how the ISN threshold is affected by changes in J_{EE}. As J_{EE} grows, the ζ_E curve goes from monotonically increasing to having a peak and then decreasing (Fig. A.4.A). Furthermore, the ISN threshold moves to lower values of ν_E as J_{EE} increases. Because of this, any value of ζ_E ranging from $-\infty$ to

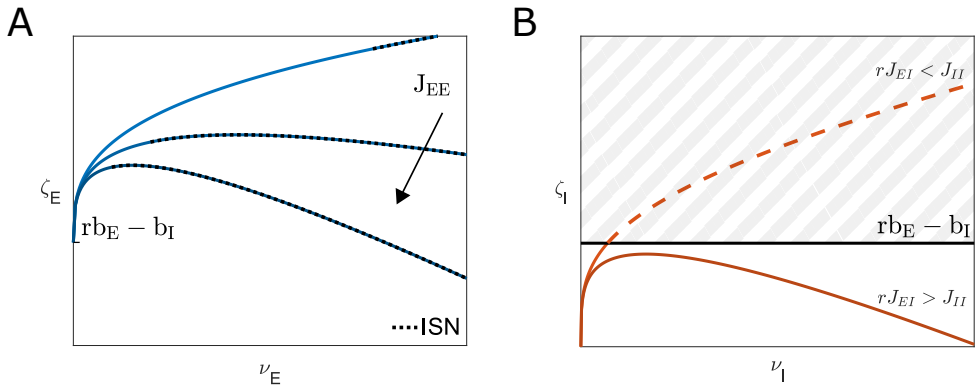


Figure A.4.: **A supersaturating network such that the *E* population is consistently suppressed cannot enter the ISN no matter the strength of the recurrent strength J_{EE}.** A network that can never be inhibition-stabilized either has no cross-connection between the E and I populations or it does not have a fixed point at the ISN threshold, no matter how strong J_{EE} is. The range of steady states accessible to a network is given by the analysis of the ζ_E and ζ_I functions (Eq. A.8). **A:** With J_{EE} as a free parameter, any value of ζ_E is accessible and any E firing rate $\nu_E > 0$ can be the onset of the ISN. To prevent the ISN no matter the value of J_{EE}, the network must not admit a stable steady state with $\nu_E > 0$. **B:** If the function ζ_I always remains below the initial point of ζ_E, $\nu_E = 0$ will always be a stable steady state, and any other steady state where ν_E is not silent is unstable. If the network does not support supersaturation ($rJ_{EI} > J_{II}$) or if the maximum of ζ_I is above the initial point of ζ_E, it will cross the domain where the ISN can be achieved (hatched area).

∞ can be made to support the ISN by tuning J_{EE}. Therefore, regardless of the parameters of ζ_I, tuning J_{EE} is guaranteed to lead to a crossing between the two curves such that $\nu_E > (a_E n_E^{n_E} J_{EE}^{n_E})^{-\frac{1}{n_E - 1}}$, the ISN threshold. In order to ensure that no J_{EE} value corresponds to the ISN, we must ensure this inevitable crossing is unstable.

A network is in the ISN if the derivative of the E activation function is between the ISN threshold (Eq. 3.1) and the instability threshold (Eq. 2.9), while satisfying the dynamic stability requirement on the trace of $\mathcal{M}$ (Eq. 2.10)

$$\frac{1}{J_{EE}} \leq f'_E < \min\left(\frac{1 + f'_I J_{II}}{J_{EE} - f'_I \det J}, \frac{1}{J_{EE}}\left(1 + \frac{\tau_{P_E}}{\tau_{P_I}}(1 + f'_I J_{II})\right)\right).$$

Here already, we see that one trivial way to prevent inhibition stabilization is to close the stability window between these values, so the ISN onset is also the onset of instability. This can occur if

- J_{EI} or J_{IE} is 0. In that case, the inhibitory feedback loop is broken and recurrent inhibition cannot stabilize the E subnetwork.

- $\frac{\tau_{p_E}}{\tau_{p_I}} = 0$. In that case, the recurrent excitation is infinitely faster than recurrent inhibition and runaway dynamics can never be tempered.

- $f'_I = 0$. In that case, the *I* population is frozen and cannot adapt dynamically to stabilize the *E* subnetwork.

These three extreme cases prevent inhibition stabilization regardless of J_{EE} by ensuring the ISN onset is already unstable. Another approach is to ensure that the ζ_E and ζ_I curves cannot cross at the ISN onset and that any further crossing will always be unstable. Since the ISN threshold occurs before the maximum of ζ_E (when this maximum exists and ζ_E is not monotonically increasing), the value of ζ_E at the ISN threshold is guaranteed to be larger than $\zeta_E(0) = rb_E - b_I$, regardless of J_{EE}. The curve ζ_I must then remain below $rb_E - b_I$, which is the limit of the ISN threshold for $J_{EE} \to \infty$. This can happen if $rb_E - b_I > 0$ and the network supports supersaturation ($rJ_{EI} > J_{II}$, see Fig. A.4.B).

For networks of which the whole curve ζ_I is lower than $rb_E - b_I$, solutions such that $\nu_E = 0$ are always supported. If $rJ_{EE} > J_{IE}$, the ζ_E curve will eventually decrease and lead to the occurrence of another solution when ζ_E reaches the maximum of ζ_I, corresponding to another fixed point. Let us consider this second crossing $\nu_E^*, \nu_I^* > 0$ such that $\zeta_E(\nu_E^*) = \zeta_I(\nu_I^*) = rb_E - b_I$, while $\zeta_I(\nu_I^*)$ is the maximum of ζ_I. We find

$$\nu_E^* = \left(\frac{1}{r} a_E^{\frac{1}{n_E}} \left(rJ_{EE} - J_{IE}\right)\right)^{\frac{n_E}{1-n_E}}$$
$$\nu_I^* = \left(n_I a_I^{\frac{1}{n_I}} \left(rJ_{EI} - J_{II}\right)\right)^{\frac{n_I}{1-n_I}},$$

which leads to

$$f_E'^* = \frac{r\, n_E}{rJ_{EE} - J_{IE}}$$
$$f_I'^* = \frac{1}{rJ_{EI} - J_{II}}.$$

Using these values, the static stability condition (Eq. 2.9) becomes

$$|T\mathcal{M}|^* = \frac{1 - n_E}{1 - \frac{J_{II}}{rJ_{EI}}},$$

which is necessarily negative for supersaturating networks with supralinear activation functions. Therefore, we find that high values of J_{EE} lead to the apparition of more solutions, but they are always unstable. That is because the solution ν_E^*, ν_I^* corresponding to the minimum non-zero ν_E solution is unstable and increasing J_{EE} will only drive the network further into instability.

Therefore, for the considered supersaturating networks such that ζ_I does not exceed $\zeta_E(0)$, the network cannot enter an inhibition stabilized regime, no matter the value of J_{EE}. In such networks, the *E* activity is always suppressed and there is no stable steady state with $\nu_E > 0$. If J_{EE} becomes large, another unstable steady state will appear. These systems cannot be in the ISN regardless of J_{EE} because the *E* activity is either entirely silent or cannot be stabilized by inhibition.

For values of b_E and b_I such that $rb_E - b_I < 0$, the network can always enter an inhibition-stabilized state by tuning J_{EE}. Interestingly, this is always the case with the values of b_E and b_I we obtained from fitting the FI curve (Table 2.5) because of the experimentally reported neuronal parameters for τ_E and τ_I [6] lead to $b_E < 0 < b_I$. Similarly, in networks which do not support supersaturation ($J_{II} > rJ_{EI}$), ζ_E always crosses ζ_I, even for low values of J_{EE}.

A.10. Connectivity parameters

	J_{EE}	J_{EI}	J_{IE}	J_{II}	Remarks
Fig. 2.3.B Fig. 4.4 Fig. A.9.A Fig. A.6	0.672	13.2	23.7	11.8	Based on layer 2/3 mouse V1
Fig. 3.1	4.75	0	0	0	
Fig. 3.2	4.75	13.2	23.7	11.8	
Fig. 3.3 Fig. 3.9 Fig. 3.10.C Fig. 3.11.A,D Fig. 4.5 Fig. A.9.B	2	12	6	1	
Fig. 3.4 Fig. 3.8.A Fig. 3.10.D Fig. 3.11.C,F Fig. 3.12.B Fig. A.7 Fig. A.8	5	10	7	11	
Fig. 3.5.A	2	3	2	1.5	
Fig. 3.5.B	10	20	9	5	
Fig. 3.6	3.75	3	3	3.75	$r = 3$
Fig. 3.7.A-C	2	3	3	2	$r = 0.5, 1, 2$
Fig. 3.7.D-F	3	2	2	3	$r = 0.5, 1, 2$
Fig. 3.8.B Fig. 3.11.B,E	1.25	1	1	1.25	$r = 3$
Fig. 3.12.A	1	2	4	4	
Fig. A.3	1	1.2	2	1.5	$J_{I_2I_1} = \frac{5}{3}, r = \frac{10}{3}$
Fig. 4.7	1.5	2.5	2	0.5	$\lambda_E = \lambda_I = \frac{\pi}{10}$ $\lambda_{ext} = \frac{\pi}{5}$
Fig. 4.8	57	140	140	170	$\lambda_E = \lambda_I = \frac{\pi}{5}$ $\lambda_{ext} = 1.6 \times \frac{\pi}{5}$

Table A.2.: Network connectivity parameters used in all figures. The connection strengths at the population level J_{XY} used in mean-field solutions are given in mV. In cases where a figure portrays the same network at different sizes, the presented values of J_{XY} correspond to the default size $N = 4000$.

A.11. Additional figures

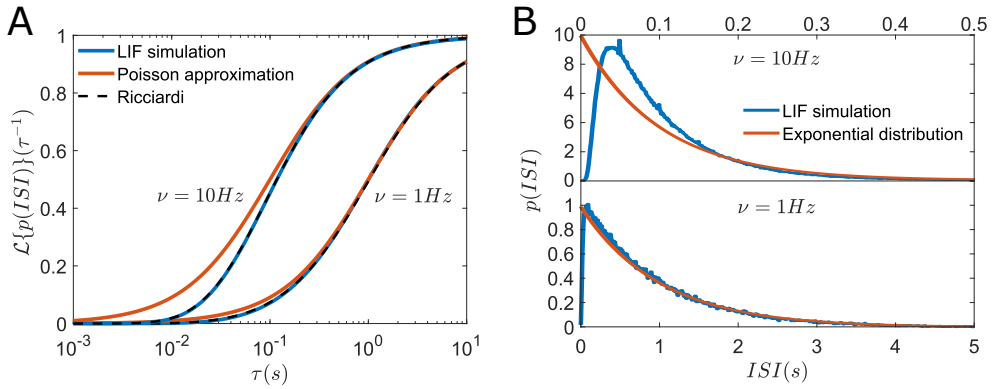


Figure A.5.: **Poisson approximation of the LIF spiking process.** The spiking statistics of LIF neurons fed with white noise can generally be well approximated with a Poisson process. However, this approximation leads to overestimating the prevalence of the shortest ISI events. **A:** The measured Laplace transform of the ISI probability density function is in agreement with the theoretical prediction based on the Ricciardi analysis [32, 118]. For high frequency in the Laplace domain, the equivalent Poisson process deviates from the simulation. **B:** In the time domain, the ISI distribution of LIF neurons can be approximated with an Exponential distribution, characteristic of Poisson processes. The main difference occurs for the near-zero ISIs, which are rare events in LIF neurons.

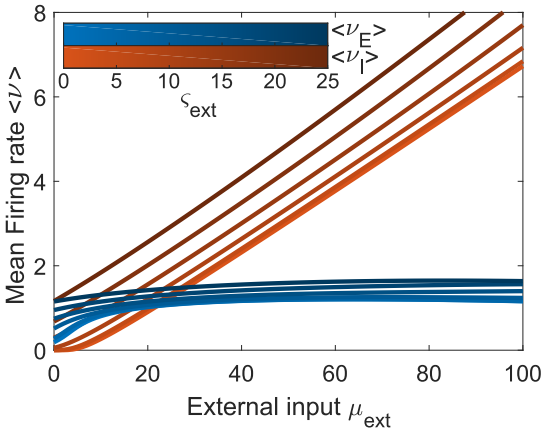


Figure A.6.: **Effect of an inhomogeneous external input on network activity.** The same network as shown in Fig. 2.3, with increasing variability in the external input. While the effect of the feedforward input variability is unnoticeable between $\varsigma = 0$ and $\varsigma = 5\,\mathrm{mV/s}$, the effect becomes increasingly stronger for higher variability. This is due to the nonlinear effect of ς on the activation function (Fig. 4.2.B) and the positive feedback of the firing rate variance in the input variability.

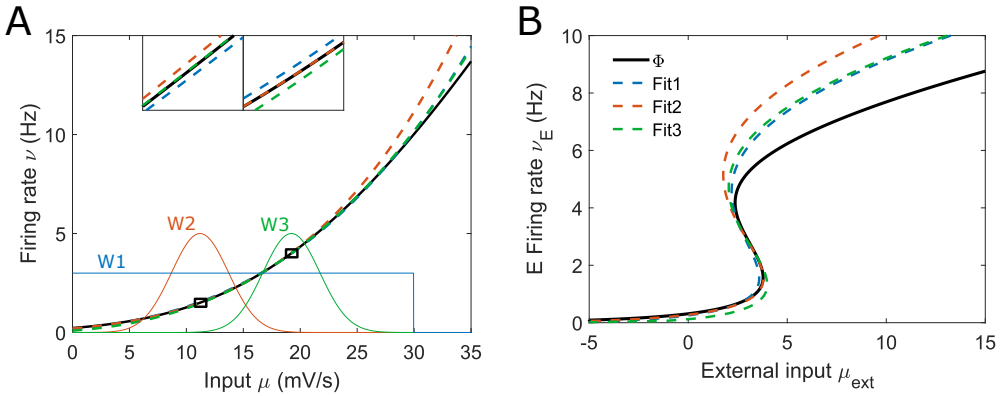


Figure A.7.: **Small errors in the fitted power law are amplified when used in a network. A:** The FI curve of the *E* population is fitted with three different weight functions. W1 is the default weight distribution (see Fig. 2.2) with a flat function until ν_E=10 Hz. W2 and W3 are mainly focused around 1.5 Hz and 4 Hz, respectively. The difference between the three fits is barely visible in the main figure. The insets show zoom-ins of the FI curve around the focal points of W2 and W3 to illustrate how each fit is locally more accurate than the others. **B:** When used to predict the network activity, the difference between the three fits becomes much more noticeable. The network chosen here is the bistable network shown in Fig. 3.4. Fit2 can very accurately match the self-consistency solution based on the Ricciardi Φ function at the bottom bend since it is precisely the level of activity where it matches the FI curve best. Similarly, Fit 3 should accurately predict the upper bend, but the error there is due to the recurrent noise, which is not accounted for in the SSN. The response of the *I* population is not shown here for clarity.

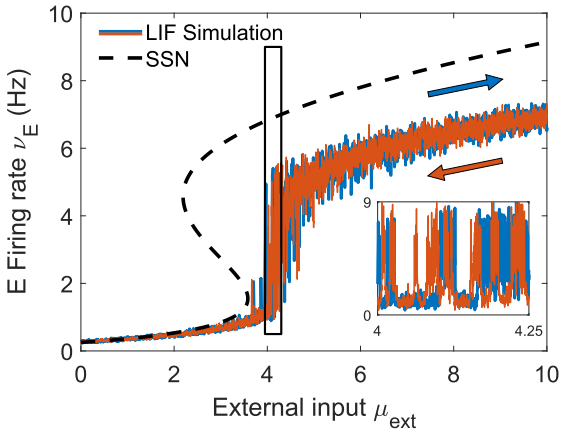


Figure A.8.: **Simulations of spiking LIF neurons require stabilizing synaptic mechanisms to enable bistability.** The same network is simulated as in Fig. 3.4, using a different synaptic transmission mechanism. Here, the simulation is carried out with instantaneous Dirac delta synapses (Eq. A.1), whereas synapses in Fig. 3.4 are distributed in time and have randomly distributed synaptic delays to prevent synchronization (Eq. A.2). Without these stabilizing mechanisms, the network does not support bistability as the forward (blue) and reverse (red) trajectories overlap. The inset shows a zoom-in of the simulated curves near the expected transition point. The spiking network is in a chaotic state where it spontaneously transitions between the low and high activity levels of the bistability. Without additional stabilizing mechanisms, the system cannot maintain a consistent activity level, which causes deviations from the mean-field since the steady state condition cannot be fulfilled. The apparent firing rates reported in the main panel lie between the low and high activity levels due to averaging over a larger time bin.

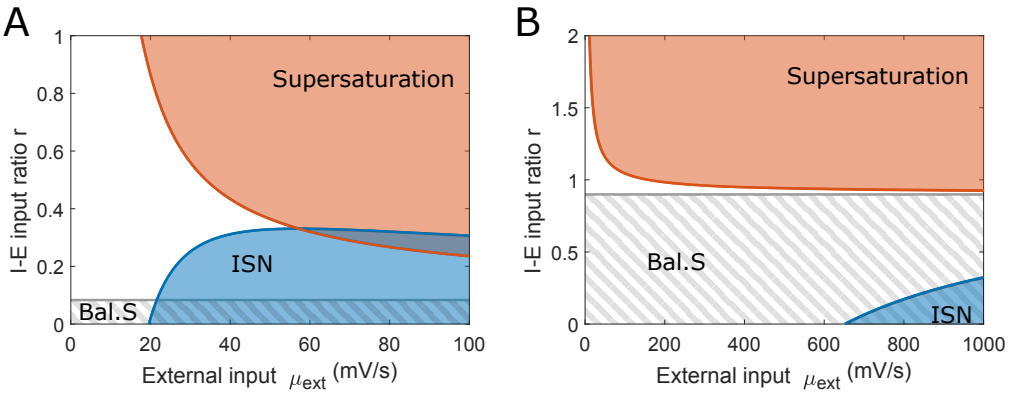


Figure A.9.: **Maps of operational regimes in input space for two additional networks.** Similar to the maps shown in Fig. 3.8. **A:** V1 network with connectivity extracted from the Allen Institute Synaptic Physiology database (see Fig. 2.3). **B:** Supersaturating network presented in Fig. 3.3

A.12. List of symbols

Symbol	Definition	First mention
ν	Firing rate	section 1.2.2
μ	Mean input	section 1.2.2
σ	Temporal input noise	section 1.2.2
V	Membrane potential	section 1.2.2
Θ	LIF spiking threshold potential	section 1.2.2
V_R	LIF reset potential	section 1.2.2
V_0	LIF resting potential	section 1.2.2
τ	LIF membrane time constant	section 1.2.2
Φ	Ricciardi activation function	section 1.2.2
Φ_{sc}	Ricciardi Self-consistent network equation	section 1.2.4
j	Single-synapse connection strength	section 1.2.4
J	Population-wise connection strength	section 1.2.4
$\det J$	Determinant of the J connectivity matrix	section 1.2.4
μ_{ext}	Mean feedforward input	section 1.2.4
σ_{ext}	Feedforward input noise	section 1.2.4
r	Ratio of external input to the *I* and *E* populations	section 1.2.4
C	In-degree, Number of incoming synaptic connections	section 1.2.4
p	Pairwise connection probability	section 1.2.4
N	Number of neurons	section 1.2.4
τ_P	Time constant of the firing-rate dynamics	section 1.2.4
a	Fitted power law parameter, prefactor	section 2.1
b	Fitted power law parameter, firing input threshold	section 2.1
n	Fitted power law parameter, exponent	section 2.1
f	Neuronal activation function	section 2.1
$\mathcal{F}$	Characteristic function	section 2.2
$\mathcal{M}$	SSN Jacobian matrix	section 2.3.1
T	Diagonal matrix of the population time constants	section 2.3.1
$\mathcal{P}$	Probability density function	section 4.1.1
$\mathcal{H}$	Heaviside function	section 4.1.1
ς	Intra-population input variability	section 4.1.1
Ψ	Population average activation function	section 4.1.2
ς_j	Variability in the strength of individual synapses	section 4.1.3
θ	Neuronal feature	section 4.2
η	Neuronal density in feature space	section 4.2
M_{ext}	Feedforward input profile in feature space	section 4.2
P	Pairwise probability of connection in feature space	section 4.2
λ	Lengthscale of recurrent connections in feature space	section 4.2.1
λ_{ext}	Lengthscale of the feedforward input profile in feature space	section 4.2.1
ω	Frequency domain of the feature space θ	section 4.2.2
w	Synaptic plasticity factor	section 4.2.3
τ_s	Synaptic time constant	section A.1

A.13. List of abbreviations

Abbreviation	Definition	First mention
AIS	Axon initial segment	section 1.1.1
PSP	postsynaptic potential	section 1.1.2
ODE	Ordinary differential equation	section 1.2.1
ISI	Inter-spike interval	section 1.2.2
FI curve	Firing rate-input relationship	section 1.2.2
LIF	Leaky-integrate and fire	section 1.2.2
E-I	Excitatory-Inhibitory	section 1.2.4
SSN	Stabilized supralinear network	section 1.2.4
ANN	Artificial Neural Network	section 1.2.5
ISN	Inhibition stabilized network	section 3.1
BF	Balanced factor	section 3.8
Pdf	Probability density function	section 4.1.1
STP	Short-term plasticity	section 4.2.3