

How Models of Canonical Microcircuits Implement Cognitive Functions

DISSERTATION

zur Erlangung des akademischen Grades Doktoringenieur
Dr.-Ing.

vorgelegt der Fakultät für Informatik und Automatisierung
der Technischen Universität Ilmenau

vorgelegt von M.Sc. Tim Kunze
geboren am 03. November 1989 in Meißen

Tag der Einreichung: 04. Januar 2018

Tag der wissenschaftlichen Aussprache: 06. Dezember 2018

Gutachter:

- 1.) Prof. Dr.-Ing. habil. Jens Haueisen
- 2.) Prof. Dr.-Ing. habil. Thomas Knösche
- 3.) Prof. Dr. Pedro Antonio Valdes-Sosa

urn:nbn:de:gbv:ilm1-2018000629

Abstract

Major cognitive functions such as language, memory, and decision-making are thought to rely on distributed networks of a large number of fundamental neural elements, called canonical microcircuits. A mechanistic understanding of the interaction of these canonical microcircuits promises a better comprehension of cognitive functions as well as their potential disorders and corresponding treatment techniques. This thesis establishes a generative modeling framework that rests on canonical microcircuits and employs it to investigate composite mechanisms of cognitive functions.

A generic, biologically plausible neural mass model was derived to parsimoniously represent conceivable architectures of canonical microcircuits. Time domain simulations and bifurcation and stability analyses were used to evaluate the model's capability for basic information processing operations in response to transient stimulations, namely signal flow gating and working memory.

Analysis shows that these basic operations rest upon the bistable activity of a neural population and the selectivity for the stimulus' intensity and temporal consistency and transiency. In the model's state space, this selectivity is marked by the distance of the system's working point to a saddle-node bifurcation and the existence of a Hopf separatrix. The local network balance, in regard of synaptic gains, is shown to modify the model's state space and thus its operational repertoire. Among the investigated architectures, only a three-population model that separates input-receiving and output-emitting excitatory populations exhibits the necessary state space characteristics. It is thus specified as minimal canonical microcircuit. In this three-population model, facilitative feedback information modifies the retention of sensory feedforward information. Consequently, meta-circuits of two hierarchically interacting minimal canonical microcircuits feature a temporal processing history that enables state-dependent processing operations. The

relevance of these composite operations is demonstrated for the neural operations of priming and structure-building. Structure-building, that is the sequential and selective activation of neural circuits, is identified as an essential mechanism in a neural network for syntax parsing. This insight into cognitive processing proves the modeling framework's potential in neurocognitive research.

This thesis substantiates the connectionist notion that higher processing operations emerge from the combination of minimal processing elements and advances the understanding how cognitive functions are implemented in the neocortical matter of the brain.

Keywords: canonical microcircuit, basic information processing operations, neural mass modeling, adaptive neural computations, network balance, disinhibition, bifurcation and stability analysis, structure-building computations, cognitive functions, syntax parsing

Kurzfassung

Kognitive Fähigkeiten wie Sprache, Gedächtnis und Entscheidungsfindung resultieren vermutlich aus der Interaktion vieler fundamentaler neuronaler Elemente, sogenannter kanonischer Schaltkreise. Eine vertiefte Einsicht in das Zusammenwirken dieser kanonischen Schaltkreise verspricht ein besseres Verständnis kognitiver Fähigkeiten, möglicher Funktionsstörungen und Therapieansätze. Die vorliegende Dissertation untersucht ein generatives Modell kanonischer Schaltkreise und erforscht mit dessen Hilfe die Zusammensetzung kognitiver Fähigkeiten aus konstitutiven Mechanismen neuronaler Interaktion.

Es wurde ein biologisch-plausibles neuronales Massenmodell erstellt, das mögliche Architekturen kanonischer Schaltkreise generisch berücksichtigt. Anhand von Simulationen sowie Bifurkations- und Stabilitätsanalysen wurde untersucht, inwiefern das Modell grundlegende Operationen der Informationsverarbeitung, nämlich Selektion und temporäre Speicherung einer transienten Stimulation, unterstützt. Die Untersuchung zeigt, dass eine bistabile Aktivität einer neuronalen Population und die Beurteilung der Salienz des Signals den grundlegenden Operationen zugrunde liegen. Die Beurteilung der Salienz beruht dabei hinsichtlich der Signalstärke auf dem Abstand des Arbeitspunktes zu einer Sattel-Knoten-Bifurkation und hinsichtlich der Signalkonsistenz und -vergänglichkeit auf einer Hopf-Separatrix im Zustandsraum des Systems. Die Netzwerkbalance modifiziert diesen Zustandsraum und damit die Funktionsfähigkeit des Modells. Nur ein Drei-Populationenmodell mit getrennten erregenden Populationen für Signalempfang und -emission weist die notwendigen Bedingungen im Zustandsraum auf und genügt der Definition eines minimalen kanonischen Schaltkreises. In diesem Drei-Populationenmodell erleichtert ein Feedbacksignal die Speicherfähigkeit für sensorische Feedforwardsignale. Dementsprechend weisen hierarchisch interagierende minimale kanonische Schaltkreise ein zeitliches Verarbeitungs-

gedächtnis auf, das zustandsabhängige Verarbeitungsoperationen erlaubt. Die Bedeutung dieser konstitutiven Operationen wird für die neuronalen Operationen Priming und Strukturbildung verdeutlicht. Letztere wurde als wichtiger Mechanismus in einem Netzwerk zur Syntaxanalyse identifiziert und belegt das Potential des Modellansatzes für die neurokognitive Forschung.

Diese Dissertation konkretisiert die konnektionistische Ansicht höhere Verarbeitungsoperationen als Ergebnis der Kombination minimaler Verarbeitungselemente zu verstehen und befördert das Verständnis für die Frage wie kognitive Fähigkeiten im Nervengewebe des Gehirns implementiert sind.

Schlüsselwörter: kanonische Schaltkreise, grundlegende Informationsverarbeitungsoperationen, neuronale Massenmodelle, adaptive neuronale Operationen, Netzwerkbalance, Disinhibition, Bifurkations- und Stabilitätsanalyse, strukturetablierende Operationen, kognitive Funktionen, Syntaxparsing

Preface

The research took place at the Max Planck Institute for Human Cognitive and Brain Sciences (Leipzig, Germany), at St. Vincent's Hospital (Melbourne, Australia), and the Institute of Biomedical Engineering and Informatics of Ilmenau University of Technology (Ilmenau, Germany), under the supervision of Prof. Dr. Jens Haueisen and Dr. Thomas R. Knösche. It was funded by the Max Planck Society and the Deutscher Akademischer Austauschdienst.

The present thesis comprises research from two projects. The first project focuses on the local stimulation-induced information processing of a canonical microcircuit model and its application in a simple syntax parsing model. My collaborators were Dr. Andre Peterson (University of Melbourne, Melbourne, Australia), Prof. Dr. habil. Jens Haueisen (Ilmenau University of Technology, Ilmenau, Germany), and PD Dr. habil. Thomas R. Knösche (Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany). This piece of research is published and can be found in:

Kunze, T., Peterson, A.D.H, Haueisen, J., and Knösche, T.R., 2017. A model of individualized canonical microcircuits supporting cognitive operations. PLoS One, 12(12):e0188003, 2017.

The second project focuses on the interaction of two canonical microcircuits via hierarchical signals and its application in an advanced syntax parsing model. My collaborators were Prof. Dr. habil. Jens Haueisen, and PD Dr. habil. Thomas R. Knösche. This piece of research is submitted for publication:

Kunze, T., Haueisen, J., and Knösche, T.R., (submitted). Emergence of cognitive priming and structure-building from the hierarchical interaction of canonical microcircuit models.

I myself am responsible for the overall production of this piece of work.

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen:

1. PD Dr. habil. Thomas R. Knösche: Interpretation der Analysen, Ausarbeitung der Transformation zweier neuronaler Massen zu einer neuronalen Masse
2. Prof. Dr.-Ing. habil. Jens Haueisen: Interpretation der Analysen
3. Dr. Andre Peterson: Diskussion der Bifurkationsanalyse und Konzept der Netzwerkbalance

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt. Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch angesehen wird und den erfolglosen Abbruch des Promotionsverfahrens zu Folge hat.

Leipzig, 04. Januar 2018

Tim Kunze

Acknowledgement

I would like to express my deepest gratitude to my supervisors and mentors Jens Haueisen and Thomas Knösche. I would like to thank them for their guidance and patience, their wide spread thematic competence, their impulse and drive to promote – be it ideas, projects or students – as well as their organizational skills and their ability to give space without losing focus.

I am very grateful to Andre Peterson and Stefan Lau for their methodological and organizational guidance. This thesis would not exist in this form without the milestones I was able to gain during my time in Melbourne.

It has been a privilege to work at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig. I would like to thank Angela Friederici for this opportunity as well as for insightful discussions, and my colleagues of the CNCF Research Group for the many discussions, advices, and support: Hermann Sonntag, Shih-Cheng (Vincent) Chien, Burkhard Maess, and Konstantin Weise.

Many non-scientific friends from Leipzig, Meißen, Ilmenau, and Marseille unconsciously supported this thesis by providing the rich environment that stimulated me, prevented dwelling in local minimal, and kept me moving by their many opinions and perspectives. Special thanks go to Claudia Herpertz for the hours she spent proofreading and to Maximilian Reich, my steady source of inspiration.

I am also thankful for my parents Angelika and Jörg, my sister Kristin, and my brother Matthias for their understanding, wisdom, and continuous support. It is a privilege to have a home that has unwavering faith in my studies and me.

This thesis would not have been possible without my wonderful partner Annika. I admire and thank her for her delighting spirit, her zest for action and amusement, and the support that she has provided me.

Symbols and Abbreviations

Abbreviation	Meaning
BA	Brodmann area
ECFS	Extreme capsule fiber system
EEG	Electroencephalography
EIN E	Excitatory interneuron cell population
fMRI	Functional magnetic resonance imaging
FOP	Frontal operculum
IIN I	Inhibitory interneuron cell population
ODE	Ordinary differential equations
MEG	Magnetoencephalography
MTG	Middle temporal gyrus
PAC	Primary auditory cortex
PMC	Premotor cortex
P_E	Fraction of neurons that are mapped from EIN to Py'
P_P	Fraction of neurons in Py' (originally located in Py)
Py P	Pyramidal cell population
Py'	Excitatory cell population (Py and EIN)
(a/p) STG	(anterior/posterior) superior temporal gyrus
UF	Uncinate fascicle

Variable	Meaning	Unit
t	Time	s
$V(t)$	Mean membrane potential of a neural mass	mV
$\phi(t)$	Mean firing rate of a neural mass	s^{-1}

Operator	Meaning
*	Convolution
$\theta(t)$	Heaviside function
$D_{e,i}$	Excitatory/inhibitory second order temporal differential operator
$S(\cdot)$	Sigmoidal activation function (logistic function)
$h_{e,i}(t)$	Excitatory/inhibitory synaptic response kernel

Parameter	Meaning	Unit
$H_{e,i}$	Excitatory/inhibitory synaptic gain	mV
$\tau_{e,i}$	Characteristic time constant of an excitatory/inhibitory neural mass	ms
e_0	Half of the highest achievable firing rate	s^{-1}
r	Maximum slope of the sigmoidal activation function	mV^{-1}
v_0	Membrane potential for which half of the maximum firing rate is invoked	mV
b_1	Topological parameter controlling target of extrinsic feedforward input and choice between direct or indirect excitatory local feedback of the pyramidal cells	
b_2	Topological parameter controlling the presence of recurrent local feedback of the inhibitory interneurons	
b_3	Topological parameter controlling presence of extrinsic feedback input at the pyramidal cells	
p_{ff}	Extrinsic feedforward input	s^{-1}
p_{fb}	Extrinsic feedback input	s^{-1}
N_{IP}	Strength of Py-to-IIN connection	
N_{PI}	Strength of IIN-to-Py connection	
N_{II}	Strength of IIN-to-IIN connection	
M_E	Number of neurons in EIN population (gradual mapping)	
M_P	Number of neurons in Py population (gradual mapping)	
β	Number of neurons mapped from EIN to Py' (gradual mapping)	
α	Linear scaling factor (gradual mapping)	
N_{EP}	Connectivity gain representing the average number of neurons in Py that project to a single neuron in EIN	
N_{PE}	Connectivity gain representing the average number of neurons in EIN that project to a single neuron in Py	
$N_{P_P E}$	Connectivity gain representing the average number of neurons in EIN that project to a single neuron in P _P	
$N_{E P_P}$	Connectivity gain representing the average number of neurons in P _P that project to a single neuron in EIN	
N_{PP}	Connectivity gain representing the average number of neurons projecting within Py' (self-connections)	
$N_{P_E P_P}$	Connectivity gain representing the average number of neurons in P _P that project to a single neuron in P _E	
$N_{P_P P_E}$	Connectivity gain representing the average number of neurons in P _E that project to a single neuron in P _P	
$N_{Py'E}$	Connectivity gain representing the average number of neurons in EIN that project to a single neuron in Py'	

Parameter	Meaning	Unit
$N_{Py'P_P}$	Connectivity gain representing the average number of neurons in P_P that project to a single neuron in Py'	s mV
$N_{Py'P_E}$	Connectivity gain representing the average number of neurons in P_E that project to a single neuron in Py'	
c_f	Inter-circuit feedforward connectivity gain (priming meta-circuit)	
c_b	Inter-circuit feedback connectivity gain (priming meta-circuit)	
t_{dur}	Duration of the rectangular stimulation	
u_{th}	Firing threshold	

Signifier and Subscripts	Meaning
A_1 / A_2	Higher/lower level minimal canonical microcircuit in initial prototypical meta-circuit
A_1^* / A_2^*	Higher/lower level minimal canonical microcircuit in prototypical meta-circuit for priming
$\tilde{A}_1 / \tilde{A}_2 / \tilde{A}_3$	Higher/lower level minimal canonical microcircuit in prototypical meta-circuit for structure-building
$p_{fac,in/out}$	Facilitative feedback signal (considered as input or output)
a	Commonly denoting a target
b	Commonly denoting a source
in	Commonly denoting an input
c	Denoting the type of a neural mass
e	Commonly denoting an excitatory parameter
i	Commonly denoting an inhibitory parameter
s	Substitution parameter

Contents

Abstract	i
Kurzfassung	iii
Preface	v
Erklärung	vi
Acknowledgement	vii
Symbols and Abbreviations	viii
1 Introduction	1
1.1 Motivation	1
1.2 Aims and Emphases	3
1.3 Contributions	6
1.4 Structure	8
2 Neural Canonical Microcircuits in Cognitive Information Processing	9
2.1 Neuronal Interaction Embodies Cognitive Information Processing	9
2.2 Linking Cell Assembly Theory with the Concept of Canonical Microcircuits	11
2.3 Structural and Functional Characteristics of Canonical Microcircuits	12
2.4 Analysis of a Minimal Canonical Microcircuit	15
2.5 Interaction in Hierarchic Networks	17
3 Methods	19
3.1 The Canonical Microcircuit Model – Architecture and Parameterization	19

3.1.1	Determination of Potential Minimal Canonical Microcircuit Architectures	19
3.1.2	Gradual Mapping of Excitatory Populations	21
3.1.3	Formularization of the Canonical Microcircuit Model . . .	24
3.1.4	Local Network Balance	28
3.2	Model Simulation and Data Analysis	29
3.2.1	Time Domain Simulations	29
3.2.2	Bifurcation and Stability Analysis	31
4	Basic Operations in the Canonical Microcircuit Model	33
4.1	Introduction	33
4.2	Stimulation-induced Response Behaviors and Basic Operations . .	34
4.3	Network Balance Variation	38
4.4	Variations in Local Topological Feedback	42
4.4.1	Transformation of the Three-population Model to the Two-population Model	42
4.4.2	Direct Excitatory Local Feedback Architecture	44
4.4.3	Introduction of Disinhibition in the Two-population Model	47
4.4.4	Direct Excitatory Local Feedback Architecture with Disinhibition	48
4.5	Response to Extrinsic Feedback Input	51
4.6	Discussion	53
5	State-dependent Processing in Interacting Canonical Microcircuit Models	57
5.1	Introduction	57
5.2	Simultaneous Stimulation Modulates Response Behavior	59
5.3	Definition of Prototypical Meta-circuits	62
5.4	State-dependent Priming	63
5.4.1	Principle Priming Mechanism	63
5.4.2	Priming Mechanism Analysis	64
5.4.3	Analysis Results	66
5.5	State-dependent Structure-building	71
5.6	Discussion	73
6	Syntax Parsing in Networks of Canonical Microcircuit Models	77
6.1	Introduction to Syntax Parsing as Part of Sentence Processing . . .	77

6.1.1	The Cortical Language Circuit	77
6.1.2	Neuronal Representation and Mechanisms of Language Processing	79
6.2	Syntax Parsing Based on Feedforward Conditioning	81
6.2.1	Model Architecture and Mechanism	81
6.2.2	Model Analysis	83
6.2.3	Model Assessment	86
6.3	Advanced Syntax Parsing Based on Hierarchical Conditioning . .	87
6.3.1	Model Architecture and Mechanism	87
6.3.2	Model Analysis	88
6.3.3	Model Assessment	89
6.4	Discussion	90
7	Discussion, Future Research, and Conclusion	95
7.1	Discussion	95
7.2	Future Research	98
7.3	Conclusion	99
A	Appendix	100
A.1	Deriving a Dimensionless Form of the Temporal Differential Operator of an Alpha Function	100
A.2	State Space Modulation due to Variation of Synaptic Gains	103
B	Glossary	107
	References	xiv
	List of Figures	xxv
	List of Tables	xxvii

Chapter 1

Introduction

“The organism of intelligence is the neocortex. It basically solves all these different things we do with the same basic methodology, the same basic algorithm. The neural tissue that implements vision, language, and touch — they all work on the same principle. From the brain’s point of view, there’s nothing really different about vision and touch and hearing. It’s mathematics. This is why we are capable of doing so many things we didn’t evolve to do.”

– Jeff Hawkins, 2017

1.1 Motivation

Cognition, or cognitive functions, embrace all mental acts the healthy brain of an organism can perform including reasoning, memory, communication through a natural language, problem solving, decision-making, emotion, or imagination. Human neuroscience and its derived disciplines strive to characterize the nature and origin of cognitive functions, to search for the causes of related neurological disorders, to improve their individual prophylaxis, diagnosis and treatment. The present thesis aims to contribute to these matters by addressing the question of how cognitive functions are mechanistically implemented in the brain’s neural matter.

Most neuroscientific research projects use experimental studies to investigate the information processing underlying cognitive brain functions. Besides behavioral measures, physiological measurements such as magnetoencephalography (MEG),

electroencephalography (EEG) or functional magnetic resonance imaging (fMRI) are used to relate the brain's quantifiable output to distinct experimental input conditions. An essential part of these experimental studies are theoretic models that capture the many underlying assumptions about neural signal generation and transmission. For example, *generative models* capture formalized¹ hypotheses about the genesis of measurable neural activity – usually as a result of neuronal interaction in the neural matter, i.e. the source space [1]. Moreover, forward models, including *volume conduction models* and *electrode models*, account for the signal transmission from the source space through the diverse head tissues (cerebrospinal fluid, skull, and scalp) to the surface of the head, i.e. the sensor space [2, 3], and consider the influence of sensor dynamics to a measured signal.

Often, experimental studies draw on inverted volume conduction and electrode models in order to derive conclusions about the internal distribution and strength of neural activity within the brain from recorded data. Despite the undisputable success of this approach there are also limitations. First, statements about the distribution or strength of neural activity have a limited explanatory power concerning the actual underlying mechanisms that produce this activity. Although macroscopically observable neural activity can confine cognitive functions to distinct brain areas, the mechanisms of signal generation and cognitive functions remain unclear [4]. Second, each measurement method captures a distinct spatial and temporal scale and records with a distinct resolution. The brain processes information on many spatial and temporal scales simultaneously and while single measurements can only reflect limited aspects of this processing, various other aspects are kept hidden.

Biologically plausible generative models address these points by making statements about how the neural matter generates measurable neural activity by means of multi-level processing mechanisms and how cognitive functions are accomplished. These models include the characterization of involved neural elements, their state variables and parameters, their interaction, and the type of received, transmitted, and emitted physiological signals. Through inversion and fitting to recorded data, generative models detect and quantify both hidden and known parameters as well as mechanisms in support of established neurocognitive brain theories, i.e. concepts explaining the nature of neural interaction that leads to cog-

¹ Here, formalized means that mathematical formalisms are used to reflect physiological processes.

nitive functions. Thus, by describing the generation of measurable physiological signals, generative models are important to explain experimental observations [5]. Moreover, generative models serve in a rather explorative way to substantiate and further develop neurocognitive brain theories. Often, neurocognitive brain theories rest on descriptive symbolic models [6]. In contrast, generative models account for measurable physiological signals and thus facilitate the experimental evidence for neurocognitive brain theories.

Succeeding past scientific disputes on how diverse mental abilities may be implemented in the brain [7], modern neurocognitive brain theories assume that higher cognitive functions emerge from the constructive interaction of a high number of fundamental units in distributed neural networks – a major neurocognitive concept called *connectionism* [8]. Despite a likely consensus among neuroscientists on this general definition of connectionism, its concrete implementation in the neural matter raises many open questions; one being the neurobiological identification of the fundamental units. The theory of *canonical microcircuits* [9–11] postulates the existence of such fundamental units in the neocortex, which can be effectively described by a common basic circuitry. However, it is not free of dispute [12, 13]. Other open questions concern the interaction of the fundamental units in terms of signals and signaling pathways as well as the understanding of purposeful mechanisms arising from their interaction. Generative models can help to formulate and formalize the individual assumptions and hypotheses of diverse research approaches and provide for their comparability. That way the development and analysis of biologically plausible generative models, as described in this thesis, advances the development of neurocognitive brain theories [14].

1.2 Aims and Emphases

This thesis pursues two major aims: first, to formulate and establish a neurobiologically plausible generative modeling framework of interacting canonical microcircuits that supports systematic experiment-based investigations of cognitive functions; second, to investigate (i) the significance of canonical microcircuits as fundamental neuronal units for cognitive processing, (ii) the framework’s potential to elucidate generic organizational principles and mechanisms of the brain’s information processing for the example of syntax parsing during language processing, and (iii) its relation with the established neurocognitive brain theory of neural cell

assemblies. To achieve these goals I aimed to answer the following guiding questions:

- How can a model of a canonical microcircuit be defined and formalized?
- Which local repertoire of basic information processing operations does this canonical microcircuit model support?
- How do local topological variations of canonical microcircuit models and the relationship of excitation and inhibition (i.e., the network balance) affect these basic operations?
- How do canonical microcircuit models interact across hierarchical levels?
- How can neurocognitive operations such as priming and structure-building be mechanistically explained through meta-circuits of interacting canonical microcircuit models?
- How do these neurocognitive operations relate to higher cognitive functions?

Throughout the thesis I combine descriptive time domain simulations with explanatory mappings of the system's state space, including bifurcation diagrams, in order to not only describe *what* processes are occurring but also *why*, thus stressing the underlying mechanisms.

This thesis features three main emphases that are (i) the definition and formalization of a canonical microcircuit model as well as the characterization of its local basic information processing operations in plausible local feedback topologies, (ii) the expansion of the dynamic repertoire and the system's processing capacity for two hierarchically interacting canonical microcircuit models that support adaptive, state-dependent processing mechanisms, and (iii) the utilization of these dynamics to implement syntax parsing in a network of canonical microcircuit models.

The **first emphasis** derives a canonical microcircuit model and examines its local dynamic repertoire. A generic model architecture, reflecting three topologies of interest, is implemented in form of a neural mass model [15–17] that captures the mesoscopic spatial scale of neural populations at which the uniformity of canonical microcircuits is arguably established. Based on transient stimulation-induced response behaviors, the canonical microcircuit model is shown to perform basic information processing operations, namely signal flow gating and working memory. These reflect fundamental operations of information processing networks as

proposed to be weighting, storing, and switching of information [18]. The system's state space is examined for characteristic properties that condition the basic operations. Moreover, it is shown how model configurations, namely local synaptic gains, topological parameters, and the separate application of feedforward and feedback information, modify the system's state space.

The **second emphasis** investigates how interacting canonical microcircuit models that exchange feedforward and feedback information across hierarchical levels embody neural operations. The analysis focuses on the modulatory role of feedback input which facilitates the perception and memorization of feedforward stimuli. It is shown how this state-dependent processing gives rise to a broad set of conceivable adaptive processing mechanisms. I exemplify two of these adaptive mechanisms: (i) *perceptual priming* [19], involving the dynamic shift of the perception threshold and (ii) *structure-building computations*, i.e. the selective establishment of sustained spatiotemporal neural activity patterns. I examine how synaptic gains, stimulation characteristics, and hierarchical connectivity strengths affect these mechanisms. It turns out that state-dependent processing mechanisms effectively condition the progressive activation of canonical microcircuits. Accordingly, I argue that canonical microcircuit might represent nodes of associative networks, also called *operational cell assemblies* [20].

The **third emphasis** applies the gathered findings to a network of canonical microcircuit models that supports syntax parsing, that is syntactic structure-building that involves decoding and temporal storage of syntactic information from a continuous stream of words. Canonical microcircuit models represent phonological word information and syntactic information using a place code [21,22]. Syntax parsing is modeled as sequential and selective activation of these canonical microcircuit models (binding) whose spatiotemporally distributed activity pattern represent the syntax prototype. Comparative simulations of two distinct networks of different complexity demonstrate processing advantages when phonological word information and syntactic information are separately represented. These include multiple representations of single words and the consideration of complex syntax structures. Essentially, networks of canonical microcircuit models implement an input-driven mechanism for functional binding of discrete elements. This enables the generation of infinite sequences out of a limited number of discrete elements. This principle is called *infinite recursion* [23] and is conceivable for many cognitive operations.

1.3 Contributions

The overall contribution of this thesis is the establishment and analysis of a generative modeling framework that mechanistically explains basic information processing operations of canonical microcircuits and composite operations in networks of these circuits. The framework allows examining generic processing principles of the nervous system and its emerging processing capacity, specifically in the cerebral cortex – and represents a prototype for the fusion of neurocomputational and neurocognitive research. Moreover, the model’s multiscale paradigm substantiates connectionism by showing that simple excitatory and inhibitory signal loops constitute specialized neural operations and neurocognitive functions on higher organizational levels by means of their interaction. Further scientific contributions are as follows.

New methodological developments:

1. Approach to represent multiple architectures of canonical microcircuits in a single generic canonical microcircuit model and to compare the system behavior (Section 3.1.3). In particular, the system behavior is evaluated in terms of (i) proposed basic information processing operations, namely working memory and signal flow gating, in response to transient rectangular stimulations (Section 4.2 and 5.2) and (ii) composite operations that emerge from the interaction of canonical microcircuit models (Sections 5.4, 5.5, 6.2, and 6.3). Notably, the analysis includes:
 - (i) classification of temporal response behavior (nonresponsive, transfer, and working memory)
 - (ii) documentation of temporal behavior (characteristic fingerprints)
 - (iii) assessment of sensitivity to parameter variations concerning the temporal behavior (dynamic function map) and the system’s state space (bifurcation diagrams).
2. Proposal of an approach to regroup two excitatory model populations to a single one to flexibly reflect different canonical microcircuit topologies of the same underlying cortical microarchitecture (Section 3.1.2).

3. Proposal of a processing network that considers the sequential and selective activation of canonical microcircuit models to represent syntax parsing at the neuronal level (Section 6.2 and 6.3).

New findings:

4. A bistable activity (of pyramidal cells) and salience selectivity (perception threshold and Hopf separatrix) are state space characteristics that underlie the basic operations of the canonical microcircuit model (Section 4.2). Identification of these characteristics in other neural models would indicate their ability for basic operations. The local network balance (regarding synaptic gains, Section 4.3) and the local feedback topology (Section 4.4) modify the model's state space. Hence, they constrain the capability for basic operations and composite operations, such as structure-building (Section 6.2). On the other hand, these factors might serve for the individualization of canonical microcircuits to specific tasks.
5. The canonical microcircuit model can store stimulation events when transmitted by feedforward connections but not when transmitted via feedback connections (Section 4.5). Moreover, the application of constant feedback information facilitates the sensitivity of a canonical microcircuit model to transient feedforward input and allows for allow state-dependent processing operations (Section 5.2). Hence, networks of interacting canonical microcircuit models possess a temporal processing history that allows the present processing behavior to depend on past processing steps (Section 5.6).
6. Cognitive priming is proposed to rest on an adaptive perception threshold of interacting canonical microcircuit model (Section 5.4).
7. Structure-building, that is the establishment of spatiotemporal patterns of neural activity, is proposed to rest on a selective activation of coupled minimal canonical microcircuits. This selectivity rests on superposition of hierarchical information flows (hierarchical conditioning) and differentiates this mechanism from others, such as stable heteroclinic channels and syn-fire chains (Section 5.5).
8. An explicit separation and hierarchical organization of lexical and syntactic information allows the syntax parsing model to account for multiple instantiations of words and complex syntactic structures (Section 6.3).

1.4 Structure

In addition to the current chapter this thesis comprises of six more chapters. The second chapter lays the conceptual groundwork of this thesis. It reviews essential concepts of neuronal interaction in associative networks and links the theory of neural cell assemblies to the notion of low-level regular circuits, i.e. canonical microcircuits. Moreover, the second chapter includes a definition of a minimal canonical microcircuit and provides provable arguments for the hypotheses that minimal canonical microcircuit (i) constitute nodes in models of distributed associative networks and (ii) perform local processing operations that are subfunctions of cognitive tasks.

The third chapter formularizes the canonical microcircuit model and introduces the methodological approaches and techniques for its analysis.

The fourth chapter is concerned with the information processing of a single canonical microcircuit model. The model's capability for basic operations is evaluated in three selected topologies and for varying levels of the local network balance.

The fifth chapter reports findings on the hierarchical interaction of canonical microcircuit models. It is shown how feedforward and feedback information flows mutually modulate each other and how the facilitating effect of feedback information supports the state-dependent processing of feedforward information.

The sixth chapter applies the collected findings to a larger network of interacting canonical microcircuit models that supports syntax parsing during sentence perception. Following a brief introduction to language processing and syntax parsing, two conceivable syntax parsing models of different complexity are compared.

The seventh chapter discusses the general approach and main findings, puts forth important points about future research questions, and concludes this thesis.

Chapter 2

Neural Canonical Microcircuits in Cognitive Information Processing

“Thus the machinery may be roughly uniform over the whole striate cortex, the difference being in the inputs. A given region of cortex simply digests what is brought to it, and the process is the same everywhere. [...] It may be that there is a great developmental advantage in designing such machinery once only, and repeating it over and over monotonously, like a crystal”

– Hubel and Wiesel, 1974

2.1 Neuronal Interaction Embodies Cognitive Information Processing

When contemplating about how the neural matter of the cerebral cortex gives rise to cognitive functions, one refers to a dense conglomerate of billions of glia cells and different types of neurons, each of which features thousands of synaptic contacts that allow communication via electrical and chemical signals. In a simplifying view on neuronal interaction, the spatiotemporal integration of excitatory and inhibitory postsynaptic potentials at the dendritic tree of a neuron shifts its membrane potential. If excited strongly enough, the neuron emits an action potential that is

sent via the neuron's axon to the dendritic trees of other connected neurons and leads to postsynaptic potentials again [24]. The effect and strength of afferent action potentials are determined by synaptic characteristics of excitatory or inhibitory operating neurons, for example the specificity of neuronal transmitters and sensitivity of receptors.

Based on this elementary mode of communication, neurons form circuits of relatively uniform structural architecture on the local level that give rise to more complex and specialized neural circuits on more global levels by virtue of long-range connectivity [25–29]. The propagation of neuronal activity in these hierarchical networks is assumed to reflect cognitive information processing [22, 30], as corroborated by physiological measurements (see for example [31, 32]). This gradual propagation of neuronal activity that establishes complex patterns in time and space is referred to as *structure-building*. *Structure-building computations* denote distinct mechanisms that describe the successive activation of neuronal units. To this end neural structure-building computations must ensure a quick, selective, and reliable composition (and decomposition) of neural activity despite the relatively fixed structural connectivity in the neocortex on a short time scale. Neurocomputational research has developed several theoretical constructs on structure-building computations [22, 33–36]. For instance, structure-building computations that sequentially activate neuronal units have been proposed to rest on *stable heteroclinic cycles* [37, 38], where activity transfer is guided through inhibitory networks. Alternative mechanisms are *sequence detectors* [21] and *synfire chains* [30, 31, 33] that employ a dynamic excitation of neuronal units.

Cell assembly theory is a neurocognitive brain theory supporting the notion that cognitive processing is reflected by the propagation of neuronal activity at the mesoscopic level, which is situated between the microscopic level of single neurons and the macroscopic level of cortical areas [4, 39–42]. Cell assembly theory assumes that cognitive functions correspond with specific patterns of observable neuronal activity whose establishment must follow distinct rules. The nature of this neuronal representation and the mechanisms guiding activity propagation, i.e. structure-building computations, are major points of cell assembly research. Generally, a *cell assembly*, or sometimes neuronal assembly, refers to a group of neurons with strong mutual excitatory connections that together reflect objects and more abstract entities of thought and thus form building blocks of cognition [4, 42]. Cell assemblies evolve through correlations of involved cells, such as through *Heb-*

bian coincidence learning, which strengthen the mutual connections [20, 40]. The progressive activation of these cell assemblies, called *association*, establishes the neural activity traces in the neocortex that embody cognitive processing. In 2006, Wennekers and colleagues [20] typify three different kinds of cell assembly associations: (i) *auto-associations*, where cell assemblies project onto themselves to enhance a localized activation (i.e., attractor states), (ii) *hetero-associations*, where two or more cell assemblies sequentially activate each other autonomously, similar to the concept of synfire chains, (iii) and *conditioned associations* that extend the spontaneous transitions of hetero-associations by an additional input that is required to gate the activation of cell assemblies and support a selective set up of neural activity patterns (termed *synfire graph*). Cell assemblies and their associations have been proposed to support various cognitive tasks including perception (e.g., pattern completion through *ignition* of a cell assembly), short-term memory (i.e., reverberations within or between assemblies maintain a persistent activity), long-term memory (i.e., formation of new cell assemblies), decision making, attention, and language tasks [4, 20].

2.2 Linking Cell Assembly Theory with the Concept of Canonical Microcircuits

By explaining the emergence of cognitive functions on the basis of interacting neuronal units, neural cell assembly theory is in line with the concept of connectionism. However, as noted by Pulvermüller [4], the spatial expansion of cell assemblies is not definite and has been addressed in different ways covering smaller pieces of brain, such as for the representation of a phonological word form [21], up to distributed *associative networks* (also called *operational cell assemblies* [20]) that spread over multiple cortical areas. Thus, cell assembly theory would benefit from a link of its spatially indefinite cell assemblies to concrete neurobiological structures that match the suggested association functionality.

In 2014, Palm and colleagues [41] raise the idea that nodes in associative networks may relate to hypercolumns [43], i.e. radial vertical columns in the neocortex. A similar concept of columnar processing units is reflected by *cortical columns*² [27, 44]. Cortical columns contain all major cortical neuronal pheno-

² "A cortical column is a complex processing and distributing unit that links a number of inputs to a number of outputs via overlapping internal processing chains" [44].

types, including pyramidal, basket, double bouquet, and stellate cells [45] that are vertically interconnected across the six cortical layers. Notably, neocortical neurons exhibit a distinct uniformity in terms of (i) local connectivity, (ii) long-range connectivity, (iii) intrinsic physiology, (iv) in vivo activity patterns, and (v) gene expression [43, 46–48]. Based on these properties, neocortical neurons are divisible into common groups, or *neuronal populations*, that were proposed to be the effective units of operation – rather than single neurons [44, 49]. The notions of *neural mass action* and uniformity inspired computational neuroscientists to investigate structure-function relationships in idealized population models [16, 17, 49]. On the other hand, the notion of uniformity, or homogeneity [20], of the neocortex gave rise to the notion of *canonical microcircuits*, fundamental neuronal circuits that feature a recurrent, basic architecture [9–11, 26, 50]. The local architecture of canonical microcircuits is attributed to a limited set of *stereotypic functions* [47, 51]. Despite their limited internal structural and functional complexity, canonical microcircuits have been suggested to accomplish the observable complexity of cognitive processing by virtue of their number and interaction within the hierarchically organized topology of the network [23, 52, 53].

So far, I pointed out how cognitive processing is supposed to rest on associative networks of cell assemblies whose nodes are linked to canonical microcircuits on the basis of recurrent connectivity patterns in cortical columns. In the remainder of this chapter, I will review structural and functional characteristics of canonical microcircuits, define the concept of a *minimal canonical microcircuit* that captures the functional abilities of cell assemblies, and identify central points of its investigation. The consequent analysis of minimal canonical microcircuits in this thesis will affirm the significance of canonical microcircuits in cognitive processing.

2.3 Structural and Functional Characteristics of Canonical Microcircuits

Neurobiological studies in the striate cortex of cats identified detailed recurrent connectivity schemes among major neuronal populations [29, 45, 54, 55]. Schemes of canonical microcircuits that have been formalized in computational models are called *canonical microcircuit models* (see the glossary in Appendix B). Canonical microcircuit models feature many conceivable architectures. Canonical microcir-

circuit models are used in detailed [56, 57] as well as simplified form [11, 58, 59] to investigate their inherent local functionality (see Fig. 2.1). Douglas and colleagues [11] considered a rather simple canonical microcircuit scheme of only three interacting excitatory and inhibitory populations receiving excitatory thalamic input. More detailed anatomical findings by Thomson and colleagues [29] and Binzegger and colleagues [45] disclosed separate excitatory and inhibitory neuronal populations for each cortical layer and their prevalent *intrinsic* connections. These schemes also incorporated the notion that canonical microcircuits participate in the postulated hierarchical organization of the neocortex [25]. The respective *hierarchical* connections allow a canonical microcircuit to integrate novel information from lower levels via *feedforward connections* and conceptual information from higher levels via *feedback connections* [58, 60].

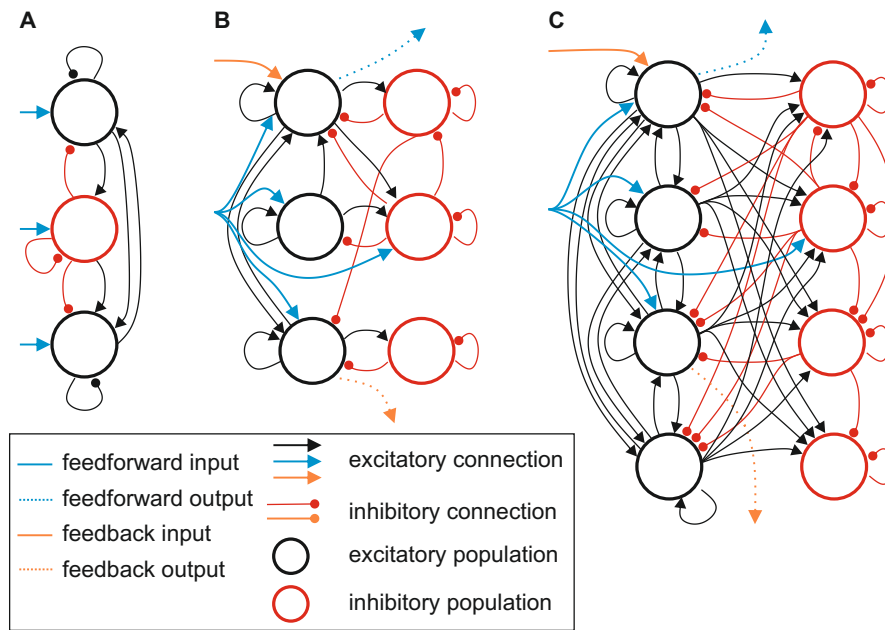


Figure 2.1: Canonical microcircuit models. Canonical microcircuit models describe recurrent connectivity patterns among neuronal populations. They feature intrinsic excitatory (black arrows) and inhibitory (red arrows) as well as hierarchical excitatory feedforward (light blue arrows) and feedback (orange arrows) connections that arrive at (solid lines) or depart (dashed lines) from the canonical microcircuit. A) Simplified canonical microcircuit model by Douglas [11]. B)-C) Detailed canonical microcircuit models as considered by Haeusler and colleagues [56] that is based on measured data by Thomson [29] (B) and Binzegger [45] (C).

Canonical microcircuit schemes such as reported by Binzegger [45] and Thomson [29] resolved intrinsic and hierarchical connections in a high level of detail. However, the universal validity of these detailed schemes was put into question based on non-striate sections of the neocortex [12, 13]. Moreover, the fact that no definitive description of a canonical microcircuit in terms of topology and the involved neuronal populations seems to be found yet [10] gave rise to the conjecture that the observable microcircuits are variants of a single idealized *homologous neocortical circuit* [47] that features structural and functional core characteristics. This notion is corroborated by the well-conserved organization of neocortex among vertebrates [61]. Following the idea of a reduced complexity of canonical microcircuit models, Bastos and colleagues [58] considered merely one inhibitory and three excitatory neuronal populations, namely superficial and deep pyramidal cells and excitatory interneurons, to investigate hierarchical processing among canonical microcircuits. Their study showed that relatively simple positive and negative feedback loops suffice to investigate cortical information processing.

Canonical microcircuits are supposed to support stereotypical functions [51], also referred to as canonical neural computations [62], universal computational capabilities [56] or computational primitives [10]. These stereotypical functions range from mechanisms for gain control and signal restoration [63], amplification and signal normalization [12, 62], selectivity and computation of gain [53] to linear (e.g., summation, division, and sign inversion) and nonlinear operations (e.g., winner-takes-all, invariance, and multistability) [10]. Based on computational models of data-based microcircuits [29, 45], Häusler and Maass [56, 64] showed that computational properties (spike pattern classification, memory tasks, and nonlinear fusion) rest on the stereotypic functions of accumulating, holding, and fusing of information. Moreover, they emphasized the value of non-fading memory abilities arising from feedback connections [57].

These proposed stereotypic functions coincide with the basic functionalities of information processing networks, namely weighting, storing and switching of information [18]. Based on these basic functionalities, I define *signal flow gating* (combining weighting and switching) and *working memory* (representing storing) as a minimal set of basic information processing operations or shortly *basic operations* [59], that is fundamental stereotypic functions a canonical microcircuit needs to feature to perform elemental information processing. In particular, signal flow gating controls the transmissibility of neural signals and allows for a differentiated

response to comparable input. This selectivity is likely to depend (i) in a bottom-up way on the properties of the input signal itself and (ii) on feedback modulation of the canonical microcircuit by the global network. The selection of input according to its salience is an important processing property, for example, associated with steering visual attention [65], selective reaction to sensory input, and determination of processing pathways [66]. Moreover, working memory enables the processing of temporally structured information in the brain. This requires a fast memory mechanism that does not rely on structural (e.g., synaptic) changes [67] but is sufficiently time stable [57].

2.4 Analysis of a Minimal Canonical Microcircuit

Combining structural and functional characteristics of canonical microcircuits, I define a *minimal canonical microcircuit* as a canonical microcircuit model that (i) considers interacting excitatory and inhibitory neuronal populations, (ii) whose structural topology reflects the homogeneity of neocortical matter, and (iii) whose functional repertoire features the basic operations of signal flow gating and working memory (see the glossary in Appendix B). Hence, minimal canonical microcircuits are a subset of canonical microcircuit models. This thesis investigates three out of many conceivable architectures of canonical microcircuit models for their conformity with the concept of minimal canonical microcircuits.

It is a central working hypothesis of this thesis that minimal canonical microcircuits constitute nodes in models of distributed associative networks. To substantiate this hypothesis, the investigation of a single canonical microcircuit model needs to prove its capability for the basic operations of signal flow gating and working memory and demonstrate their correspondence to the three types of cell assembly associations. Moreover, the minimal model architecture that supports working memory and signal flow gating as basic operations needs to be determined.

In accordance with former approaches [58], a model of a cortical microcircuit should feature pyramidal cells with long axons projecting to distant cortical areas, that is hierarchic connections, as well as local excitatory and inhibitory feedback loops. Such architectures have been represented in the parsimonious form of neural mass/field models [15, 16, 49, 68, 69]. Extensive investigations of such models have convincingly demonstrated that the basic features of local cortical architecture

do indeed support the aforementioned basic operations [70]. That is, they exhibit bistability (memory behavior) and bifurcations (switching/gating behavior). Further, various analyses of neural mass models have investigated local dynamics and charted their dependence on specific parameters such as the number and type of considered populations or the target of afferent input [70–73]. However, a number of key issues remain insufficiently investigated, in particular with respect to conceivable minimal canonical microcircuit architectures and the basic operations discussed above:

- **Indirect versus direct excitatory local feedback.** Most neural mass and field models only consider a single excitatory neural population (e.g., [68]) where the excitatory local feedback loop is modeled as a recurrent (direct) feedback. Other approaches (e.g., [15]) distinguish between pyramidal cells, which form the main sources of EEG and MEG, and excitatory interneurons, which are supposed to be the main receiver of bottom-up afferent input [25, 53]. In this three-population model the excitatory local feedback³ to the pyramidal cells is indirect via the excitatory interneurons. Neurobiological studies have established that excitatory interneurons and pyramidal cells populate separate cortical layers [50]: pyramidal cells mainly in layers 3, 5, and 6, excitatory interneurons predominantly in layer 4. Furthermore, the finding that bottom-up sensory information tends to be fed mainly into the layer 4 neurons, whereas top-down modulating signals target pyramidal cells [25], supports the consideration of two separate excitatory populations. Garnier and colleagues [72] examined the consequences of direct compared to indirect excitatory local feedback and reported that an indirect feedback path provided additional dynamics. However, the relevance of these additional dynamics needs to be balanced against the costs of increased model complexity and should be assessed according to the specific modeling requirements [53]. It is therefore important to assess the sensitivity of the basic operations of a canonical microcircuit model with respect to this choice.
- **Recurrent inhibitory local feedback.** The axons of inhibitory interneurons form collaterals that target the same or other inhibitory neurons [44]. This ‘inhibition of inhibition’ is incorporated into some models [68], while disregarded in others [15]. As the nonlinear effect of this disinhibition with

³ Here, local feedback refers to the intrinsic connectivity, that is, within a microcircuit – as opposed to feedback mediated through extrinsic (inter-circuit) connectivity, that is, between microcircuits.

respect to the basic operations is difficult to predict theoretically, it needs to be investigated by numerical simulations.

- **Local network balance.** An important characteristic of neural circuits and central for the interaction of neuronal populations [74] is the local relationship of inhibition and excitation, often referred to as *network balance*. The healthy brain establishes a dynamic network balance of excitation and inhibition, as shown theoretically [75] and experimentally in both in vitro [76] and in vivo studies [77, 78]. Haider and colleagues [78] propose that changing the network balance modulates the functionally relevant neuronal responsiveness through a transition of stable network states. This also means that a network imbalance disturbs cortical processing mechanisms, impairs cognitive function, and can lead to severe brain malfunctions and disorders, such as epilepsy [77, 79], autism [80–82], schizophrenia [83, 84], and excitotoxicity [85]. It is, therefore, of high interest to assess the information processing capacity of canonical microcircuits in the light of a variable local network balance.

2.5 Interaction in Hierarchic Networks

Cognitive tasks have been proposed to emanate from hierarchically operating cell assemblies [20]. Accordingly, minimal canonical microcircuits receive and send hierarchical, namely feedforward and feedback, information in order to perform composite processing operations that are subfunctions of cognitive tasks. To substantiate this hypothesis, the investigation of interacting minimal canonical microcircuits needs to address the following points. Firstly, in order to operate hierarchically, a minimal canonical microcircuit needs to respond differently to feedforward and feedback input [86]. Secondly, interacting minimal canonical microcircuits need to be able to perform local processing operations in minimal meta-circuits. Various computational models have been proposed that investigate the cooperation of multiple canonical microcircuits to realize higher cognitive functionality, such as attentional processing [87], predictive coding [86], language processing [4, 20], and visual steering [88]. Extending this research, I investigate *priming*, that is the adaptive processing of inputs in dependence on previous events, and *structure-building*, that is the sequential establishment of spatiotemporal neural activity patterns. Moreover, I focus on how local processing operations emerge from the hi-

erarchical interaction of minimal canonical microcircuits. Thirdly, it needs to be shown that cognitive tasks may emerge from a combination of local processing operations. To this end, the parsing of syntactic information during the perception of a spoken sentence [89] is demonstrated to result from cascaded structure-building computations within a network of many interacting canonical microcircuits.

Chapter 3

Methods

“Aut viam inveniam aut faciam.” (I shall either find a way or make one.)

– Hannibal, 218 v. Chr

3.1 The Canonical Microcircuit Model – Architecture and Parameterization

3.1.1 Determination of Potential Minimal Canonical Microcircuit Architectures

In this chapter, I present the modeling framework that was used to investigate minimal canonical microcircuits. This modeling framework comprises the determination of a generic canonical microcircuit model that accounts for topological choices in terms of intrinsic and hierarchical connectivity, its formularization and parameterization as well as simulation and analysis procedures.

Simple circuits of local excitatory and inhibitory feedback loops proved adequate to investigate cortical information processing [58] and to meet the requirements for basic operations, i.e. bistability and bifurcations [70]. Hence, I consider three major neuronal populations, namely excitatory pyramidal cells⁴ (Py), excitatory interneurons⁵ (EIN), and inhibitory interneurons⁶ (IIN) to compose a canonical

⁴ representing pyramidal cells in cortical layers 2/3 and 5/6

⁵ representing stellate cells in layer 4

⁶ representing a common pool of all inhibitory neurons

microcircuit model. The mean membrane potential of the Py, integrating both positive and negative feedback, forms the observable signal of the circuit (e.g., by EEG) and at the same time gives rise to the output signal to distant areas through the activation function. Hence, the description of observable dynamics is centered on this principal cell population. Two major points of investigation are identified concerning the architecture of a minimal canonical microcircuit (see Section 2.4): (i) the consequence of direct versus indirect excitatory local feedback and (ii) the influence of disinhibition through recurrent inhibitory local feedback. Thus, the following local feedback topologies were considered:

- The **three-population model** with an indirect excitatory local feedback path through the EIN (see Fig. 3.1A): this reference architecture is likely to support the basic operations as it exhibits bistability and bifurcations [70]. By separating populations that receive feedforward (to the EIN) and feedback input (to the Py) [25] this architecture represents a simplified, though established [86], reflection of hierarchical connections. Output is emitted by the Py.
- The **two-population model with direct excitatory local feedback** through self-connections of the pyramidal cells (Fig. 3.1B): for this compressed architecture the two excitatory populations, Py and EIN, are gradually mapped into a single one, Py', that both receives and emits hierarchical signals.
- The **two-population model with direct excitatory local feedback and recurrent inhibitory feedback** of the IIN (see Fig. 3.1C): this architecture accounts for disinhibition. Py' both receives and emits hierarchical signals.

The canonical microcircuit model of this thesis commonly accounts for these different local feedback topologies. I assume that all topologies reflect the same cortical microstructure, but differently group the model elements. Hence, to flexibly reflect the topologies without any qualitative change of the underlying cortical microarchitecture, two excitatory populations, EIN and Py, need to be regrouped to a single excitatory population, Py'. Notably, this continuous mapping denotes a weighting of direct and indirect local feedback of the pyramidal cells and features the three-population model and the two-population model as extreme cases. The following section presents the mapping.

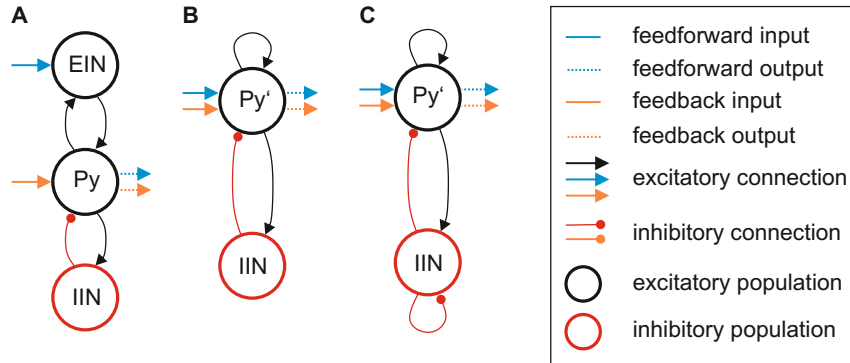


Figure 3.1: Local feedback topologies of the canonical microcircuit model. A) The three-population model with an indirect excitatory local feedback path through the EIN. B) The two-population model with direct excitatory local feedback of the Py'. C) The two-population model with direct excitatory local feedback and recurrent inhibitory feedback of the IIN.

3.1.2 Gradual Mapping of Excitatory Populations

Mapping two excitatory populations into a single excitatory population without any qualitative change of the underlying cortical microarchitecture demands the conservation of: (i) the total number of neurons in the system, (ii) the sum of flowing currents, and (iii) all connections between single neurons.

In the two-population case, I assume that the EIN population contains M_E neurons and the Py population contains M_P neurons. In the following, the connectivity gain N_{ab} denotes the average number of neurons of population b from which a single neuron in population a receives input. When mapping the two-populations onto a single population, one gradually merges more and more neurons from EIN, a fraction denoted by $\beta = \alpha \cdot M_E$, with neurons from Py into a growing population Py'. α denotes a linear scaling factor ranging from 0 to 1.

For $\alpha = 0$, $\beta = \alpha \cdot M_E = 0$, no neurons have been transferred yet and the two excitatory populations feature the regular connectivity strengths N_{EP} and N_{PE} (top panel in Fig. 3.2A).

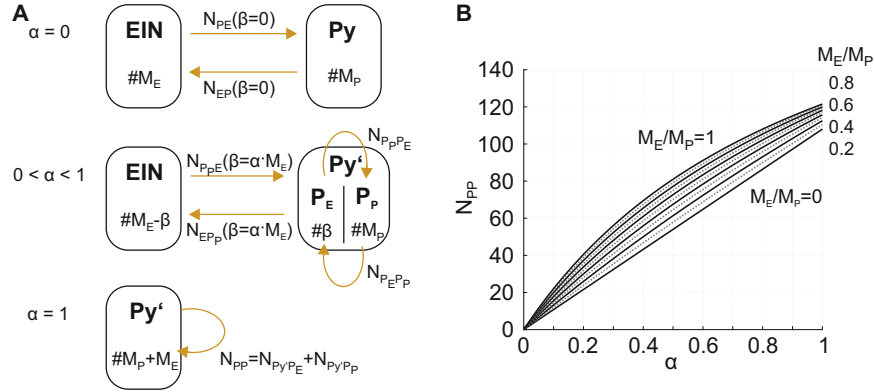


Figure 3.2: Regrouping two excitatory populations to a single population. A) Top: For $\alpha = 0$, the two excitatory populations EIN (of number M_E) and Py (of number M_P) are coupled through connections of strength N_{PE} and N_{EP} , respectively. Middle: For $0 < \alpha < 1$, a fraction of neurons (P_E) of number $\beta = \alpha \cdot M_E$, which was previously associated with the EIN, merges with former neurons of the Py, P_P , to a growing population Py'. While the connection strength N_{EP} stays the same (P_P still fire) internal feedback $N_{P_E P_P}$ is established within population Py'. Moreover, feedback from EIN to Py', $N_{P'E}$, is reduced in favor of self-connections, N_{PP} , within population Py'. Bottom: For $\alpha = 1$, all former EIN neurons are merged with former Py neurons constituting population Py' that features self-feedback of strength N_{PP} . B) Development of N_{PP} over α , implying a co-existence of direct and indirect excitatory local feedback in the model, for various proportions of excitatory interneurons and pyramidal cells M_E/M_P .

When $0 < \alpha < 1$, neurons of number $\beta = \alpha \cdot M_E$ are transferred from EIN to Py', which are denoted P_E . Within Py' there is now a fraction P_P of number M_P and a fraction P_E of number β . Thus, the total number of neurons in EIN is $M_E - \beta$ and the total number of neurons in Py' is $M_P + \beta$ (middle panel in Fig. 3.2A). The connectivity has changed in the following way: the number of connections from EIN to P_P (former N_{PE}) is scaled by a factor $(M_E - \beta)/M_E$, because fewer neurons in EIN are available to project onto Py' (source fraction). This is reflected by:

$$\begin{aligned} N_{P_P E}(\beta = \alpha \cdot M_E) &= \frac{M_E - \beta}{M_E} \cdot N_{PE}(\beta = 0) \\ &= (1 - \alpha) \cdot N_{PE}(\beta = 0) \end{aligned} \quad (3.1)$$

Connectivity from P_P to EIN (former N_{EP}) remains the same, because the number of neurons, which originally projected from Py to EIN, now bounded in P_P , remains constant at M_P (source fraction). Thus, N_{EP} reads:

$$N_{EP_P}(\beta = \alpha \cdot M_E) = N_{EP}(\beta = 0) \quad (3.2)$$

The transfer of neurons gives rise to self-connections within Py' , namely N_{PP} . These are represented by the sum of projections from P_P to P_E ($N_{P_E P_P}$) and projections from P_E to P_P ($N_{P_P P_E}$). Considering $N_{P_E P_P}$, similar to the fully separated case, P_P (source fraction) still projects onto P_E :

$$N_{P_E P_P}(\beta = \alpha \cdot M_E) = N_{EP}(\beta = 0) \quad (3.3)$$

For $N_{P_P P_E}$ the P_E project onto P_P but are scaled by a factor β/M_E reflecting their reduced number (source fraction) as in:

$$N_{P_P P_E}(\beta = \alpha \cdot M_E) = \frac{\beta}{M_E} \cdot N_{PE}(\beta = 0) \quad (3.4)$$

As the population Py' is growing, the incoming input is distributed among more and more neurons. This is reflected by an additional scaling of the connectivity gains (distribution factor). Connections from EIN, that is $N_{P_P E}$ (Equation 3.1), are scaled by a distribution factor $M_P/(M_P + \beta)$. Thus $N_{Py'E}$ reads:

$$\begin{aligned} N_{Py'E}(\beta = \alpha \cdot M_E) &= \frac{M_P}{M_P + \beta} \cdot N_{P_P E}(\beta = 0) \\ &= \frac{M_P}{M_P + \alpha \cdot M_E} \cdot (1 - \alpha) \cdot N_{PE}(\beta = 0) \\ &= \frac{1 - \alpha}{1 + \alpha \cdot M_E/M_P} \cdot N_{PE}(\beta = 0) \end{aligned} \quad (3.5)$$

Within Py' , $N_{P_E P_P}$ (Equation 3.3) is scaled by a distribution factor $\beta/(M_P + \beta)$. Thus, $N_{Py'P_P}$ reads:

$$\begin{aligned} N_{Py'P_P}(\beta = \alpha \cdot M_E) &= \frac{\beta}{M_P + \beta} \cdot N_{P_E P_P}(\beta = 0) \\ &= \frac{\alpha}{M_P/M_E + \alpha} \cdot N_{EP}(\beta = 0) \end{aligned} \quad (3.6)$$

Also within Py' , $N_{P_P P_E}$ (Equation 3.4) is scaled by a distribution factor $M_P/(M_P + \beta)$. Thus, $N_{P_{Y'} P_E}$ reads:

$$\begin{aligned} N_{P_{Y'} P_E}(\beta = \alpha \cdot M_E) &= \frac{M_P}{M_P + \beta} \cdot N_{P_P P_E}(\beta = 0) \\ &= \frac{M_P}{M_P + \beta} \cdot \frac{\beta}{M_E} \cdot N_{PE}(\beta = 0) \\ &= \frac{\alpha}{1 + \alpha \cdot M_E/M_P} \cdot N_{PE}(\beta = 0) \end{aligned} \quad (3.7)$$

Thus, the total number of self-connections in Py' is:

$$\begin{aligned} N_{PP}(\beta = \alpha \cdot M_E) &= N_{P_{Y'} P_E} + N_{P_{Y'} P_P} \\ &= \frac{\alpha}{1 + \alpha \cdot M_E/M_P} \cdot N_{PE}(\beta = 0) + \frac{\alpha}{M_P/M_E + \alpha} \cdot N_{EP}(\beta = 0) \end{aligned} \quad (3.8)$$

Assuming that the ratio between number of excitatory interneurons and number of pyramidal cells, M_E/M_P , is 0.25 [90] and that $N_{PE} = 108$ and $N_{EP} = 135$, N_{PP} equals 113.4 in case of the completed mapping (see bottom panel in Fig. 3.2A). The diagram in Figure 3.2B shows the function $N_{PP}(\alpha)$, for different M_E/M_P . In essence, this mapping represents a regrouping of neurons in the populations that still make the same basic assumptions. The resulting model topologies have been analyzed separately. Consequent changes in the bifurcation structure of the system are reported in Section 4.4.1.

3.1.3 Formularization of the Canonical Microcircuit Model

The concept of canonical microcircuits considers the interaction of neuronal populations instead of single neurons. Neural mass models [15, 17] consider neuronal populations and typify their mean field descriptions as *neural masses* [1]. Accordingly, I chose this type of model to formalize a canonical microcircuit model. In the past, neural mass models already served to explain electroencephalography data [15, 71], to elucidate epileptogenic processes [73, 91] and electrical brain stimulation [92, 93], and to describe the local steady-state system behavior of a circumscribed neural area [70, 94, 95]. Moreover, neural mass models served to reveal

diverse mechanisms in processing systems (i.e., dynamic causal modeling [5, 71]) and to describe a potential realization of predictive coding [86]. Notably, neural mass models strike a balance between neurobiological plausibility and mathematical tractability and interpretability.

The present canonical microcircuit model uses three neural masses to account for the selected local feedback topologies (see Section 3.1.1) – one neural mass each for the pyramidal cells, excitatory interneurons, and inhibitory interneurons. In each neural mass the afferent mean firing rate $\phi(t)$ arriving at the dendritic tree of a neuronal population is transformed to a respective mean membrane potential $V(t)$ by convolving the firing rate with a synaptic response kernel $h_{e,i}(t)$ as in:

$$V_{e,i}(t) = \phi(t) * h_{e,i}(t) \quad (3.9)$$

where the index $e(i)$ denotes the synaptic response kernel of an excitatory (inhibitory) neural mass. The synaptic response kernel is modeled as an alpha-function:

$$h_{e,i}(t) = \frac{H_{e,i}}{\tau_{e,i}} \cdot t \cdot \theta(t) \cdot e^{-t/\tau_{e,i}} \quad (3.10)$$

where $\theta(t)$ denotes the Heaviside function, $H_{e,i}$ is the synaptic gain, reflecting number and efficacy of synaptic contacts, and $\tau_{e,i}$ is the characteristic time constant of either excitatory or inhibitory operating neural masses. The mean membrane potential $V_c(t)$, $c \in [Py, E, I]$ of the respective neural masses then depends on the sum of all incoming input components. Using Green's function this is expressed as:

$$D_{e,i} \cdot V_c = \sum \phi(t) \quad (3.11)$$

where $D_{e,i}$ denotes an excitatory (inhibitory) second order temporal differential operator and reads:

$$D_{e,i} = \frac{\tau_{e,i}}{H_{e,i}} \left(\frac{d^2}{dt^2} + \frac{2}{\tau_{e,i}} \cdot \frac{d}{dt} + \frac{1}{\tau_{e,i}^2} \right) \quad (3.12)$$

where Equation 3.10 represents the Green's function of this differential operator. This operator is then decoupled into two first-order differential equations. The

transformation of a mean membrane potential to a mean firing rate, representing the processes occurring at the axonal hillock of a neuron, is modeled by a sigmoidal activation function $S(\cdot)$, in this case the logistic function:

$$\phi(t) = S(V_c(t)) = \frac{2e_0}{1 + e^{r(v_0 - V_c(t))}} \quad (3.13)$$

Whereas e_0 represents the half of the highest achievable firing rate, r is the maximum slope of the sigmoidal activation function and v_0 denotes the membrane potential for which half of the maximum firing rate is invoked.

The canonical microcircuit model features a model structure (Fig. 3.3A) that flexibly accounts for the chosen local feedback topologies (see Section 3.1.1 and Fig. 3.1) through variation of two architectural parameters b_1 , b_2 . The first parameter b_1 allows a gradual transition between the three-population model with feedforward input (p_{ff}) received by the excitatory interneurons EIN ($b_1 = 1$, Fig. 3.3B) and the two-population models with feedforward input (p_{ff}) received by the excitatory population Py' ($b_1 = 0$, Figs. 3.3D and E). Moreover, b_1 weights direct and indirect excitatory local feedback of the Py. Importantly, intermediate situations can be modeled where both excitatory populations receive extrinsic input and both direct and indirect excitatory local feedback loops co-exist ($0 < b_1 < 1$). This approach enables the treatment of a principle choice in model structure (two vs. three-population model) as a continuous parameter, such as for a bifurcation analysis. The second architectural parameter (b_2) controls the presence of the recurrent feedback loop for the IINs ($b_2 = 0$, Fig. 3.3E). Another architectural parameter b_3 regulates the presence of feedback input (p_{fb}) at the Py of the three-population model ($b_3 = 1$, Fig. 3.3C). This hierarchical input reflects physiological findings, namely that thalamic or feedforward input arrives in layer 4 (at the EIN) and lateral or feedback input arrives in layers 2/3 (at the Py) [25,64]. The membrane potential of the pyramidal cells, $V_{Py}(t)$, represents the output of the model that is sent to connected circuits and is detectable, for example, by EEG.

Accordingly, the system of governing equations of the canonical microcircuit model read:

$$\begin{aligned}
D_e \cdot V_1 &= N_{EP} \cdot \phi_P + b_1 \cdot p_{ff} \\
D_e \cdot V_2 &= b_1 \cdot N_{PE} \cdot \phi_E + (1 - b_1) \cdot N_{pp} \cdot \phi_P + (1 - b_1) \cdot p_{ff} + b_3 \cdot p_{fb} \\
D_i \cdot V_3 &= N_{PI} \cdot \phi_I \\
D_e \cdot V_4 &= N_{IP} \cdot \phi_P \\
D_i \cdot V_5 &= (1 - b_2) \cdot N_{II} \cdot \phi_I \\
\phi_P &= S(V_{Py}) = S(V_2 - V_3) \\
\phi_E &= S(V_E) = S(V_1) \\
\phi_I &= S(V_I) = S(V_4 - V_5)
\end{aligned} \tag{3.14}$$

The parameters N_{ab} denote the connectivity gains between the source population b and the target population a , where $a, b \in [P, E, I]$. This system of non-linearly coupled linear ordinary differential equations, in the following called governing system equations, is transformed into a dimensionless form, see Appendix A.1. The system is initially parameterized according to a previously used configuration [15], here referred to as the default parameterization, see Table 3.1. Further, the connectivity gains N_{PP} and N_{II} are determined to be $N_{PP} = 113.4$ and $N_{II} = 33.25$, similar to other inhibitory connection strengths.

Table 3.1: Default parameterization of the canonical microcircuit model.

Parameter	Value	Unit	Parameter	Value	Unit
H_e	3.25	mV	N_{PE}	$0.8 \cdot N_{EP}$	-
H_i	22	mV	N_{IP}	$0.25 \cdot N_{EP}$	-
τ_e	10	ms	N_{PI}	$0.25 \cdot N_{EP}$	-
τ_i	20	ms	N_{PP}	113.4	-
r	0.56	mV^{-1}	N_{EP}	135	-
v_0	6	mV	N_{II}	33.25	-
e_0	2.5	s^{-1}			

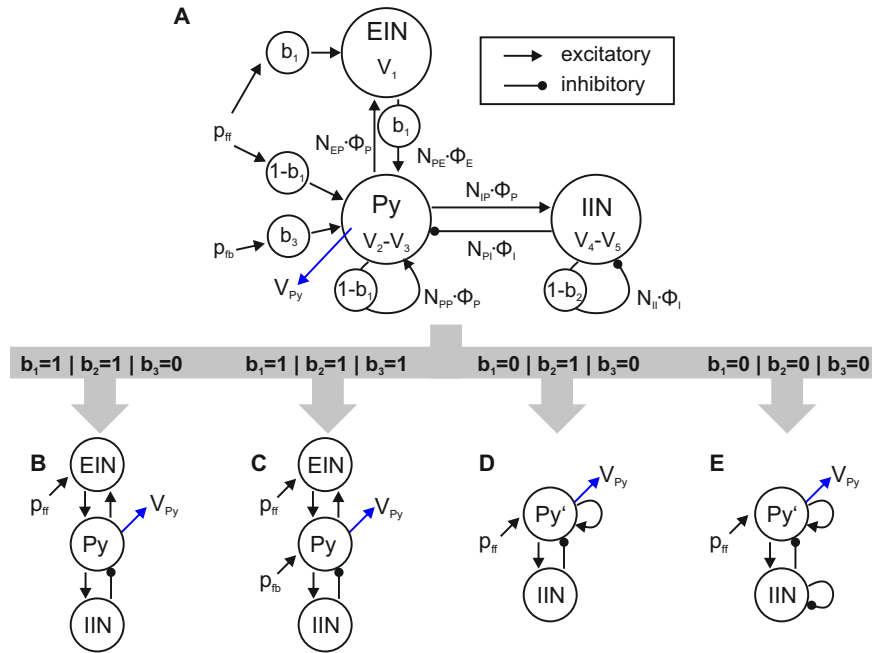


Figure 3.3: Generalized architecture of the canonical microcircuit model. A) The canonical microcircuit model accounts for excitatory interneurons (EIN), inhibitory interneurons (IIN), and pyramidal cells (Py). The architectural parameter b_1 controls the deployment of direct and indirect excitatory local feedback as well as the feedforward input receiving population. The consideration of inhibitory collaterals is governed by the architectural parameter b_2 . Moreover, b_3 regulates the presence of extrinsic feedback input in the three-population model. This parameterization allows for a comparative investigation of relevant changes in the dynamical behavior among the three distinct architectures: a three-population model (B and C) and two-population models without (D) and with recurrent inhibitory feedback of the IIN (E). The transmitted mean firing rates $\phi(t)$ are scaled by connectivity gains N_{ab} between the source population, b , and the targeted population, a , respectively. The membrane potential of the pyramidal cells, $V_{Py}(t) = V_2(t) - V_3(t)$, represents the output of the model (indicated by blue arrows), that is detectable, for example, by EEG.

3.1.4 Local Network Balance

A central point of analysis is the parameterization of the local network balance in the canonical microcircuit model. However, the concept of network balance is ambiguous and difficult to quantify simply as ratio of excitation and inhibition in a neural system. This is due to the multiple spatial and temporal scales in the brain [74, 82] and the multiple structural and functional aspects that could be con-

sidered. A description of network balance on a mesoscopic level of interacting neuronal populations may focus on structural influences (i.e., topology, number, and efficacy of synaptic contacts), functional features (i.e., conveyed firing rates) or factors of the synaptic response. In computational models of other studies, proposed approaches relate excitatory and inhibitory charges, conductances [78,79,82] or membrane potentials [74] to each other. Often the network balance is defined in a network context as the ratio of recurrent inhibition to excitation [96,97]. However, this is limited to two-population models and becomes ambiguous for multiple population models such as the present canonical microcircuit model.

In the present canonical microcircuit model, the relevant parameters for excitation and inhibition include: (i) the synaptic response function (time constants, synaptic gains), (ii) the external input to all three populations, (iii) the parameters of the sigmoidal activation function, and (iv) the connectivity gains between the populations. Among these parameters, the synaptic gains $H_{e,i}$, the connectivity gains, i.e. N_{EP} , N_{PE} , N_{PI} , N_{IP} , N_{PP} , and N_{II} , and the external inputs have the most direct and biologically plausible effect on the excitation and inhibition of the system. However, note the formal redundancy, yet conceptual difference among synaptic gains and connectivity gains in the system equations. According to the governing system equations H_i , N_{PI} , and N_{II} reflect a gain of inhibitory feedback just with different scaling. Also, varying H_e is equivalent to synchronously varying N_{EP} , N_{PE} , N_{IP} , and N_{PP} . In the interest of tractability it is, thus, sufficient to consider a parsimonious set of parameters for the investigation of the modulating influence of excitation and inhibition to the local dynamics. Hence, the analysis focuses on the influence of H_e and H_i , which are interpreted to represent efficacy and density of excitatory, e.g. AMPA, and inhibitory, e.g. GABA_A, neurotransmitter receptors. This is equivalent to the number and strength of the synaptic weights.

3.2 Model Simulation and Data Analysis

3.2.1 Time Domain Simulations

The described canonical microcircuit model represents a *dynamical system* as it features a set of state variables whose temporal evolution is described by an evolutionary law [98]. Dynamical systems theory provides a mathematical framework and associated analysis procedures to systematically investigate the dynamical be-

havior of dynamical systems. In the following, I assume the reader's familiarity with basic aspects of dynamical systems theory [98, 99].

All analysis steps involved time domain simulations for which the governing system equations were numerically integrated using Heun's method. The stability for the integration interval of 1 ms was verified. All state variables were initialized with a zero vector and without external stimulation so that the system consistently resided on the lower branch of the S-shaped fixed point curve in the case of a bistable regime (see Section 4.2). Unless stated differently, a simulation lasted for 5 s during which a neuronal population of the model received a rectangular impulse of defined intensity and duration. The impulse was added to the feedforward or feedback input with zero baseline after a 1 s settling time. Feedforward impulses mimicked input from upstream sources, such as sensory information stemming from primary cortical areas, or higher level information, such as spoken words or phonemes. Feedback impulses mimicked input from downstream sources, such as conceptual information stemming from higher associative cortical areas.

The maximum of the membrane potential of the Py, $V_{Py}(t)$, was compared to a *firing threshold* in three consecutive analysis windows, namely prestimulus, immediate response, and asymptotic (Fig. 3.4A). Based on the sigmoidal activation function, the firing threshold was defined relative to the maximum firing rate of 5 s^{-1} , so that about 25% of the maximum firing rate was reached at the threshold of 4 mV. In each time window the system was considered to be active if the maximum activity exceeded the threshold and to be inactive if the maximum activity was below the threshold. Three discriminable types of response behavior were defined that together constituted the basic operations of signal flow gating and working memory: (i) *nonresponsive* for sub-threshold transient deflections in $V_{Py}(t)$, (ii) *transfer* for transient supra-threshold deflections, and (iii) *memory* for sustained supra-threshold deflections, see Figure 3.4B. Note that in some cases the activity oscillated around the threshold. In these cases the population activated postsynaptic populations at least for part of the time and was therefore considered to be above threshold. The response behaviors were mapped to the stimulus' properties (i.e., stimulus duration and intensity) in so-called *characteristic fingerprints*. These characteristic fingerprints typified the dynamical response repertoire of the respective parameterization and allowed the simple assessment of accessible basic operations. Collections of characteristic fingerprints with different parameteriza-

tions (e.g., synaptic gains) were arranged in *dynamic function maps*. All simulation and analysis scripts are written in Python.

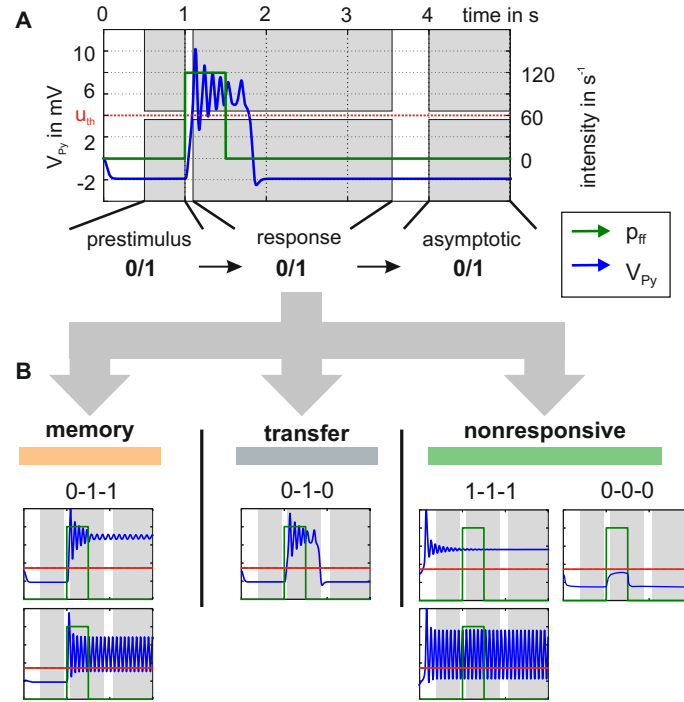


Figure 3.4: Stimulation principle and categorization of the dynamic response behavior. A) A neuronal population of the canonical microcircuit model received a rectangular stimulation of varying intensity and duration (green line). The maximum of the mean membrane potential of the pyramidal cell population (blue line) was recorded in three time windows, i.e. prestimulus window (0.5 s - 1 s), immediate response window (1.1 s - 3.5 s), and the asymptotic window (4 s - 5 s), as indicated by gray shaded areas. To classify the response behavior, these activity values were compared to a firing threshold of 4 mV (red horizontal line, u_{th}) in each window – ‘0’ denoting a subthreshold activation and ‘1’ denoting an activation exceeding the threshold. B) The combined evaluation of the activities (e.g., ‘0-1-1’) led to three distinct classes of response behaviors: memory, transfer, and nonresponsive behavior. The plotted curves exemplify the behavior of the three-population model with default parameters (see Table 3.1) stimulated with feedforward input.

3.2.2 Bifurcation and Stability Analysis

Bifurcation diagrams and explanatory mappings were used to investigate the canonical microcircuit model’s state space. To this end, the governing system equations were implemented in a dimensionless form in Matlab (The MathWorks Inc., Natick, Massachusetts, USA) in order to perform a bifurcation and stability

analysis using the numerical continuation tool DDE-BIFTOOL [100]. Standard methods to compute fixed point curves were used, i.e. computation of fixed points, derivation of Jacobian matrix, linearization of the system around the fixed points, and evaluation of the eigenvalues to determine the local stability [98].

Local bifurcations, that is sudden dynamic state transitions for which the system's largest eigenvalue crosses the imaginary axis, were evaluated for either one or two continuously varied parameters, the so-called bifurcation parameters. Relevant bifurcation parameters were the external input to a population, p_{ff} and p_{fb} (to assess their individual effect), the synaptic gains H_e and H_i (to evaluate the influence of the network balance), and the architectural parameters b_1 and b_2 (to transform local feedback topologies).

Chapter 4

Basic Operations in the Canonical Microcircuit Model

“A mechanism is an organized system of component parts and component operations. The mechanism’s components and their organization produce its behavior, thereby instantiating a phenomenon.”

– William Bechtel, 2005

4.1 Introduction

This chapter examines the information processing capacity of a single canonical microcircuit model (Section 3.1.3) in consideration of the selected local feedback topologies (see Section 3.1.1) and the local network balance. The analysis discloses the conditions under which the canonical microcircuit model supports mechanisms for signal flow gating and working memory and allows the nomination of an architecture that matches the definition of a minimal canonical microcircuit.

I first show how the three-population model processes feedforward information by means of its basic operations signal flow gating and working memory - and address the necessary state space conditions. Subsequently, I show how variations in the local network balance affect the availability of basic operations via modifications of the system’s state space. This analysis procedure is repeated for the alternative local feedback topologies, including brief statements on state space modifications

during their transformation into each other. Finally, I examine how feedback information is processed in the three-population model, which is the only architecture that provides separate excitatory populations for hierarchical inputs.

4.2 Stimulation-induced Response Behaviors and Basic Operations

Rectangular bursts with a distinct intensity and duration were applied to the EIN of the three-population model with indirect excitatory local feedback (i.e., $b_1 = b_2 = 1$, $b_3 = 0$, see Fig. 3.3B). Figure 4.1A shows the distinct response behaviors: The system responds to weak and brief stimuli with a small deflection of the Py membrane potential (nonresponsive behavior). For stronger, though still brief, stimuli the system responds with a large transient, exceeding the firing threshold of 4 mV (transfer behavior). This transient activation, typical for perception, does not affect the sensitivity to further stimulation (i.e., no hysteresis). For both response behaviors the system settles down to its original activity state shortly after the stimulus is turned off. In contrast, for longer lasting stimuli the system settles down in a stable state of higher activity and remains insensitive to further stimuli or noise (memory behavior). This memorization of the stimulation event might affect processing even long after the stimulation itself. Being in this stimulus-selective high-activity state, a brief input to the IIN can actively reset the system to the lower activated state.

The characteristic fingerprint in Fig. 4.1B summarizes this dependency of the response behaviors on the salience of the applied stimulus in terms of intensity and duration. Whereas subliminal stimuli do not evoke a significant output signal (i.e., nonresponsive behavior, green area), supraliminal stimuli evoke a transient output signal that can be used for subsequent processing (i.e., transfer behavior, gray area). In the three-population model with default parameterization, the respective *perception threshold* (i.e., the border between nonresponsive and transfer/memory behavior) is located at a stimulation intensity of 78 s^{-1} . Salient stimuli activate the system (i.e., memory behavior, orange area) and make it insensitive to further stimuli. However, the system can be reset by means of impulses to the IIN (Fig. 4.1C). Together the observed response behaviors reflect the basic operations of signal flow

gating (combination of nonresponsive and transfer behavior) and working memory (memory behavior).

Figure 4.1B shows stripe-like patterns in the transition zone between areas of transfer and memory behavior. These stripes signify a dependence of the system's response on the stimulus' switch-off time relative to the phase of the system's intrinsic oscillations. This observation is explained in more detail below.

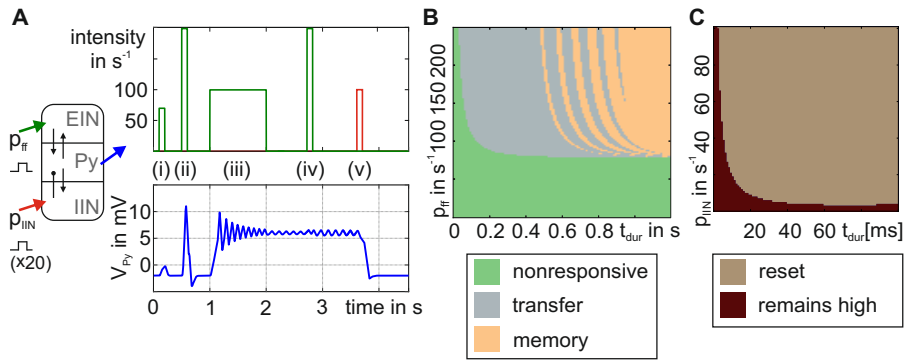


Figure 4.1: Response behaviors of the three-population model for feedforward stimulation. A) Depending on the salience of the applied stimuli in terms of duration and intensity, three distinct response behaviors are observable: (1) a nonresponsive behavior following weak and brief stimulation, where the Py 's membrane potential, V_{Py} , responds only with a small deflection below the firing threshold (impulse (i)), (2) a transfer behavior following a strong and brief stimulation where V_{Py} exceeds the firing threshold (impulse (ii)), and (3) a memory behavior following longer stimulation of medium intensity (impulse (iii)), for which the system settles down on a stable state of higher activity. In this state the system is insensitive to further stimuli, or noise, (impulse (iv)), but can be actively reset through a weak and brief impulse to the IIN, clearing the memory trace (impulse (v)). Please note that this IIN impulse is enlarged by a factor of 20 to improve clarity. B) The response behaviors depend on the salience of the input. Nonresponsive behavior is shown for intensities below $78 s^{-1}$ (green region). Exceeding this perception threshold, a longer stimulus reliably evokes a memory behavior (orange region). The shorter the stimulus the more likely becomes a transfer behavior (gray region). The stripe-like patterns signify a dependency of the behavior (transfer or memory) on the phase relation between stimulus switch-off time and the intrinsic system oscillation. C) Sufficiently long and strong impulses to the IIN (sandy area) reset a memorized stimulation, i.e. transfer the system from a high to a low activity – in contrast to brief and weak stimuli (brown area).

The observed behaviors are based on a rather simple mechanism in the system's state space. Figure 4.2A shows the fixed point curve of the steady state behavior of the three-population model with default parameterization and feedforward input p_{ff} as bifurcation parameter. Fold bifurcations are located at each turning point of

the fixed point curve – one of saddle-node type (unstable-stable) and one of saddle-saddle type (unstable-unstable). Further, a subcritical Hopf bifurcation is located at $p_{ff} = -5.9 \text{ s}^{-1}$. The resulting separatrix marks an unstable manifold that repels local trajectories in the state space close to it.

If no input is fed into the system ($p_{ff} = 0$), the system resides on the lower branch of the fixed point curve with $V_{Py} \approx -2 \text{ mV}$, see Figure 4.2A. For a weak impulse to the EIN ($p_{ff} < 78 \text{ s}^{-1}$) the system is not able to pass the lower fold bifurcation and settles down on the lower branch of the fixed point curve again, defining the nonresponsive behavior (Fig. 4.2B). The required intensity to pass this fold bifurcation reflects the perception threshold in the characteristic fingerprint (see Fig. 4.1B). For an input $p_{ff} > 78 \text{ s}^{-1}$ the system passes the lower fold bifurcation and oscillates around the upper branch of the fixed point curve (exceeding the firing threshold of 4 mV). A pair of complex conjugate eigenvalues with negative real part causes a damped oscillation. When the stimulus is switched off, the system's input returns to its original value ($p_{ff} = 0 \text{ s}^{-1}$). If the system's trajectory is located outside the Hopf separatrix (i.e., the system has not been damped sufficiently) the system settles down at the lower branch of the fixed point curve, defining the transfer behavior (Fig. 4.2C). If, however, the system's trajectory is located inside the Hopf separatrix (i.e., the oscillations have been damped sufficiently due to ample settling time) the system settles down on the upper branch of the fixed point curve, defining the memory behavior (Fig. 4.2D).

This insight explains the stripe-like patterns in Figure 4.1B. Notably, the Hopf separatrix is irregularly shaped in the six-dimensional state space. Hence, for distinct phases of the system's oscillation, the system's phase point is either inside or outside the basin of attraction of the Hopf separatrix when the stimulus is switched off (see Fig. 4.3). Hence, the stripes reflect the period of the system's oscillation. However, the longer the stimulation is switched on, the stronger is the system damped. This increases the likeliness of residing inside the Hopf separatrix (i.e., memory behavior) and causes the wider orange stripes for larger stimulus durations in Figure 4.1B.

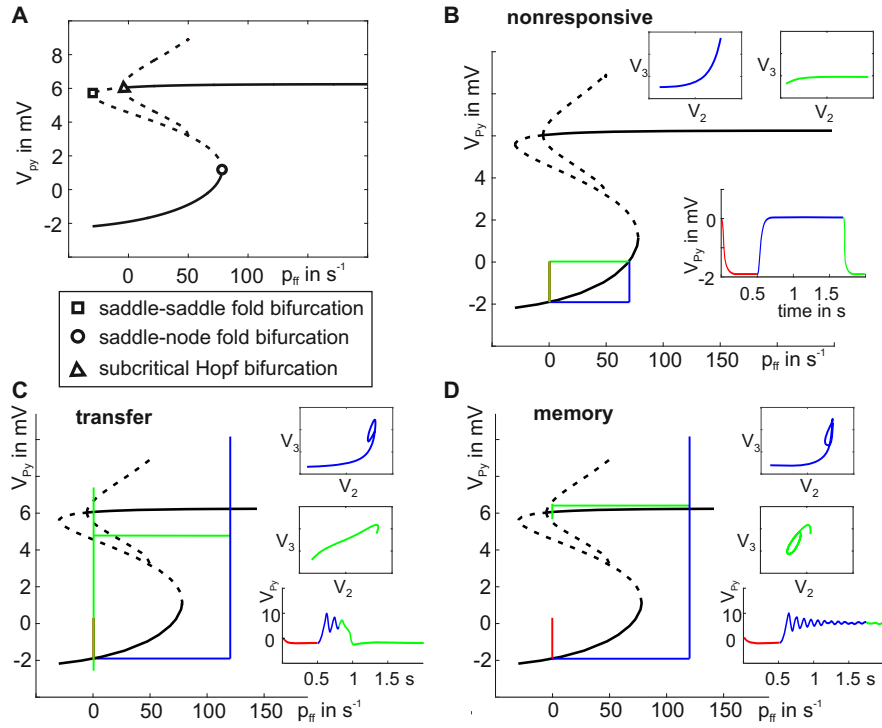


Figure 4.2: Dynamics of the distinct response behaviors in a projection of the state space. A) The S-shaped fixed point curve features stable (solid line) and unstable (dashed line) fixed points for varying input strengths to the EIN. Two fold bifurcations (saddle-node and saddle-saddle) and a subcritical Hopf bifurcation have been identified. B-D) Projections of the response behaviors in the bifurcation diagrams with inlets illustrating state space trajectories and the respective time courses: nonresponsive (B), transfer (C), and memory (D) behavior. Color-coding distinguishes prestimulus (red line), response (blue line), and asymptotic (green line) mean membrane potentials $V_{Py}(t)$.

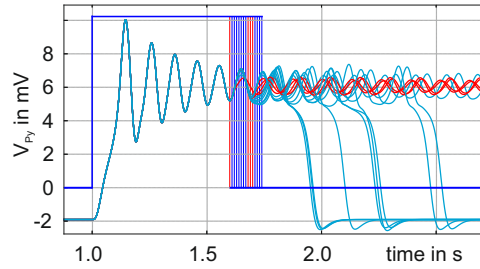


Figure 4.3: Phase-dependency between the system's intrinsic oscillation and stimulus switch off time. The diagram shows a collection of system responses to stimuli of distinct intensity (100 s^{-1}) and durations ranging between 600-750 ms. The stimulus offset times are marked by vertical lines. Blue lines denote stimuli and system responses for which the system returns to the lower activity state after the stimulus was switched off (i.e., transfer behavior). In this case the system's phase point was located outside the separatrix when the stimulus was switched off. Red lines denote stimuli and responses for which the system remains in the higher activity state (i.e., memory behavior). In this case the system's phase point was located inside the separatrix when the stimulus was switched off. Note that the two blue time courses were still high at the end of the time window but fell down eventually.

In summary, the hallmarks of the described mechanism are (i) a bistable activity of the Py population (high and low state), (ii) selectivity for salient stimuli (perception threshold and Hopf separatrix), (iii) a reduced sensitivity to further stimuli in the high state, (iv) relative robustness to noisy fluctuations in the external input in each state, and (v) a phase-dependency of the stimulus offset. Whereas certain phases allow the system to settle down in the high activity state other phases do not.

4.3 Network Balance Variation

In the following, I assess how an adjusted network balance, controlled by variations of the excitatory and inhibitory synaptic gains $H_{e,i}$, effects the described response behavior. For each variation of the excitatory and inhibitory synaptic gains $H_{e,i}$ the characteristic fingerprint was computed and summarized in a dynamic function map that was qualified in the state space.

The dynamic function map (Fig. 4.4A) charts the classified response behaviors in the parameter space spanned by H_e and H_i . The map shows parameterizations for which the system is dominated by the nonresponsive (bright green, anthracite and cyan regions), transfer (gray regions), and memory behaviors (orange and rose regions) as well as their compositions. The default parameterization ($H_e = 3.25$

mV and $H_i = 22$ mV) is located at the tip of a larger memory-dominated region. Moreover, this default parameterization adjoins a region dominated by transfer behavior. Such proximity of a system's state to major transition zones of the system's behavior (i.e., bifurcations) has been referred to as *criticality*. Criticality is considered beneficial for the system's information processing capacity, because small parameter changes may produce large changes in behavior [101, 102]. Hence, the local network balance is a very sensitive determinant of the system's criticality. Moreover, the local network balance controls the sensitivity of the system by regulating the distance between the working point and the lower fold bifurcation (i.e., the perception threshold, compare Figs. 4.4G and H). Around the default parameterization, decreasing H_e raised the perception threshold; meaning a reduced sensitivity to external stimuli. In turn, increasing H_e lowered the perception threshold; meaning an increased sensitivity to external stimuli. However, note the complex relationship: an exclusive increase of H_e results in a transfer-dominated behavior, whereas an increased sensitivity in favor of memory behavior requires a simultaneous decrease in inhibition (Fig. 4.4A).

In order to mechanistically understand how the network balance affects the response behaviors I evaluated the system's state space at its working point $p_{ff} = 0$, that is the input strength before and after stimulation. Keeping p_{ff} at zero, I systematically varied H_e and H_i and tracked the relevant bifurcations. In Figure 4.5 a light red background color denotes monostability but oscillations around the firing threshold, light blue denotes monostability and no oscillations around the firing threshold, and dark blue denotes bistability and no oscillations around the firing threshold at the system's initialization. The default parameter values for H_e and H_i are indicated on the axes. Note that the emerging graphs reflect the borders of the response behavior regions from Figure 4.4A.

The dark blue line indicates the lower fold bifurcation for $p_{ff} = 0$. Below the line this bifurcation is located at $p_{ff} > 0$ and above it at $p_{ff} < 0$ (compare Figs. 4.5C and D). Likewise, the cyan line indicates the upper fold bifurcation for $p_{ff} = 0$, which is located at $p_{ff} < 0$ above that line and at $p_{ff} > 0$ below the line (Figs. 4.5D and G). In consequence, only for the area between the two fold bifurcation branches, the point $p_{ff} = 0$ is located between the two fold bifurcations – a necessary condition for bistability at that point. It is a necessary prerequisite for the memory behavior that bistability existed without input. Therefore, the two fold bifurcation lines delimit the area in the $H_e - H_i$ plane where memory behavior is

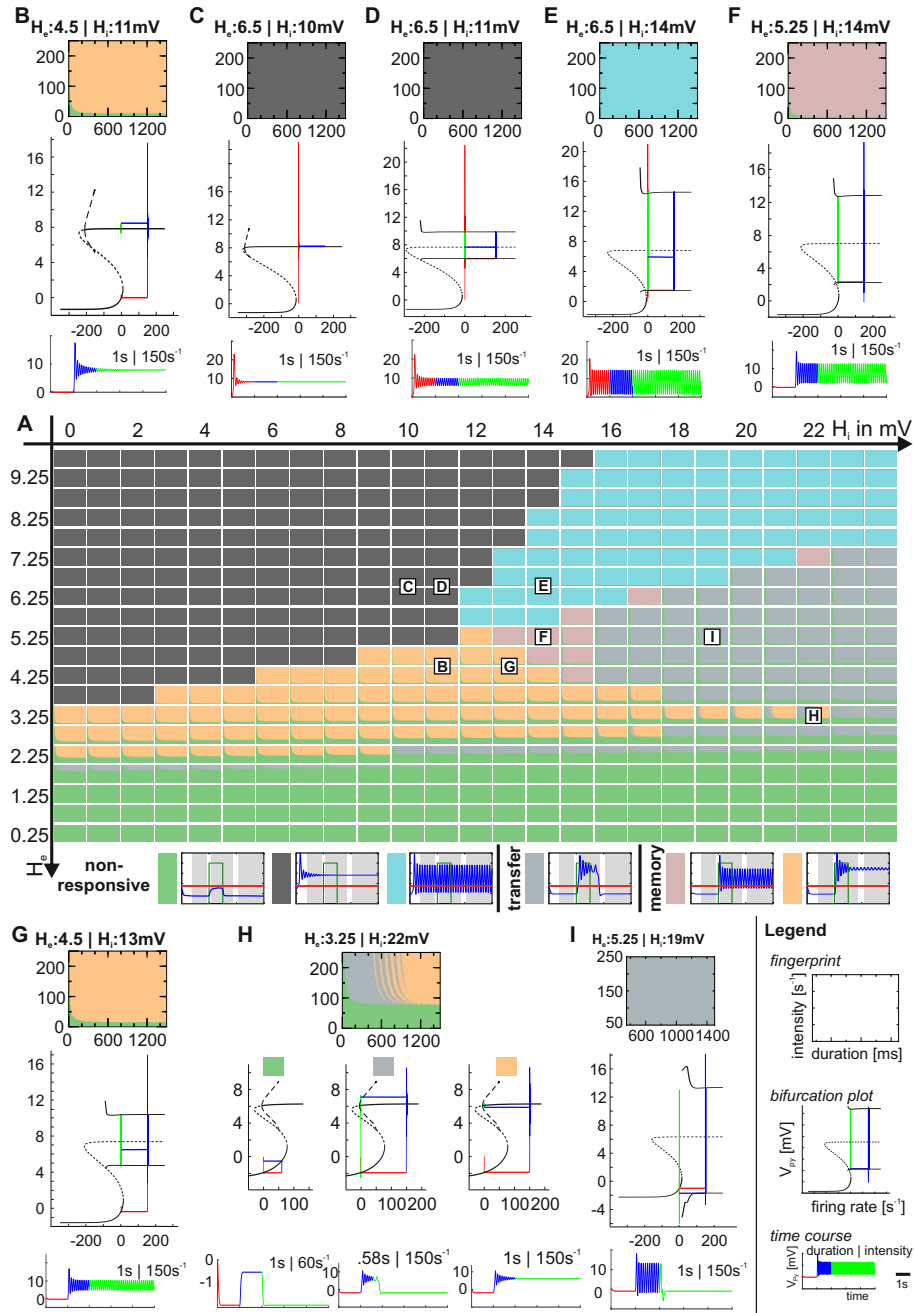


Figure 4.4: Dynamic function map for the three-population model with indirect excitatory local feedback architecture (Fig. 3.3B). A) Collection of characteristic fingerprints for varying excitatory (H_e) and inhibitory (H_i) synaptic gains. Colors code the observed response behaviors: nonresponsive (bright green, anthracite and cyan regions), transfer (gray regions), and memory (orange and rose regions). The local network balance controls the dominance of the behaviors and tunes the criticality of the system. B-J) Exemplary parameterizations featuring fingerprints, time courses, and projections thereof in a bifurcation plot. Time course color codes are similar to Figure 4.2.

possible. This is in agreement to the observed behavior shown in Figure 4.4A. Further, between those lines the distance between the upper and lower fold bifurcations on the p_{ff} axis (smaller subplots in Fig. 4.5) determines the robustness of the system to noise by scaling the width of the bistability. Moreover, the system's location relative to the lower fold bifurcation determines its sensitivity to stimuli.

However, not the entire area between the two fold bifurcation branches actually exhibits bistability. As H_i was increased at some point, the upper fold bifurcation (saddle-node bifurcation, solid cyan line in Fig. 4.5A) splits into a fold bifurcation between two unstable fixed points (saddle-saddle bifurcation, dashed cyan line) and a subcritical Hopf bifurcation (dashed orange curve). To the left of that curve the subcritical Hopf bifurcation occurs at $p_{ff} < 0$ leading to bistability at $p_{ff} = 0$ in form of a stable focus on the upper branch of the fixed point curve. To the right of that curve the Hopf bifurcation is located at $p_{ff} > 0$. Consequently, the upper branch of the fixed point curve is unstable for $p_{ff} = 0$ and bistability is abolished (Fig. 4.5B). At some point along this subcritical Hopf bifurcation branch the subcritical Hopf bifurcation turns into a supercritical Hopf bifurcation (solid purple curve in Fig. 4.5A). Instead of the stable focus it is then the stable limit cycle, associated with the supercritical Hopf bifurcation, that accounts for the bistability at $p_{ff} = 0$. In consequence, bistability also occurs along a narrow strip on the right side of the supercritical Hopf bifurcation branch in Figure 4.5A with an oscillatory upper state.

The mentioned subcritical and supercritical Hopf bifurcations collide at positive/negative values of p_{ff} within a small region close to the dashed/solid purple line, respectively. This collision leads to a stable limit cycle that surrounds small bits of the overlapping part of the fixed point curve (e.g., Fig. 4.4F), giving rise to bistability and thus memory behavior. In this case, the system either oscillates around the firing threshold (classified rose in Fig. 4.4F) or above the firing threshold of 4 mV (classified orange in Fig. 4.4G), both indicating memory behavior. For a slightly higher H_i , the stable limit cycle reaches just up to the lower fold bifurcation, abolishing bistability and leading to transfer behavior (see Fig. 4.4I).

In summary, the dark blue area and small parts of the light blue area in Figure 4.5 account for bistability and allow for memory behavior, as corroborated by Figure 4.4. For a tuning of the network balance outside of this bistable region, the canonical microcircuit model does not feature the memory behavior anymore and, thus, loses an integral part of its basic operations.

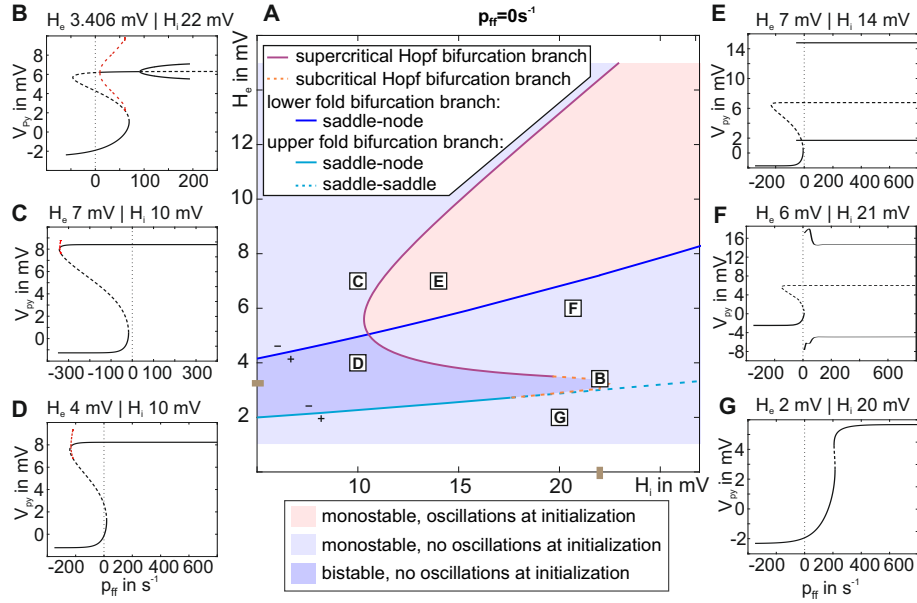


Figure 4.5: Two-parameter bifurcation plot of the three-population model. A) Identified bifurcations at $p_{ff} = 0$ are tracked through the parameter space with respect to H_e and H_i . The background is colored light red for oscillating behavior in the low state at $p_{ff} = 0$, light blue for non-oscillatory behavior and monostability, and dark blue for no oscillations and bistability. Brown marks at the axis denote the default parameters for H_e and H_i . The solid purple line denotes the supercritical Hopf bifurcation branch. The dashed orange line denotes the subcritical Hopf bifurcation branch that is important for the transfer behavior of the system. Together, these branches mark the border between dominant regions of memory and transfer behavior (compare Fig. 4.4). Upper (cyan line) and lower (blue line) fold bifurcations and the Hopf bifurcation branches delimit a region that exhibits bistability (enabling memory behavior). Signs $+$ and $-$ indicate whether the particular fold bifurcation is located at positive or negative values of p_{ff} . B)-G) Bifurcation diagrams characterizing stable and unstable fixed points for a range of p_{ff} values. H_e and H_i tune robustness and sensitivity of the system to external stimulation.

4.4 Variations in Local Topological Feedback

4.4.1 Transformation of the Three-population Model to the Two-population Model

In the following subsections I repeat the analysis of the stimulation-induced response behaviors for the other local feedback topologies, namely the two-population models with and without inhibitory self-feedback.

To this end, the three-population model with indirect excitatory local feedback (Fig. 3.3B) was gradually transformed into a two-population model with direct excitatory local feedback (Fig. 3.3D) by mapping two excitatory populations into a single one, see Section 3.1.2. In the canonical microcircuit model this transformation meant to decrease the architectural parameter b_1 from one to zero and to adopt the connectivity weights. Accordingly, the strength of the excitatory self-feedback of the pyramidal cells (N_{PP}) increases, whereas the strength of the local feedback from the excitatory interneurons to the pyramidal cells (N_{PE}) gradually vanishes. Moreover, extrinsic feedforward input (p_{ff}) progressively arrives at the pyramidal cells. Notably, for values $0 < b_1 < 1$ direct and indirect excitatory local feedback coexist. Figure 4.6 shows the consequent changes of this transformation in the system's state space. Strengthening the self-feedback of the Py moves a subcritical Hopf bifurcation to larger p_{ff} values and a supercritical Hopf bifurcation to lower p_{ff} values (Fig. 4.6C) and eventually leads to their collision. Moreover, the increasing self-feedback of the Py moves the upper fold bifurcation to positive p_{ff} values and decreases its distance to the lower fold bifurcation. Lower and upper fold bifurcation collide in a cusp bifurcation (Fig. 4.6E), straighten the fixed point curve, and abolish bistability. In the extreme case of exclusive self-feedback of the Py (Fig. 4.6G), there is only a very small range of bistability – far off the working point. There, two stable states are present, namely a stable limit cycle that encompasses a stable focus.

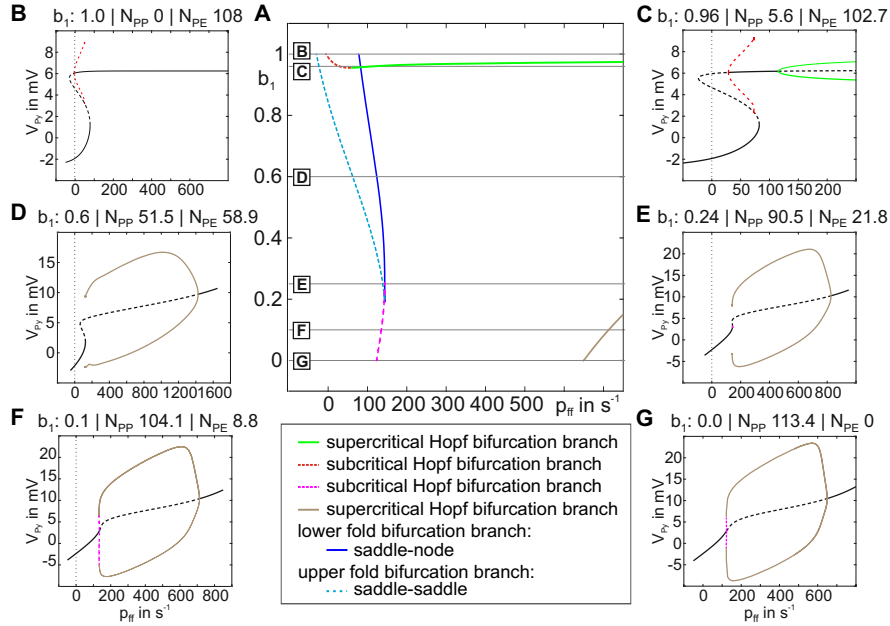


Figure 4.6: Gradual transformation of the indirect excitatory local feedback to a direct excitatory local feedback. A) The two-parameter bifurcation plot tracks bifurcations along p_{ff} when b_1 was changed from 1 (three-population model) to 0 (two-population model). For this mapping, the network balance is held constant at values $H_e = 3.25$ mV and $H_i = 22$ mV. B-G). The single parameter bifurcation plots show the fixed point curve (V_{py}) and local bifurcations along p_{ff} for different values of b_1 (and accordingly N_{PP} and N_{PE}).

4.4.2 Direct Excitatory Local Feedback Architecture

The architectural parameters of the obtained two-population model without disinhibition were $b_1 = 0$, $b_2 = 1$, $b_3 = 0$ (see Fig. 3.3D). Notably, this two-population model does not divide the excitatory neurons into an input and output layer. As shown in Figure 4.6, the bistability around the working point – based on the S-shaped fixed point curve – vanishes when the indirect excitatory local feedback was mapped to direct excitatory local feedback with default values for H_e and H_i .

The dynamic function map for the direct excitatory local feedback architecture was computed and is depicted in Figure 4.7A. As no oscillatory behavior is observed around the working point of $p_{ff} = 0$, this map exhibits fewer variants as compared to the indirect feedback architecture (Fig. 4.4), but still contains all types of the classified response behaviors. However, no parameterizations for which all three types of response behaviors are robustly present in a single fingerprint are observed. Memory and nonresponsive behaviors still occur for single parameterizations and

depend on the intensity, but not on the duration of the stimulus anymore. Moreover, the system does not respond with a damped oscillation to external input anymore, but settles down immediately. Like in the case of the indirect excitatory local feedback, the system, in general, is more sensitive to changes in the excitatory synaptic gain than to changes in the inhibitory synaptic gain.

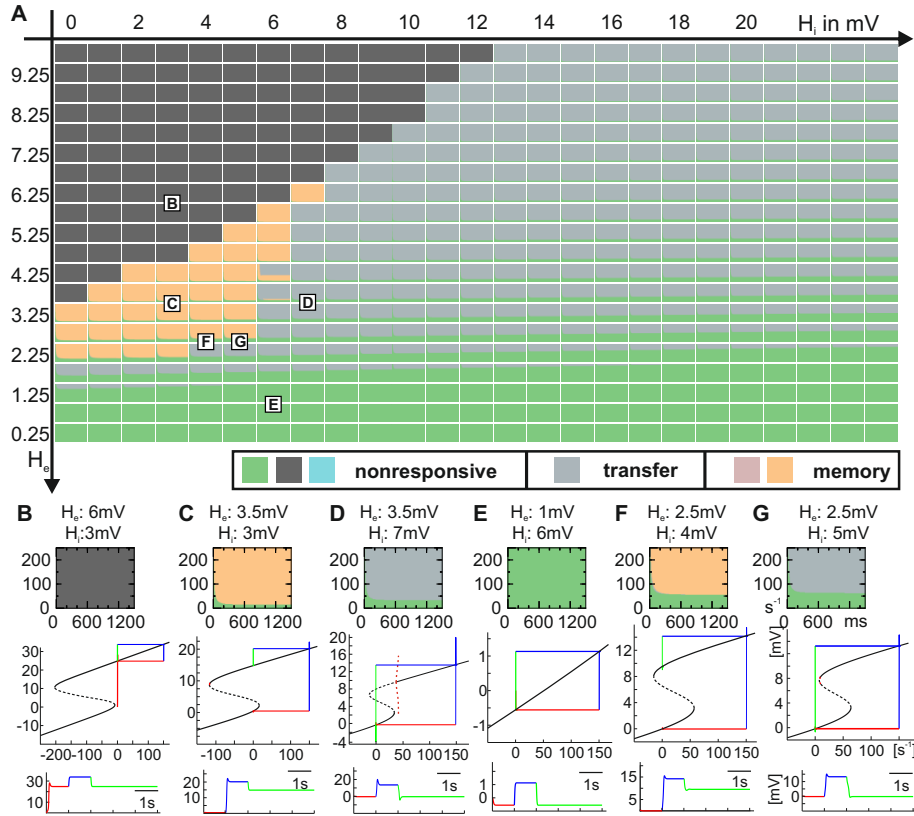


Figure 4.7: Dynamic function map for the direct excitatory local feedback architecture (Fig. 3.3D). A) Collection of characteristic fingerprints for varying excitatory (H_e) and inhibitory (H_i) synaptic gains. Colors code the observed response behaviors: nonresponsive (bright green and anthracite), transfer (gray), and memory (orange). The variety of observed behaviors is reduced compared to the three-population case (Fig. 4.4). Although all three response behaviors are present they do not occur for a single value of the local network balance. B)-G) Selected parameterizations featuring fingerprints, time courses, and projections thereof in bifurcation plots.

As for the indirect excitatory local feedback architecture, I examined the system's state space for bifurcations at the working point $p_{ff} = 0$, see Figure 4.8A. Again, the bifurcation branches reflect the borders of dominant response behaviors in the dynamic function map (Fig. 4.7A).

As described for the indirect feedback architecture, bistability and memory behavior are generally delimited by the parts of upper and the lower fold bifurcation branches that reflect saddle-node bifurcations. Also similar to the indirect feedback case: at about $H_i = 6$ mV the upper fold bifurcation (solid cyan) split into a saddle-saddle (dashed cyan) and a subcritical Hopf bifurcation (dashed purple), see Figure 4.8A. Bistability at the working point exists only for $H_e - H_i$ values left of that Hopf bifurcation branch, compare Figures 4.8C and D. In general, the membrane potentials in the high state are quite high (>10 mV) and close to the saturation threshold of the sigmoid function at about 14.8 mV. Hence, the efferent firing rates of such high membrane potentials led to almost invariant transmitted firing rates of 5 s^{-1} . In consequence, graduated efferent firing rates, as observed of the indirect excitatory local feedback architecture, are exceptional.

In summary, the two-population model without inhibitory self-feedback exhibits all types of response behaviors but for lower ranges of the inhibitory synaptic gain. A concomitant existence of all types of response behaviors for a single value of the local network balance is not observed.

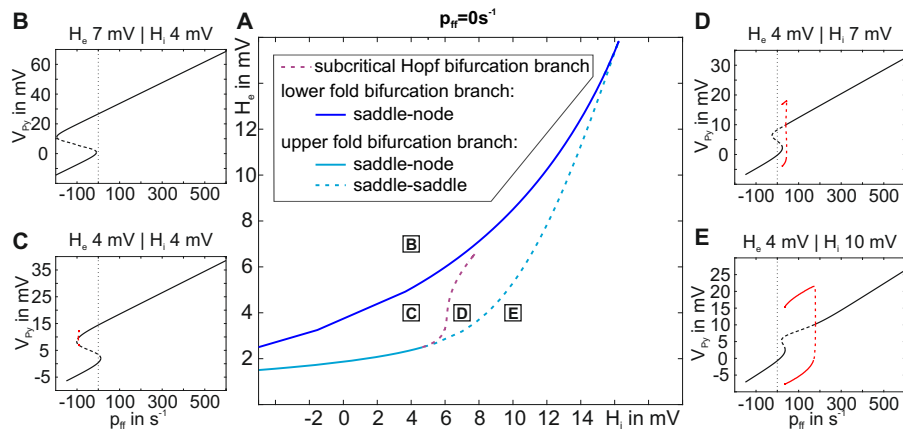


Figure 4.8: Bifurcation plot of the two-population model with direct excitatory local feedback. A) Identified bifurcations at $p_{ff} = 0$ are tracked through the parameter space with respect to H_e and H_i . The upper (cyan line) and lower (blue line) fold bifurcation branch delimit the region for which a bistable fixed point curve occurs. However, the subcritical Hopf bifurcation (purple line) renders parts of the fixed point curve unstable and prevents an actual bistability at $p_{ff} = 0$. This suppresses a memory behavior in favor of a transfer behavior (see Fig. 4.7). B-E) The single parameter bifurcation plots show the fixed point curve (V_{py}) and local bifurcations along p_{ff} for different values of H_e and H_i .

4.4.3 Introduction of Disinhibition in the Two-population Model

An often neglected property of neural model architectures is the existence of recurrent inhibitory collaterals, leading to inhibitory self-feedback and resulting in a disinhibition effect. Whereas inhibition dampens an excited system, disinhibition reduces this damping and makes the inhibitory feedback path less effective. Decreasing the architectural parameter b_2 to zero and keeping $b_1 = 0$ introduces self-feedback of the inhibitory interneuron population in the canonical microcircuit model and yields a two-population model with direct excitatory local feedback and recurrent inhibitory feedback of strength N_{II} (Fig 3.3E).

Figure 4.9 reports the consequent state space changes and bifurcations when N_{II} was continuously increased. Weak inhibitory self-feedback moves the supercritical Hopf bifurcation (green line in Fig. 4.9A) to lower values of p_{ff} . Stronger inhibitory self-feedback reintroduces bistability around the working point ($p_{ff} = 0$) via an S-shaped fixed point curve as consequence of a cusp bifurcation that generates two fold bifurcations – one of saddle-saddle type (upper fold bifurcation) and one of saddle-node type (lower fold bifurcation). For $N_{II} > 10$ the bistable part of the fixed point curve increases because the upper fold bifurcation moves to lower values of p_{ff} . For very strong inhibitory self-feedback the fixed point curve, characterized by the two fold bifurcations and the generated firing rates at the pyramidal cell population (via the activation function), converge to a distinct form independent from the network balance parameters (Fig. 4.9F). Since the dynamics do not change significantly for large levels of disinhibition, I fixed N_{II} to $N_{II} = 33.25$, similar to the level of inhibitory connections that target excitatory populations.

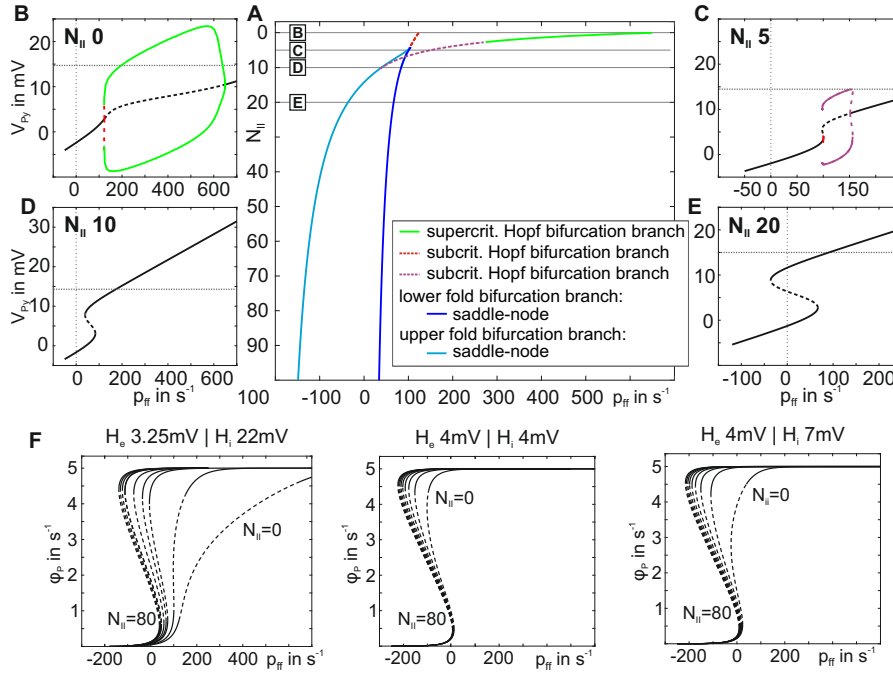


Figure 4.9: Increasing recurrent inhibitory self-feedback in the two-population model. A) The two-parameter bifurcation plot tracks the occurring bifurcations along p_{ff} when the recurrent inhibitory self-feedback N_{II} was increased. The network balance was held constant at values $H_e = 3.25$ mV and $H_i = 22$ mV. B-E) The single-parameter bifurcation plots show the fixed point curve (V_{Py}) and local bifurcations along p_{ff} for different values of N_{II} . F) Fixed point curves (p_{ff} , mean firing rate $\phi_P(t)$) for different values of the local network balance.

4.4.4 Direct Excitatory Local Feedback Architecture with Disinhibition

The dynamic function map for the direct excitatory local feedback architecture with inhibitory recurrent feedback (Fig. 3.3E) was computed and is shown in Fig. 4.10A. Again, all types of the classified response behaviors are observed. For some parameterizations two behaviors occur in dependence on the stimulus' intensity. The memory behavior features membrane potential that range from high values ($H_e = 5$ mV, $H_i = 18$ mV), putting the excitatory local feedback path into saturation (Fig. 4.10D), to lower values ($H_e = 2.5$ mV, $H_i = 18$ mV), avoiding a saturation (Fig. 4.10E). Again, the system is more sensitive to changes in the excitatory synaptic gain than to changes in the inhibitory synaptic gain. Bifurcations at

the working point $p_{ff} = 0$ reflect the borders of the dominant response behaviors in the dynamic function map (Fig. 4.11). Again, the upper and lower fold bifurcations delimit the multistability range of the fixed point curve. The upper fold bifurcation remains a saddle-node bifurcation for increasing inhibitory synaptic gains and preserves stability of the upper part of the fixed point curve. This was in contrast to both alternative topologies, for which the saddle-node bifurcation splits into a saddle-saddle and a subcritical Hopf bifurcation (Figs. 4.5 and 4.8). The preserved bistability and the absence of a separatrix (filtering out brief stimuli, see Fig. 4.5) explains the considerably larger range of working memory in the two-population model with disinhibition. In terms of signal flow gating, the two-population model with disinhibition filters stimuli according to their intensity, but disregards their temporal consistency and transiency.

In summary, disinhibition promotes saturation of the excitatory local feedback path and leads to stationary dynamics for which the membrane potential of the Py rises in linear dependence on the external input. This effect is even stronger for high excitatory (increase in positive feedback) and inhibitory (increase of disinhibition) synaptic gains.

Refer to Appendix A.2 for supplementary information on how the variation of synaptic gains modulated the state space in the considered local feedback topologies.

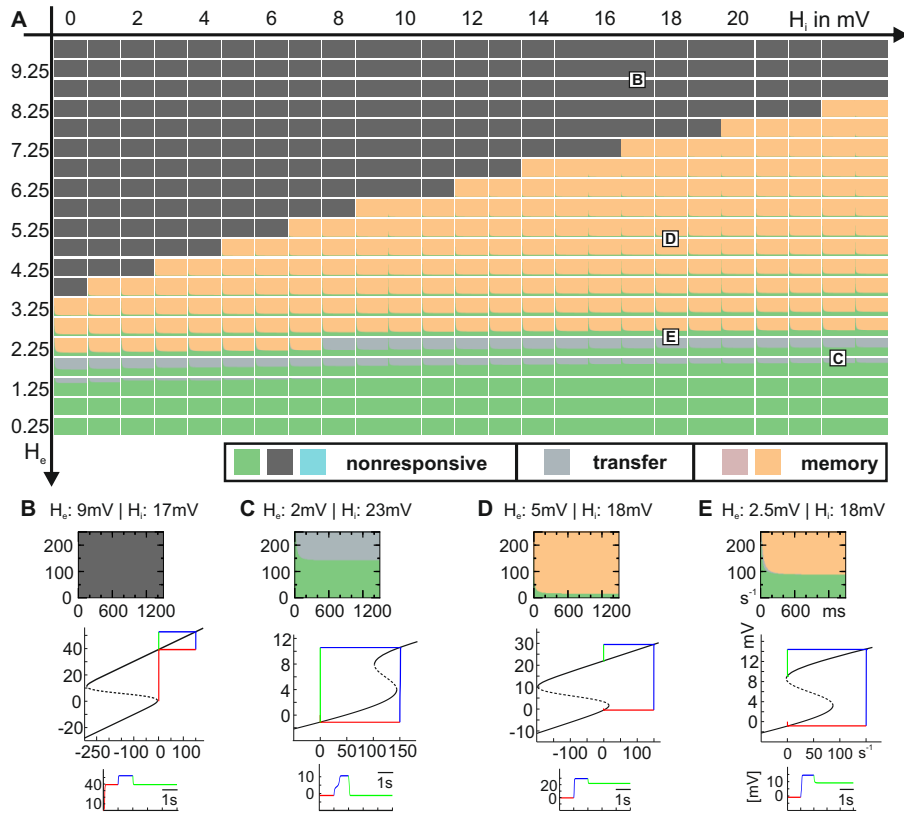


Figure 4.10: Dynamic function map for the two-population model with disinhibition (Fig. 3.3E). A) Collection of characteristic fingerprints for varying excitatory (H_e) and inhibitory (H_i) synaptic gains. Colors code the observed response behaviors: nonresponsive (bright green and anthracite), transfer (gray), and memory (orange). The variety of observed behaviors is reduced compared to the three-population case (Fig. 4.4). However, all three response behaviors are observable. B-E) Selected parameterizations featuring fingerprints, time courses, and their projections in a bifurcation plot.

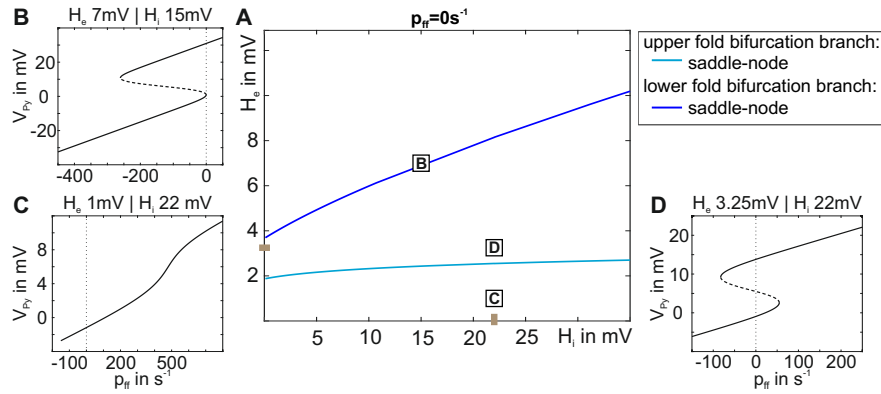


Figure 4.11: Two-parameter bifurcation plot for the two-population model with disinhibition. A) Identified bifurcations at $p_{ff} = 0$ are tracked through the parameter space with respect to H_e and H_i . The region between the upper (cyan line) and lower (blue line) fold bifurcation delimit the parameter range for a bistable fixed point curve. These bifurcation branches reflect the borders of nonresponsive, transfer, and memory behavior in Fig. 4.10. B)-D) The single parameter bifurcation plots show the fixed point curve (V_{Py}) and local bifurcations along p_{ff} for distinct values of the local network balance.

4.5 Response to Extrinsic Feedback Input

So far, I analyzed the response behavior of the canonical microcircuit model when it is stimulated by extrinsic feedforward input p_{ff} . In the following, I extend this analysis to extrinsic feedback input p_{fb} that – similar to lateral input – arrives at the pyramidal cell population [25]. In the following, I only consider the three-population model (i.e., $b_1 = b_2 = b_3 = 1$, see Fig. 3.3C) that features two separate populations to receive feedforward and feedback information. This is in contrast to the two-population models. Note that the three-population model with input to its pyramidal cell population reflects a model configuration that has been intensively investigated before [70, 94, 95], though not systematically in terms of transient response behaviors.

Rectangular bursts with a distinct intensity and duration were applied to the Py and the respective response behaviors (nonresponsive, transfer, and memory) were mapped to characteristic fingerprints. In contrast to feedforward stimulations (see Section 4.2), these feedback stimulations lead to large ranges of nonresponsive behavior and transient behavior that embeds sparse traces of memory behavior (Fig. 4.12A). These response behaviors indicate distinguishable operational roles for the separate input channels. Whereas both channels are able to gate information via a

perception threshold (green-gray border in Figs. 4.1B and 4.12A), only the feed-forward input channel (EIN) allows the storing of information. This differentiated processing documents the model's capability to react differently to afferent information streams.

As before, I analyzed the underlying state space by means of a bifurcation diagram to explain the observed response behaviors. Compared to the feedforward stimulation of EIN (Fig. 4.2A), the state space features two stable limit cycles that add sustained oscillations to the dynamical repertoire of the canonical microcircuit model (Fig. 4.12B). The large amplitudes of these oscillations prevent the system from settling down inside the separatrix (arising from a subcritical Hopf bifurcation) when the stimulation ends ($p_{fb} = 0$). This difficulty to express a high activity state causes the near absence of memory response behaviors in the characteristic fingerprint (Fig. 4.12A).

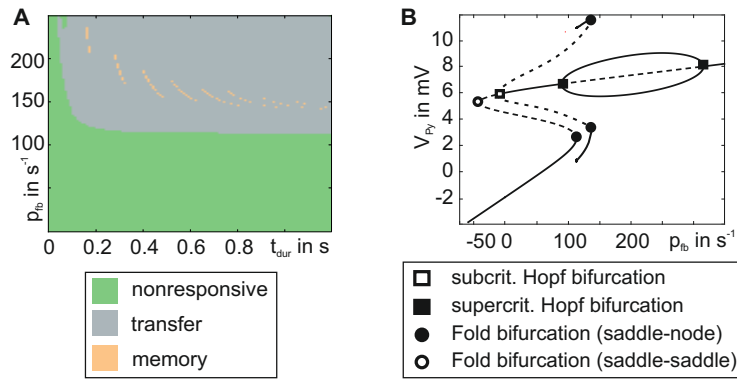


Figure 4.12: Characteristic fingerprints for transient stimulation of Py. A) In the the three-population model a transient Py stimulation (feedback information) evokes transient sub-threshold deflections (nonresponsive behavior, green) for weak stimuli, transient supra-threshold deflections (transfer behavior, gray) for strong stimuli, and sparse traces of sustained supra-threshold deflections (memory behavior, orange). B) The S-shaped fixed point curves with stable (solid line) and instable (dashed line) sections features two fold bifurcations (saddle-node and saddle-saddle) at the turning points of the fixed point curve. Moreover, a subcritical Hopf bifurcation gives rise to an unstable manifold (separatrix) and a stable limit cycle via another fold bifurcation. Two more supercritical Hopf bifurcations establish another stable limit cycle.

4.6 Discussion

The analysis of stimulus-driven information processing in the canonical microcircuit model corroborates the ability for basic operations, namely signal flow gating and working memory. It also shows that these are based on three distinct response behaviors. The performed state space analysis identifies a bistable activity of the Py population and the ability for stimulus saliency detection (via an intensity threshold and a Hopf separatrix) as necessary conditions to support these basic operations. The examination of different local feedback topologies shows that only the three-population model exhibits the coexistence of all three response behaviors. Hence, the three-population model can selectively block, transmit or memorize a stimulus based on its properties. Only this three-population architecture meets the demands of a minimal canonical microcircuit (see Section 2.4). Moreover, the analysis shows that a variation of the local network balance effectively individualizes the local functionality of the canonical microcircuit model and that the three-population model reacts differently to hierarchical information streams. These results are now discussed in more detail.

Canonical microcircuits have been previously associated with basic operations, or stereotypic functions [51]. The presence of such operations has been investigated in large-scale spiking neuron networks [56, 57, 64, 103] as well as in mean field models [4, 20, 104], which form the basis of the majority of attempts to model neurocognitive experiments (e.g., DCM; [5, 58]). In particular, neural mass models [1, 70] suit the examination of canonical microcircuits as they capture the mesoscopic spatial scale of neuronal populations, at which the uniformity of canonical microcircuits is arguably established. Notably, the three-population architecture that is analyzed here has already been studied for its local steady-state system behavior [70, 94, 95]. However, the explicit demonstration of basic operations at the level of a canonical microcircuit as well as their sensitivity to topological features and levels of excitation and inhibition is original for this type of neural mass model.

In contrast to more detailed computational models [56, 57, 64, 103] the present canonical microcircuit model emphasizes the canonicity of a neural circuit not in the strict reproduction of the laminar architecture of a cortical column but rather in the minimal realization of internal positive and negative feedback loops. These simple feedback loops give rise to the meaningful basic operations of signal flow gating and working memory. In the canonical microcircuit model, the local mem-

ory behavior is based on bistability (of the Py population), which is one mechanism of short-term (or working) memory. The recurrent and self-sustaining activity depends on the network balance and is quickly initiated and terminated by afferent inputs, a behavior that has been shown *in vitro* [76]. In contrast to long-term memory that rests on synaptic plasticity, short-term (or working) memory is thought to rely on mechanisms that do not change the underlying connectivity structure. Besides bistability, working memory has been proposed to be based on delays and time constants [67, 105], such as through *synfire chains* [33], recurrent excitatory networks [40] or cellular properties [106]. In these mechanisms the period of storage (forgetting time) depends on relatively fixed structural aspects of the network, whereas with bistability the item can be kept in memory in principle for any period until it is actively switched off.

In regard to signal flow gating, the canonical microcircuit employs an intensity threshold (i.e., the perception threshold denoting the distance between working point and lower fold bifurcation) and a Hopf separatrix (evaluating the input's temporal consistency and transiency) to select input signals in dependence on their respective properties. Notably, signals can be distinguished from noise by their higher amplitude and temporal smoothness. Thus, the canonical microcircuit model is robust to noise but sensitive to afferent input at the same time. It is shown that the intensity threshold depends on the local network balance, which is therefore a key parameter for governing the tradeoff between robustness to noise and sensitivity to stimuli.

Whereas all considered local feedback topologies feature this intensity threshold, only the three-population model concurrently features the Hopf separatrix. Hence, the transfer behavior of the two-population models only depends on the strength of the applied stimuli and not on their duration, as opposed to the three-population model, where it depends on both. Consequently, only the three-population model exhibits the coexistence of all three response behaviors for a fixed value of the network balance. Thus, through nonrestrictive support of the basic operations of signal flow gating and working memory, the three-population model is the only of the considered architectures that represents a *minimal canonical microcircuit*⁷. In this sense, the consideration of an indirect excitatory feedback loop of the Py

⁷ A minimal canonical microcircuit is defined as a canonical microcircuit model that (i) considers interacting excitatory and inhibitory neuronal populations, (ii) whose structural topology reflects the homogeneity of neocortical matter, and (iii) whose functional repertoire features the basic operations of signal flow gating and working memory (Section 2.4).

population increases the diversity and biological realism of the model's dynamics in very important ways. Notably, the present canonical microcircuit model disregards a direct interaction between the excitatory and inhibitory interneurons as well as the self-feedback among the excitatory interneurons. Both of these connections certainly exist [56]. These features are among the important extensions of the model that need to be investigated in future studies.

With this work I follow the notion that the brain's information processing rests on a hierarchically organized network structure across multiple organizational scales [107]. The concept of canonical microcircuits mostly refers to the level of cortical columns [50, 56]. However, the canonical microcircuit model considered here emphasizes positive and negative feedback loops and can thus be regarded as valid on multiple spatial levels of brain networks, ranging from single neurons to entire cortical areas. Accordingly, I interpret the results as not specific for a distinct type of neocortex [12], or for elaborated architectures [56]. Though not trivial, I consider it part of a set of generic neural mechanisms of the brain's processing.

Besides the finding that the local feedback topology constrains local functionality, the analysis shows that the local network balance is a biologically plausible means to individualize local functionality. In particular, the network balance determines the dominant response behavior of a canonical microcircuit, that is whether it selectively forwards an impulse (transfer behavior), switches into a higher persistent activity (memory behavior) or does not respond to the impulse at all (non-responsive behavior). Note that in this case the local network balance is treated as a lumped parameter that actually encapsulates a large number of physiological and anatomical properties and mechanisms, such as neurotransmitter kinetics and neuroreceptor densities, dendritic arborization, synaptic and extra-cellular ionic dynamics as well as local and non-local network connectivity [108, 109].

Cell assemblies are proposed to process information through auto-associations, hetero-associations, and conditioned associations (see Section 2.1). Auto-association (a self-projection that enhances a localized activation when the cell assembly is stimulated) is represented in minimal canonical microcircuits through the basic operation of working memory – corresponding to a persistent increased activity following adequate stimulation. Hetero-association (autonomous sequential activation of two or more cell assemblies) can be achieved in minimal canonical microcircuits through cascading them with strong connectivity gains. A persistent activity in one microcircuit automatically leads to a persistent activity

in a connected circuit (similar to a synfire chain). Moreover, cascaded minimal canonical microcircuits can account for conditioned associations of cell assemblies for which an additional input is required to gate the activation of a cell assembly. The simplest form to do this is *feedforward conditioning*. In feedforward conditioning an afferent input pre-activates a minimal canonical microcircuit and moves the system close to – however below – its perception threshold. The presence of this pre-activation then conditions the effect of received activity from a cascaded minimal canonical microcircuit. This functional compatibility of cell assemblies and minimal canonical microcircuits substantiates the hypothesis that minimal canonical microcircuits may constitute neurobiological plausible nodes in models of distributed assembly networks (see Section 2.4).

Interestingly, the finding that the three-population model reacts differently to feedforward than to feedback information indicates distinguishable operational roles for the separate input channels. Although both channels gate information via a perception threshold, only the feedforward input channel (EIN) allows storing of input information. The following chapter further investigates the role of feedback signals and examines the interaction of multiple canonical microcircuit models.

Chapter 5

State-dependent Processing in Interacting Canonical Microcircuit Models

“The process of identifying unsuspected levels of organization situated between existing levels and conceptualizing the types of operations that occur there is a recurring step in the development of mechanistic explanation in the life sciences.”

– William Bechtel, 2005

5.1 Introduction

Hierarchically interacting canonical microcircuits simultaneously receive feedforward and feedback information, effectively integrating low-level sensory and high-level conceptual information. Accordingly, extending the analysis of a single canonical microcircuit model, this chapter examines the hierarchical interaction of canonical microcircuit models, in particular the three-population model (Fig. 3.3C). Note that the following analysis uses the term *minimal canonical microcircuit* synonymously to this three-population model. This is done because the analysis accounts for the three-population model, though the findings are thought to hold for minimal canonical microcircuits in general. The three-population model allows the separate application of feedforward and feedback information and is shown to

respond differently to both information streams, which is a necessary requirement for hierarchical communication.

Various computational models have been proposed that investigate the cooperation of multiple canonical microcircuit models to realize higher cognitive functionality, such as attentional processing [87], predictive coding [86], language processing [4, 20], and visual steering [88]. Extending this research, I investigate how the neural operations of priming and structure-building emerge from the hierarchical interaction of minimal canonical microcircuits.

Priming refers to the influence of past experience on the current processing performance [19, 110, 111]. In particular, priming reflects the facilitation of perception, attention, and responding as evident through increased stimulus identification probabilities or decreased response latencies. Due to shared characteristics, priming is considered an unconscious form of human memory but operates independently from explicit memory systems, such as procedural and semantic memory [110, 111]. Priming is a phenomenon that is difficult to localize, occurs for various sensory modalities, and modulates processing on multiple levels of perception [19, 110]. The neural mechanistic underpinnings of this multi-scale processing principle remain elusive. Some evidence is provided by visual search tasks that suggest priming to share characteristics of top-down attentional guidance and that “priming patterns may operate as a memory system for the deployment of attention” [19].

Structure-building refers to the gradual propagation of neural activity that establishes patterns in time and space (see Section 2.1). Similar to priming, structure-building involves some sort of *conditioning* in the sense that *a priori* information unlocks subsequent processing steps. This notion of conditioning indicates a containment of subsequent processing steps (in terms of their probabilities) and thus predicts them. This predictive processing is believed to be a fundamental processing principle also known as *predictive coding* [60, 112]. Generally, predictive coding suggests that the brain is continuously predicting future states based on an internal model of the environment. This internal model integrates novel sensory and established conceptual information via forward and backward connections from different levels of the cortical hierarchy [60, 86, 113, 114].

The following analysis aims at the depiction of composite neural operations. Firstly, the analysis substantiates the modulatory character of feedback signals and shows how facilitative feedback signals establish state-dependent operations – effectively constraining the access to the basic operations. Subsequently, prototypical meta-circuits that describe the minimal hierarchical interaction of canonical microcircuit models are shown to support the neural operations of priming and structure-building.

5.2 Simultaneous Stimulation Modulates Response Behavior

In order to evaluate the characteristic behavior of minimal canonical microcircuits that receive hierarchical information flows, both feedforward and feedback inputs were simultaneously applied to a single minimal canonical microcircuit.

Figure 5.1A displays a two-parameter bifurcation diagram of the canonical microcircuit model with constant feedforward (p_{ff}) and feedback (p_{fb}) input as bifurcation parameters. Color-coded sections reflect monostable (yellow and gray) or bistable (green) states of the model. Horizontal cross sections reflect single-parameter bifurcation diagrams. For example, the solid line at $p_{fb} = 0$ reflects pure feedforward input (left panel of Fig. 5.1B). Vertical cross sections reflect single-parameter bifurcation diagrams. For example, the solid line at $p_{ff} = 0$ reflects pure feedback input (left panel of Fig. 5.1H). At the crossing of both solid lines (the default working point) the model features bistability with constant output (light green section). For no feedback input, the subcritical Hopf bifurcation (dashed orange line) and the lower fold bifurcation (solid blue line) delimit the system's bistability. Increasing the feedback input ($0 < p_{fb} < 60$) shifts the bistable region to lower values of p_{ff} and widens the Hopf separatrix at the working point $p_{ff} = 0$ (compare Figs. 5.1B and E). This causes a lower perception threshold, an increased robustness of bistability around the working point, and a lower required stimulation duration for the phase point to remain within the separatrix when the stimulation ceases. For a further increase of p_{fb} the subcritical Hopf bifurcation splits into a supercritical Hopf bifurcation (orange solid line) and a fold bifurcation (labeled as Hopf fold I, gray solid line) via a (two-dimensional) fold-Hopf bifurcation. Together with the lower fold bifurcation (blue line), this Hopf fold I delimits a bistable region with constant and oscillatory output (medium green section). At about $p_{fb} \approx 90 \text{ s}^{-1}$, another fold bifurcation arises (labeled Hopf fold II, red solid line) and establishes a limit cycle which introduces an oscillatory output with larger amplitudes. The constant/oscillatory bistability (medium green section) becomes an oscillatory bistability (dark green section) when the lower fold bifurcation (blue line) moves to p_{ff} values below those of the Hopf fold I bifurcation branch (gray line).

In summary, Figure 5.1A shows that extrinsic feedback input (i) regulates existence, output (constant or oscillatory), and robustness of the bistability around the working point and (ii) lowers the distance between working point and lower fold bifurcation (perception threshold), that is it regulates the basic operations.

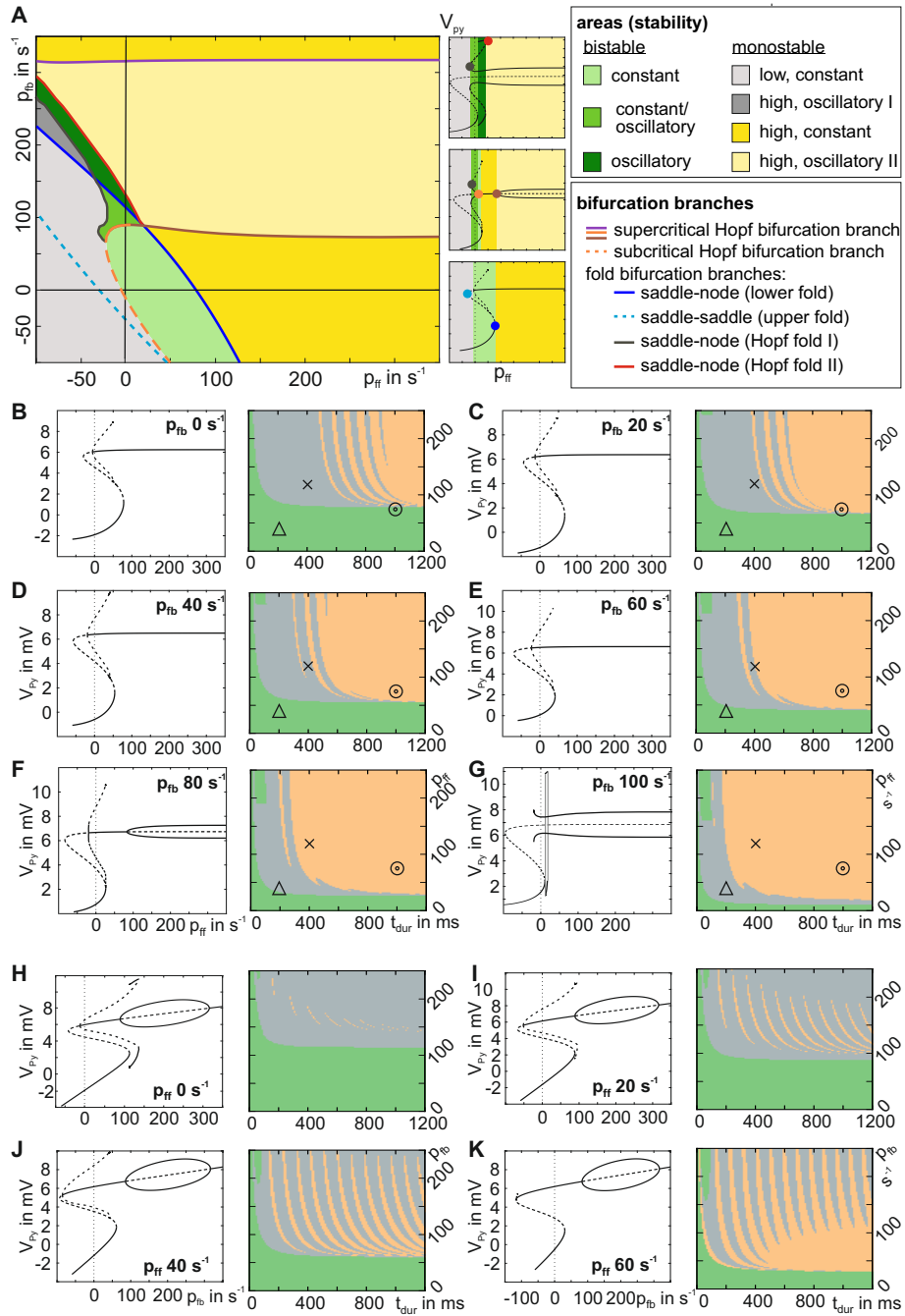


Figure 5.1: Simultaneous stimulation of the canonical microcircuit model with feedforward and feedback input. A) Two-parameter bifurcation diagram of the minimal canonical microcircuit with extrinsic feedback (p_{fb}) and feedforward (p_{ff}) input as bifurcation parameters. As explained in the main text, extrinsic feedback input (i) regulates existence, output (constant or oscillatory), and robustness of the bistability around the working point and (ii) lowered the perception threshold for feedforward input.

Figure 5.0 (previous page): B)-G) Left-hand: bifurcation diagrams exhibit stable (solid lines) and unstable (dashed lines) states of the mean membrane potential V_{Py} for a range of feedforward input values (p_{ff}) for increasing levels of constant feedback input (p_{fb}). Right-hand: characteristic fingerprints illustrate the consequent response behaviors for transient p_{ff} stimulation. Colors code the stimulation-induced response behaviors: non-responsive (green), transfer (gray), and memory behavior (orange). The markers (cross, triangle, and circle) denote distinct stimuli whose specific response behavior depends on the level of concomitant feedback input (p_{fb}). H)-K) Opposite case of increasing levels of constant feedforward input (p_{ff}) and transient feedback input (p_{fb}). Single-parameter bifurcation diagrams and characteristic fingerprints indicate an increase in memory behavior. However, these response behaviors are not robust to variations in the stimulation duration, as indicated by their striped appearance.

To demonstrate the modulating influence of extrinsic feedback information in the time course of the minimal canonical microcircuit I computed characteristic fingerprints (Fig. 5.1B-G). The model simultaneously received transient feedforward and constant feedback input. The time constant feedback levels reflect long holding times due to slower top-level processes [58]. Increasing the feedback input enlarged the range of memorized stimuli (orange area in Fig. 5.1B-G). Moreover, the range of nonresponsive behavior (green area in Fig. 5.1B-G) decreases – a consequence of the lowered perception threshold. Consequently, feedback input allows formerly unperceived stimuli to be perceived or even memorized (see markers in Fig. 5.1B-G). For comparison, I also considered the less plausible contrary case and applied increasing levels of constant feedforward input simultaneously to transient feedback stimulation. In this case, a stripe-like memory response behavior establishes that was very sensitive to variations in stimulation duration (see Fig. 5.1H-K).

In summary, feedback input makes the minimal canonical microcircuit more sensitive to feedforward input by stabilizing the memory behavior and lowering the perception threshold. Importantly, this modulatory effect causes different response behaviors for identical stimuli and designates the feedback input as a facilitative signal. This *facilitative feedback* can be used in two ways to modify the response behavior to feedforward stimulation: (i) a non-perceivable stimulus (compare triangles in Fig. 5.1B and G) becomes perceivable and (ii) an either non-perceivable (circle) or non-memorizable stimulus (cross) becomes memorizable (Fig. 5.1B and G).

5.3 Definition of Prototypical Meta-circuits

So far I showed that the minimal canonical circuit model reacts differently to feedforward and feedback information flows and that feedback input regulates access to the basic operations for feedforward input. In this methodological section, I derive prototypical meta-circuits of interacting minimal canonical microcircuits that use this processing to support priming and structure-building.

The initial meta-circuit (Fig. 5.1A) describes a minimal circuit of hierarchical information integration. A higher level minimal canonical microcircuit A_1 conveys facilitative feedback signals to a lower level minimal canonical microcircuit A_2 , which receives additional feedforward stimulations $p_{ff-Stim}$. The feedback signal is weak if A_1 resides in a low activity state (i.e., below firing threshold, see Section 4.2) and high if A_1 resides in a high activity state. This initial meta-circuit represents the minimal canonical architecture that establishes state-dependent processing in the sense that activity of one minimal canonical microcircuit (here A_1) governs the stimulation-induced response of another minimal canonical microcircuit (here A_2). This state-dependent processing operation represents a potential building block for versatile phenomena in perceptual and memory processing. In particular, it resembles predictive coding as A_2 integrates conceptual model information (feedback input) with novel sensory information (feedforward input).

From this initial meta-circuit, I derived a prototypical meta-circuit for priming by introducing a feedforward connection from the lower to the higher minimal canonical microcircuit (Fig. 5.1B). The inherent mechanism of this priming meta-circuit, namely the adaptive shift of the perception threshold, is extensively studied for its boundary conditions in Section 5.4.

I further derived a prototypical meta-circuit for structure-building (Fig. 5.1C). \tilde{A}_2 is conceptually treated as a higher level microcircuit that still receives a facilitative feedback signal $p_{fac,in}$ from \tilde{A}_1 . The feedforward stimulation signal $p_{ff-Stim}$ originates from another low-level canonical microcircuit \tilde{A}_3 which in turn receives feedforward stimulations at its EIN. The facilitative feedback signal $p_{fac,in}$ effectively regulates the availability of working memory in \tilde{A}_2 – as studied in Section 5.5. By providing a facilitative signal $p_{fac,out}$ to other connected circuits, this structure-building meta-circuit supports the establishment of spatiotemporal neural activity patterns and is used as a module for the syntax parsing model that is examined in Section 6.3.

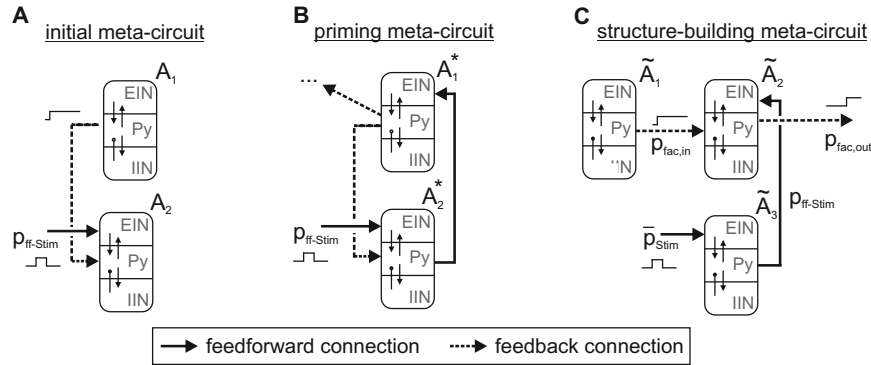


Figure 5.1: Prototypical meta-circuits of cooperating canonical microcircuit models. *A)* In the initial meta-circuit a lower minimal canonical microcircuit A_2 receives a facilitative feedback signal (dashed line) from a higher minimal canonical microcircuit A_1 . Thereby, A_1 modulates A_2 's response to a feedforward stimulation $p_{ff-Stim}$. Two computationally relevant derivations are investigated: *B)* Facilitation modifies perception (priming), where a recurrent feedforward coupling from A_2^* to A_1^* (solid line) allows A_2^* to activate A_1^* . The consequent facilitative feedback signal lowers the perception threshold in A_2^* . Note that the facilitative feedback signal, conveyed by A_1^* , may also modify the perception threshold of yet another circuit (as indicated by the three dots). *C)* Facilitation modifies memorization (structure-building), where \tilde{A}_2 is treated as a higher microcircuit but still receives a facilitative feedback signal, $p_{fac,in}$, from \tilde{A}_1 . The feedback signal conditions the memorization of a feedforward stimulation $p_{ff-Stim}$ that originates from a lower microcircuit \tilde{A}_3 . In case of an activation of \tilde{A}_2 , a facilitative feedback signal $p_{fac,out}$ is conveyed to connected circuits, effectively cascading this local operation and supporting the incremental build-up of sustained activity patterns.

5.4 State-dependent Priming

5.4.1 Principle Priming Mechanism

Priming is an ubiquitous aspect of the brain's processing abilities for which one stimulus influences the processing of subsequent stimulations [19]. In the following, I examine a neural mechanism for priming that is based on the local cooperation of two minimal canonical microcircuits A_1^* and A_2^* (see Fig. 5.1B) for its boundary conditions. At first (Fig. 5.2A), an afferent target stimulus does not considerably affect the output of a minimal canonical microcircuit A_2^* (see Figs. 5.2D and E). However, a priming stimulus with higher intensity causes a transient output in A_2^* that activates A_1^* (Fig. 5.2B). By means of a feedback connection the sustained high activity of A_1^* modulates the sensory sensitivity in A_2^* and effectively shifts the perception threshold (Fig. 5.2C). Consequently, the same target

stimuli (that had no effect before priming) evokes an output in A_2^* that is available for further processing. A deactivation of the higher level microcircuit A_1^* makes A_2^* insensitive to the target stimuli again. This homeostasis mechanism is readily implementable into the present model.

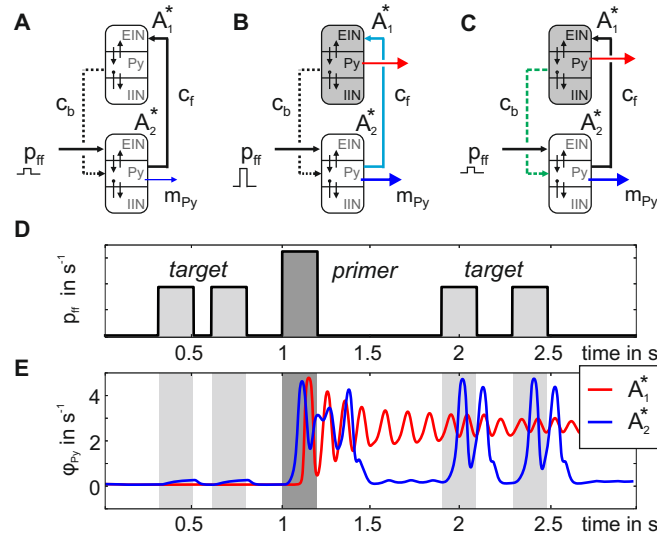


Figure 5.2: Principle of the priming mechanism. A-C) A target stimulus, applied to a minimal canonical microcircuit A_2^* , does not evoke a significant output in A_2^* . However, as soon as a priming stimulus activates the higher minimal canonical microcircuit A_1^* (light blue) A_1^* emits a facilitative feedback signal (green dashed line) that allows A_2^* perceiving the target stimulus. D) Time course of afferent feedforward stimulations of the minimal canonical microcircuit A_2^* . E) Efferent signals of the minimal canonical microcircuits.

5.4.2 Priming Mechanism Analysis

Compared to a single minimal canonical microcircuit, the priming meta-circuit conceptually separates the basic operations of signal flow gating and working memory. The lower level circuit A_2^* primarily perceives and transmits salient inputs for further processing. Meanwhile, the higher level circuit A_1^* memorizes the primer and modulates the processing in A_2^* via facilitation. In particular, the facilitative feedback signal emitted by A_1^* does not need to act on the same circuit that received the primer, that is self-priming. Instead, it may also facilitate processing in a different target-receiving microcircuit, representing the perception of another modality (as indicated in Fig. 5.1B).

I examined the boundary conditions of this priming mechanism by closer inspection of a set of key parameters:

- **The inter-circuit connectivity gains c_f and c_b :** The strength of feedforward and feedback connections determine the signal strength at the targeted microcircuits. Feedforward connections scale the feedforward input that A_1^* needs to switch from a low to a high activity state. Feedback connections scale the facilitative feedback signal that modulates the response behavior in the lower circuit A_2^* .
- **The afferent stimuli:** Intensity and duration of the target stimulus determine whether a stimulus can be perceived or not. The priming stimulus, being salient in terms of its intensity (or duration), evokes a sustained high activity (i.e., memory behavior) in A_1^* , causing the facilitative feedback signal. I examined the relationship between target and priming stimuli in terms of a varying intensity while keeping their durations equal.
- **Individual adaptation of the microcircuits:** I already showed that individual levels of the local network balance bias the response dynamics of canonical microcircuit models (see Section 4.3). In particular, a slight increase of inhibition, relative to the default parameterization, favors the transfer behavior, whereas a slight decrease of inhibition favors the memory behavior. Accordingly, I examined how inhibitory synaptic gains H_i , characterizing each population's inhibitory synaptic response, constrain the working range of the priming mechanism.

To evaluate these critical parameters in the priming mechanism, I applied individual stimulation streams to the lower level circuit A_2^* of the priming meta-circuit. Each of those streams was composed of two target stimuli that were separated by a priming stimulus. Figure 5.3 describes the analysis procedure. The response behaviors of A_1^* and A_2^* were evaluated in seven analysis windows before, during, and after the individual stimuli. The membrane potential of the pyramidal cells was compared with a firing threshold of 4 mV. A stimulation stream was defined as *effective* (i.e., evoking a shift of the perception threshold) if (i) the priming stimulus evokes a sustained high activity in A_1^* , but not in A_2^* , and (ii) the target stimulus evokes a supra-threshold transient deflection in A_2^* only after the priming stimulus. The target stimuli varied in intensity and duration, whereas the priming stimulus was of equal length yet of higher intensity in relation to the target stimuli (see Table

5.1). Supraliminal target stimuli that evoke supra-threshold deflections in A_2^* even without priming were disregarded. For all other stimulation streams, I mapped the percentage of effectual stimulation streams to the inter-circuit connectivity gains c_f and c_b , see Figure 5.4. This stimulation procedure was repeated with modified inhibitory synaptic gains H_i in A_1^* and A_2^* , signifying an altered network balance compared to the default network balance [15]. Table 5.1 lists the varied parameter values.

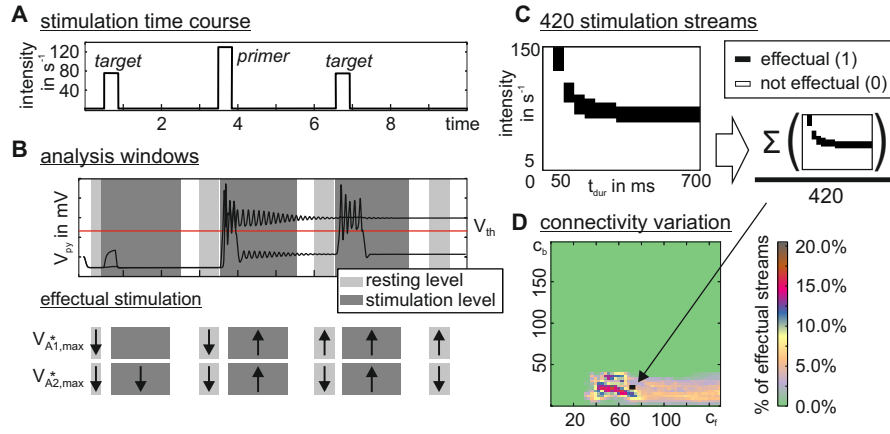


Figure 5.3: Analysis of the priming mechanism. A) For the characterization of the priming mechanism, I applied stimulation streams that comprised two target stimuli separated by a priming stimulus. B) I evaluated the response behaviors of A_1^* and A_2^* by comparing V_{py} to a firing threshold of 4 mV in seven analysis windows: before, after (i.e., resting level, light gray sections), and during stimulation (dark gray sections). As indicated by the arrows, a stimulation stream is determined as effectual if (i) the priming evokes a sustained high activity in the higher canonical microcircuit A_1^* , but not in A_2^* , and (ii) the target stimulus evokes a transient high activity only after priming. C) Target stimuli of different duration (t_{dur}) and intensity are applied. The sum of all effectual stimulation streams is scaled by the total number of considered stimulation streams (420, see Table 5.1). D) The percentage of effectual streams is mapped to the range of connectivity gains c_f and c_b in order to characterize their constraining influence on the priming mechanism.

5.4.3 Analysis Results

The connectivity gains c_f and c_b constrain the priming mechanism (see Fig. 5.4). Effectual stimulation streams occur for $c_f > 30$. This value denotes a minimum feedforward strength in order to convey A_2^* 's activity to the higher level circuit during priming. Moreover, effectual stimulation streams occur for $c_b < 45$. This value denotes a maximum feedback strength in order to limit the facilitation in the

Table 5.1: Parameter values for the assessment of the priming mechanism.

		Target stim.	Priming stim.	
Stimulus duration	minimum	50 ms	50ms	
	step	50 ms	50ms	
	maximum	700 ms	700ms	
Stimulus intensity	minimum	5 s^{-1}	Target + (20, 40, 80)%	
	step	5 s^{-1}	5 s^{-1}	
	maximum	150 s^{-1}	Target + (20, 40, 80)%	
Connectivity gains	minimum	c_f		c_b
	step	0		0
	maximum	3		3
		150		200
Synaptic gains	Default (Fig. 5.4A)	A_1^* bias (Fig. 5.4B)	A_2^* bias (Fig. 5.4C)	A_1^* & A_2^* bias (Fig. 5.4D)
$H_i(A_1^*)$	22 mV	21 mV	22 mV	21 mV
$H_i(A_2^*)$	22 mV	22 mV	23 mV	23 mV

lower level circuit and to keep it sensitive to further stimulation (i.e., prevent self-activation). However, note that the exact values are subject to the chosen model parameterization. These connectivity constraints are observed for a varying intensity of the priming stimulus (rows in Fig. 5.4) and for all considered configurations of the inhibitory synaptic gains H_i (Fig. 5.4A-D). Further, the percentage of the effectual stimulation streams increases with the relative strength of priming and target stimulus for all considered variations of the inhibitory synaptic gains H_i (Fig. 5.4A-D). For the default configuration (Fig. 5.4A), high rates of effectual stimulation streams are restricted to a small range of connectivity values. Both this range and the rate itself increases for a slight decrease of inhibition in the network balance of the higher circuit A_1^* (Fig. 5.4B) and for a slight increase of inhibition in the network balance of the lower circuit A_2^* (Fig. 5.4C). A combination of both (Fig. 5.4D) maximizes the range of suited connectivity values and the rate of effectual stimulation streams.

I identified the priming stimuli that are actually able to shift the perception threshold (Fig. 5.5). Strong and long stimuli (i.e., supraliminal) are already recognized, making priming superfluous. Furthermore, weak and brief stimuli are not suited to initiate the priming as they fail to evoke a memory behavior in the higher-level circuit A_1^* . Tuning the inhibitory synaptic gains increases the range of suited stimu-

lation parameters. Importantly, the described priming mechanism is not supported by alterantive topologies of the priming meta-circuit (see Fig. 5.6).

In summary, the proposed priming mechanism performs an adaptive processing of inputs in dependence on previous events that cause top-down facilitation. The mechanism rests on the constructive cooperation of the involved minimal canonical microcircuits and suggests their conceptual specialization to either *signal gating*, thereby avoiding insensitivity to future stimuli, or *memorization* that guides further information processing through facilitation. This functional specialization can be biased by means of the local relationship of excitation and inhibition, i.e. the local network balance, in the involved microcircuits. The topology of feedforward and feedback connections within the meta-circuit and the salience of stimuli are effective constraining factors of the priming mechanism.

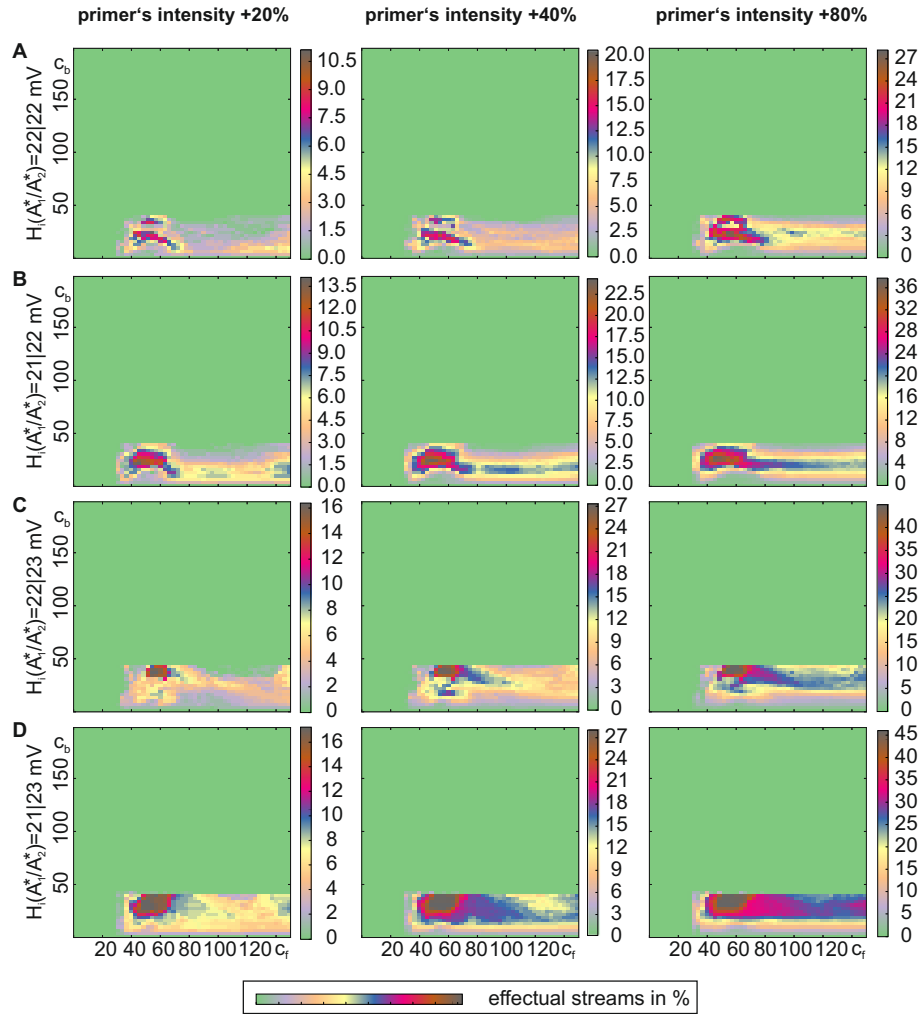


Figure 5.4: Impact of stimulus salience, connectivity gains, and network balance to the priming mechanism. The single plots color-code the percentage of effectual stimulation streams that evoke priming for varying values of the connectivity gains. Stimulations varied in duration and intensity. Priming and target stimulus were equally long, but the primer's intensity was 20% (left column), 40% (middle column) or 80% (right column) larger. Connectivity gains constrain the effectual stimulation streams. Stronger priming leads to more effectual stimulation streams both in terms of the maximum rate and the extent of suited connectivity gains. Compared to the default configuration of inhibitory synaptic gains (A), a slight decrease of inhibition in A_1^* (B), and a slight increase of inhibition in A_2^* (C) enhance the maximum rate of effectual stimulation streams and the range of suited connectivity gains. The combination of both inhibition variations maximizes the percentage of effectual stimulation streams (D). Note the different color scaling in the single subplots and that the percentages are specific for the chosen stimulation parameter ranges (see Table 5.1).

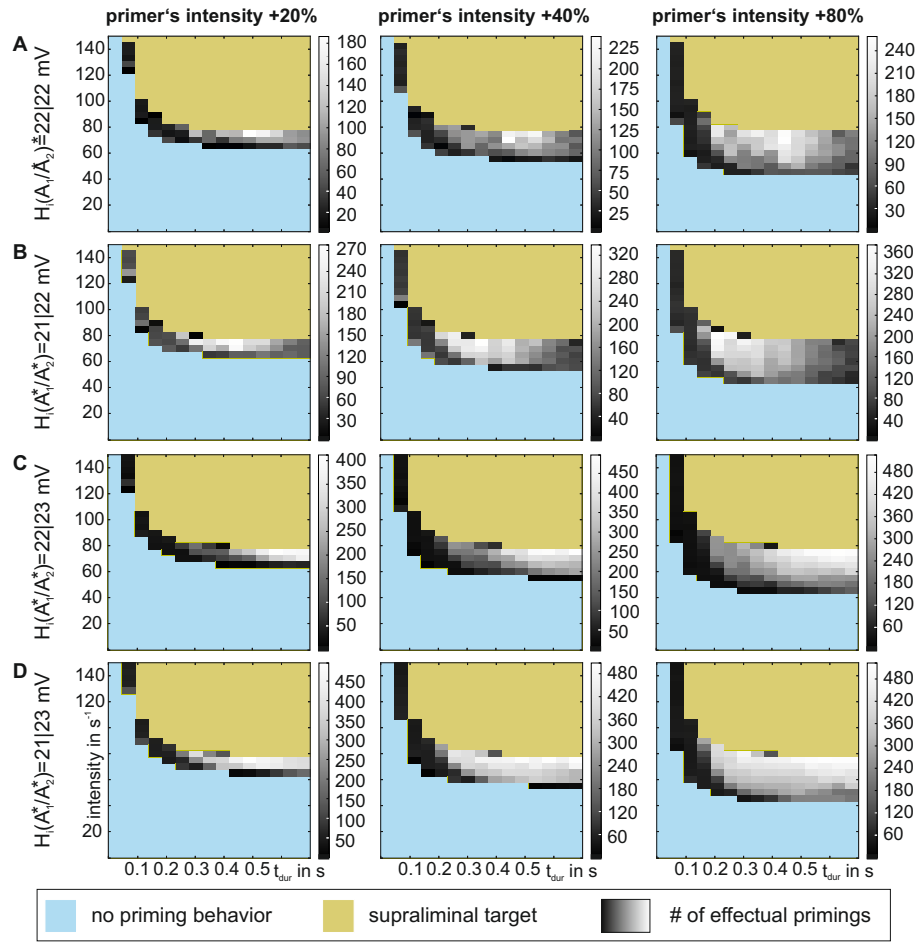


Figure 5.5: Characterization of stimuli that shift the perception threshold. Each subplot shows the number of successful threshold shifts for a specific stimulus in grayscales. Numbers sum up the respective $c_f - c_b$ combinations. Salient, supraliminal target stimuli, i.e. very strong or long stimuli, are already perceived, making priming superfluous (sandy coloring). Furthermore, weak and brief (i.e., subliminal) stimuli fail to evoke a response memory behavior in the higher circuit A_1^* (light blue coloring) and do not initiate the priming. Columns relate to the stimulation intensity, where the priming stimulus was 20% (left column), 40% (middle column) or 80% (right column) larger than the target stimulus. In general, a stronger relative priming intensity increases the range of perceived target stimuli. Compared to the default configuration (A), promoting the memory response behavior in the higher circuit A_1^* through a decreased inhibitory synaptic gain (B) increases the maximum number of effectual primings. C) An even larger increase occurs for a promotion of the transfer response behavior in the lower circuit A_2^* through an increased inhibitory synaptic gain. D) A combination of both synaptic gain modulations maximizes not only the maximum number of perceived target stimuli, but also their robustness to variations in intensity or duration, marked by the larger range of effectual stimulations. Note the different grayscales for each subplot.

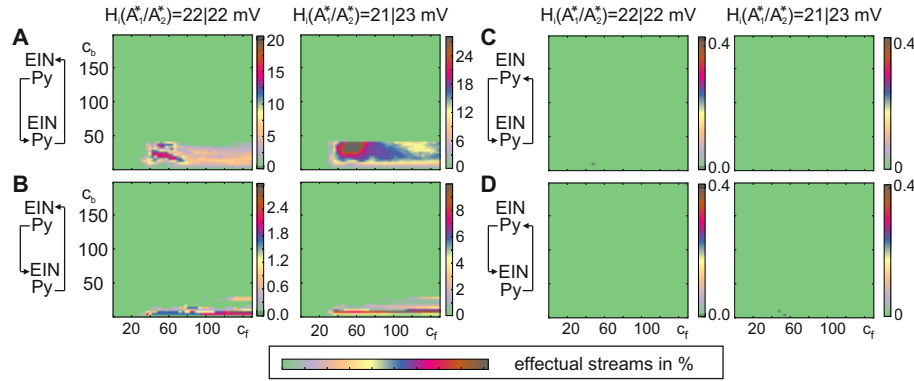


Figure 5.6: Influence of the priming meta-circuits' topology. A) For the priming mechanism, I consider a specific topology within the priming meta-circuit: the higher microcircuit connects to the lower microcircuit via a feedback connection, i.e. targeting the Py, and the lower microcircuit connects to the higher microcircuit via a feedforward connection, i.e. targeting the EIN. This arrangement supports the idea of effective functional specialization: the feedforward connection is prone to storing events and favoring memory in the higher circuit. The feedback connection robustly modulates the response behavior and qualifies the lower microcircuit for dynamic signal flow gating. In contrast, alternative topologies – solely feedforward connections (B), solely feedback connections (C) or a permutation of feedforward and feedback (D) – fail to support the priming mechanism, but may be relevant for other cooperative neural operations. For the simulations, the primer's intensity was 40% higher than that of the target stimulus.

5.5 State-dependent Structure-building

The previous section showed how top-down facilitation via feedback connections permits the dynamic perception of subliminal stimuli. In the following, I show how a facilitative feedback signal conditions the memorization of stimulation events in the structure-building meta-circuit (Fig. 5.1C).

In the meta-circuit for structure-building, I applied transient rectangular feedforward stimuli to \tilde{A}_3 and mapped the response behaviors of \tilde{A}_2 and \tilde{A}_3 (Fig. 5.7). An effectual stimulus evokes a transient activity in \tilde{A}_3 and a persistent activity in \tilde{A}_2 . Hence, \tilde{A}_3 perceives the stimulus, but remains sensitive for further stimulation, whereas \tilde{A}_2 memorizes the stimulation. Figure 5.7A shows the response behaviors for increasing levels of facilitative feedback $p_{fac,in}$. Weak and short stimuli are not able to evoke persistent activity in \tilde{A}_2 (gray area) whereas strong and long stimuli activate both \tilde{A}_2 and \tilde{A}_3 (black area). Few stimuli selectively activate \tilde{A}_2 without sustained activation of \tilde{A}_3 (green area). However, this selective activation becomes

more likely for higher levels of the facilitative feedback signal $p_{fac,in}$ (compare the red crosses in Fig. 5.7A). The selective activation can be further promoted by favoring a transfer response behavior in \tilde{A}_3 through a slight increase of inhibition (Fig. 5.7B). Hence, the top-down facilitative feedback signal $p_{fac,in}$ conditions the establishment of sustained activity in one part of the structure-building meta-circuit (\tilde{A}_2), while preserving its sensitivity in \tilde{A}_3 . The importance of these characteristics will become evident for structure-building in larger networks, as demonstrated in Chapter 6.

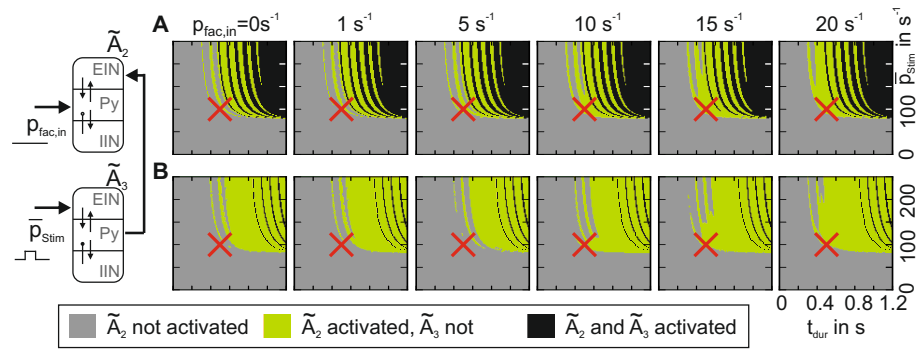


Figure 5.7: Facilitative feedback signals enable the memorization of stimulation events. A) A minimal canonical microcircuit \tilde{A}_3 receives transient feedforward stimulation. The consequent feedforward stimulation caused a sustained activation of \tilde{A}_2 while keeping \tilde{A}_3 sensitive to further stimulation (green area) if the facilitative feedback $p_{fac,in}$ is sufficiently strong. Weak and brief stimuli, applied to \tilde{A}_3 , fail to activate \tilde{A}_2 (gray area), whereas strong and long stimuli activate both \tilde{A}_2 and \tilde{A}_3 (black area). B) The selective activation of \tilde{A}_2 can be promoted by a slight increase of the inhibitory synaptic gain H_i , which favors the transfer response behavior in \tilde{A}_2 . Red crosses denote an exemplary stimulation of defined intensity and duration applied to \tilde{A}_3 .

5.6 Discussion

The performed analysis shows that extrinsic feedback stabilizes the working memory behavior that is induced by feedforward stimulation and lowers the perception threshold of a minimal canonical microcircuit, thereby increasing its sensitivity. Hence, identical stimuli evoke different response behaviors depending on the activity state of the minimal microcircuit that emits the facilitative feedback signal. This also means that extrinsic feedback input constrains the access to the basic operations. In two exemplary meta-circuit configurations, this *state-dependent* facilitative feedback is shown to influence information processing of neural signals, that is to say their perception (priming) and memorization (structure-building). Facilitative feedback signals flexibly⁸ influence the local relationship of excitation and inhibition in the minimal canonical microcircuit. Moreover, variations of the less mobile inhibitory synaptic gains also affect the function of minimal canonical microcircuits in these meta-circuits. These results are now discussed in more detail.

Facilitative feedback conditions the processing of afferent feedforward stimulations. This observation is in line with empirical studies showing the modulatory effect of top-down information that includes contextual information to the processing of sensory information [115]. The analysis shows that both feedforward and feedback input can modulate the respective other, though in an asymmetric way. This supports the notion that feedforward information is not strictly driving and feedback information is not strictly modulatory [58, 116]. However, feedback modulation appears to be more robust to variations of afferent stimuli than feedforward modulation (Fig. 5.1). This model prediction needs to be evaluated in experimental studies.

The facilitative feedback signal is thus a flexible, effective, and metabolically inexpensive way to regulate information processing without the need for synaptic adaptations. Besides the ability to select signals in a bottom-up way based on signal properties (signal flow gating, see Section 4.2), extrinsic feedback provides the canonical microcircuit models with the ability to select signals based on the state of the surrounding network. Hence, in addition to cascaded minimal canonical microcircuits that establish conditioned associations by superimposed feedforward signals (feedforward conditioning, see Section 4.6) state-dependent signal selec-

⁸ In the sense of quickly adjustable in time and space.

tion reflects the controlling influence of higher processing levels via feedback connections, here termed *hierarchical conditioning*. Conditioned associations that rest on state-dependent processing operations further corroborate the compatibility of minimal canonical microcircuits and cell assembly theory [4, 41]. Notably, I only considered stationary feedback signals. Future work should also investigate the simultaneous application of transient feedforward and feedback signals.

In interacting minimal canonical microcircuits the basic operation of working memory allows one minimal canonical microcircuit to store processing events over time and to keep this information available to modulate future processing steps of another minimal canonical microcircuit. The resulting temporal processing history gives rise to a broad set of conceivable adaptive processing mechanisms, two of which I further exemplified. The initial prototypical meta-circuit (Fig. 5.1A) is the minimal model for this form of state-dependent processing as it conceptually separates the basic operations of working memory and signal flow gating and assigns them to the two involved minimal canonical microcircuits, respectively. In other words, in the initial prototypical meta-circuit the single minimal canonical microcircuits functionally specialize in a distinct basic operation. Notably, this specialization depends on the dominance of either feedforward or feedback information (Fig. 5.6). This finding supports the notion that the development of neural microcircuits on mesoscopic levels, such as through pruning of feedforward or feedback connections, establishes a commitment of specialized areas to executable functions [46].

Given that the inhibitory synaptic gains bias the response behavior (Section 4.3), I showed how the adaptation of inhibitory synaptic gains favors the functional specialization of minimal canonical microcircuits. For instance, this adaptation improves the reliability and efficiency of the facilitative feedback signal in priming (Fig. 5.4). The functional specialization on basis of a shifted network balance does not require any structural changes. Hence, the possibility to dynamically reassign the dominant basic operations of minimal canonical microcircuits underscores the task-based adaptability of networks of canonical microcircuit models and eventually the cortical matter⁹. This might be an important aspect if a part of the cortex needs to adapt to a new functionality, such as during remapping after stroke [117] or acquisition of new skills. The present canonical microcircuit model suggests that the functional specialization is reversible at different time scales and through

⁹ Given that both feedforward and feedback channels are present and not pruned away.

different modalities that affect the local relationship of excitation and inhibition, or local network balance. Although I chose inhibitory synaptic gains, lumping together parameters of synaptic transmission [15], other conceivable neurobiological plausible parameters to regulate the local network balance include afferent information flows, altered synaptic weights, and electrical brain stimulation [92]. The neural causes that give rise to the behavioral effects of electrical brain stimulation are still unclear [118]. Hence, the present findings on functional consequences of a shifted local network balance are an interesting starting point for the investigation of functional consequences associated with electrical brain stimulation. In terms of modulating information flows, the hierarchical interaction between canonical microcircuit models can explain experimental findings about stimulus-specificity of sensory neurons [21, 119]. In a memory paradigm, macaque monkeys memorized the shape or color of stimuli. It was found that neurons feature stimulus-specific response patterns (in the sense that they respond less or do not respond when a stimulus of another color is shown) and that this specificity depends on another cortical area. In the model of canonical microcircuits, specificity was demonstrated in the characteristic fingerprint and afferent activity of a connected area was shown to modulate this particular response.

I proposed prototypical meta-circuits for two neural operations that make use of adaptive, state-dependent processing, namely priming and structure-building. Common to both phenomena is the notion that past processing steps modify the current processing. In particular, the facilitative feedback, arising from the sustained high activity of one canonical microcircuit model, either shifts the perception threshold (perceptual priming) or permits a stimulation-induced sustained high activity (structure-building) in another microcircuit. In both examples, the facilitative feedback signals confine potential subsequent processing steps. That means non-facilitated, or unpredicted, stimulations are ineffective. In other words, state-dependent processing typifies *anticipation*. Moreover, minimal canonical microcircuits integrate sensory or prior information and conceptual or model information via hierarchical connections. These characteristics suggest a conceptual proximity of state-dependent processing in minimal canonical microcircuits to predictive coding [60, 112–114]. However, in contrast to other established canonical microcircuit models that embody predictive coding [58, 86] the present framework of minimal canonical microcircuits does not consider the explicit exchange of prediction errors, potentially reflecting different mechanisms [41].

Priming refers to a behavioral phenomenon generally stating that past experience modifies the current processing performance [19, 110, 111]. Because of its ubiquity and despite its behavioral diversity, priming may depend on a small set of generic neural mechanisms that are individualized for the respective task. Based on the dependence on past states, priming must involve some kind of storage and was proposed as a special form of memory [111]. I propose a mechanism for perceptual priming that rests on the dynamic shift of the perception threshold by means of a facilitative feedback signal. Analysis of topology, connectivity weights, and stimulation characteristics suggests that the meta-circuit's architecture must support a functional specialization to working memory and signal flow gating. This is accomplished by a feedforward connection that evokes a memorized high activity and a feedback connection that mediates facilitative signals. Notably, I only considered self-priming for which a single minimal canonical microcircuit receives both target and priming stimuli. However, a conceivable separation of these receiver circuits will support cross-priming. Interestingly, similar to the present priming scenario, Ardid and colleagues [87] proposed a model for attentional processing that investigates the hierarchical cooperation of the prefrontal cortex and the visual middle-temporal area. The similarity of the computational models supports a supposed common mechanism underlying top-down attention processing and perceptual priming [19].

The analysis demonstrated that extrinsic feedback information conditions the memorization of feedforward stimulations in structure-building operations. This *hierarchical conditioning* superimposes driving feedforward and modulating feedback information. Hence, hierarchical conditioning complements feedforward conditioning (Section 4.6) to manifest conditioned associations in minimal canonical microcircuits. Cascading this local operation (i.e., a *synfire graph*, see Section 2.1) in a network of hierarchically interacting minimal canonical microcircuits – effectively representing an operational cell assembly – promises to be an effective way to establish (sparse) spatiotemporal activity patterns that reflect cognitive processing. In the next chapter, I exemplify this notion and show how the sequential and selective activation of canonical microcircuit models within a distributed network realizes incremental (i.e., word by word) syntax parsing during the perception of a sentence.

Chapter 6

Syntax Parsing in Networks of Canonical Microcircuit Models

“Language processing is a paradigm case of higher cognitive processes and the basic neuronal principles and mechanisms that are fruitful here might apply, on a broader scale, to other higher cognitive processes as well.”

– Wennekers et al., 2006

6.1 Introduction to Syntax Parsing as Part of Sentence Processing

6.1.1 The Cortical Language Circuit

In this chapter, I apply the gathered findings on minimal canonical microcircuits as represented by the three-population model to a biologically realistic network model that supports syntax parsing during sentence processing. In the following, I briefly review the cortical language circuit that is assumed to support cognitive language processing and present neurocomputational approaches that investigate representation and mechanisms of cognitive language processing at the neuronal level. In two exemplary syntax parsing models, I then prove the suitability and merit of the framework of hierarchically interacting canonical microcircuit models to account for this challenging neurocognitive task.

The comprehension of a sentence involves the decoding of semantic, syntactic, and prosodic information from the perceived stream of words [89, 120, 121]. Whereas prosodic information is supposed to be initially processed dominantly in the right hemisphere, processing semantic and syntactic information rests on a fronto-temporal network in the left hemisphere [121, 122]. This cortical language circuit includes temporal areas, namely the primary auditory cortex (PAC), the superior temporal gyrus (STG), and the middle temporal gyrus (MTG) as well as frontal areas, namely the frontal operculum (FOP), Brodmann areas (BA) 47 (pars orbitalis), BA 45 (pars triangularis), BA 44 (pars opercularis), and the premotor cortex (PMC), see Figure 6.1. Ventral and dorsal pathways support an information transfer between these cortices.

The information flow within this cortical language circuit is described by the *dynamic dual pathway model* [121]. In the following, I focus on the processing of syntactic information and disregard semantic information (but see [121]). Based on perceived auditory information, an acoustic-phonological analysis in the PAC yields phonological word forms whose syntactic word categories (e.g., noun, verb, adjective etc.) are used to construct syntactic phrases (i.e., structured sequences of words) in the anterior superior temporal cortex (aSTG). In the adult brain, available templates of different phrase structures are proposed to speed up the phrase structure building processes. For higher order syntactic computations, syntactic information is transferred from the aSTG to the FOP via ventral connections and from there to the posterior portion of the inferior frontal cortex (IFG, in particular BA44).

Notably, there is a subtle difference between the neurolinguistic concept of phrase structure building (i.e., establishing a phrase structure) and the neurocomputational concept of structure-building as used here (i.e., establishing a neural activity pattern). One aim of this chapter is to demonstrate their accordance in the way that the establishment of phrase structures relies on neurocomputational structure-building computations. This process of syntactic structure-building, that is decoding and temporal storage of syntactic information based on a continuous stream of afferent phonological word information, is referred to as *syntax parsing* [123, 124].

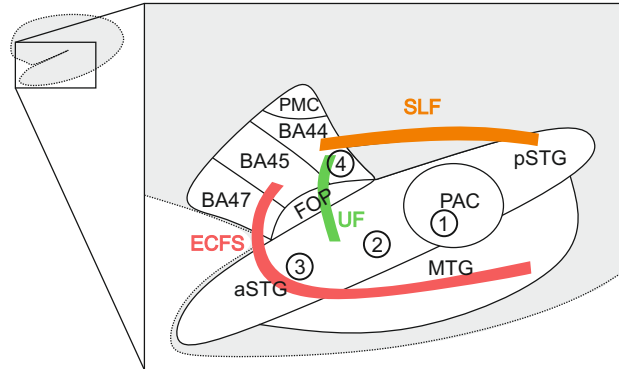


Figure 6.1: Schematic of the cortical language circuit. (1) Subsequent to the perception of auditory information in the primary auditory cortex (PAC), (2) phonological word form and syntactic word category are determined. These are the basis for (3) the initial construction of syntactic phrases in the anterior superior temporal cortex (aSTG). The middle temporal gyrus (MTG) is responsible for the lexical-semantic access of words. Via ventral pathways, namely the uncinate fasciculus (UF) and the extreme capsule fiber system (ECFS), semantic and syntactic information is transmitted from temporal to frontal areas (i.e., BA 44/45/47 and frontal operculum, FOP) for higher-order computations. These computations, for instance the syntactic and semantic integration involving the posterior superior temporal gyrus (pSTG) via dorsal pathways (superior longitudinal fasciculus, SLF), are an active field of research [122].

6.1.2 Neuronal Representation and Mechanisms of Language Processing

Establishing not just syntactic hierarchies but also the mere reproduction of a sentence after some time, requires a local representation of the parts of a sentence (words, phrases) at the neuronal level in order to store them and detect their mutual relations. Neurocomputational research conjectures about this neuronal representation and putative mechanisms for language processing.

Rolls and Deco [22] propose that words are represented as distinct states of cortical attractor networks. According to their notion, these attractor networks, or *modules*, represent the syntactic role of a word (subject, verb, and object) using place coding. Each module is thought to feature a sufficient capacity to encode a person's working vocabulary of the respective word class. Syntactic relations are encoded by the temporal order of activation of these modules (i.e., subject-verb-object, implemented through weak forward coupling). Inflections of words or prepositions may

be possible alternatives. Every word a module actually represents is determined by a deep network layer that reflects meaning.

Pulvermüller and colleagues [4, 21, 32, 125] propose another theory on language mechanisms that considers cell assemblies. Pulvermüller assumes that words are dividable into categories depending on their syntactic and semantic role in the sentence. Whereas some words refer to an actual semantic concept (content words, such as nouns, verbs or adjectives) others modify these concepts or their relation (function words, such as conjunctions and determiners) [125]. In regards of content words the actual *word* merely reflects the phonological feature of the semantic concept of an object (e.g., cat) that further involves other sensory perceptions, such as shape or smell. Each of these features is represented by individual neural units that are interlinked to coherent representations, called *functional webs*, with a distinct cortical topography. Notably, despite the conceptual equality with cell assemblies, Pulvermüller introduced the term *functional webs* to prevent misunderstandings. Three qualitatively different types of functional webs are proposed to be relevant for realizing spoken language in the cortex: (i) phonological webs, (ii) word webs, and (iii) webs governing the serial processing of parts of a sentence. Phonological webs are distributed over the perisylvian cortex and link information about acoustic percepts and articulatory movements. Phonological webs are part of word webs, which associate phonological information and information about the actions and perceptions (semantics) to which a word refers. Accordingly, word webs have diverse topographic distributions, including visual and motor areas. The third kind of web reflects the concept of *operational* cell assemblies [20]. It realizes mechanisms for the serial processing of language, such as through synfire chains (or delay lines if axonal delays are considered), sequence detectors, and declining activity dynamics.

Common to both proposals is the idea that the incremental perception of a sentence is represented by a sequential functional activation of word-representing neural assemblies. Regarding syntax parsing, the conceptual separation of (phonological) word information and syntactic roles of words as well as signal propagation mechanisms within the neural substrate are of special interest. In the notion of Rolls and Deco phonological word information is represented within the module that reflects syntactic roles [22]. This means that words with multiple conceivable syntactic roles (e.g., drink) have multiple representations. Also, if the same word occurs multiple times in a sentence this word will need to recruit multiple modules, im-

plying a redundant representation of phonological word information. In contrast, word webs consider unique phonological representations. Mechanisms for syntactic structure-building, the binding of these representations according to syntactic roles, are still subject of current research [4].

In the following, I will demonstrate how the concept of functional webs, represented as network of hierarchically interacting minimal canonical microcircuits, may decode the syntactic hierarchy of a perceived sentence. I will first implement a syntax parsing model that reflects the approach of Rolls and Deco and point out the respective limitations. In that model, word-representing minimal canonical microcircuits are grouped to modules that reflect the syntactic roles of these groups via a place code. Signal propagation is assumed to rest on simple *feedforward* conditioning (see Section 4.6). Subsequently, the syntax parsing model is further developed to overcome its limitations by using *hierarchical* conditioning, as examined in Chapter 5, and separate minimal canonical microcircuits to reflect afferent word information and syntactic roles.

6.2 Syntax Parsing Based on Feedforward Conditioning

6.2.1 Model Architecture and Mechanism

For both syntax parsing networks, I adopt the notion that words are represented by distributed neural units, here minimal canonical microcircuits, with distinct cortical topographies [21, 22]. Moreover, the incremental parsing of syntactic information is represented by a sequential and selective activation of these canonical microcircuits. The sustained and spatio-temporally distributed activity pattern represents the syntax prototype. To illustrate the parsing process I chose the sentence *I hit the thief with the club*, which is ambiguous in its syntax. The phrase *with the club* can be interpreted as adverbial phrase, further specifying *hit*, or as adjective phrase, further specifying *the thief*. This ambiguity is assumed to be solved by available contextual information (disambiguating input [126]). This contextual information includes prosodic information [127] and semantic information such as specific knowledge concerning the discourse, the topic or individual experience. The temporal order of the words provides information about the assignment of subject and object [22]. The network topologies are fixed in terms of synaptic changes and assumed to be established during language acquisition. Further, I presume a

successful acoustic-phonological analysis previous to the syntactic parsing. This analysis step is not explicitly modeled here, but see Yildiz [38] for a conceivable approach. Information about the identified phonological word forms (*afferent word information*) is represented by rectangular stimulation signals of defined intensity and duration that are delivered to the syntax parsing network via feedforward connections.

In the first syntax parsing model (Fig. 6.2), words are represented by minimal canonical microcircuits (the three-population model, Fig. 3.3C) that mutually inhibit each other. The word-representing microcircuits are categorized into modules according to their syntactical role, i.e. subjects, verbs, objects, and their modifiers. Notably, nouns are redundantly represented in both the subject and the object module. The initial expectation to perceive a sentence is modeled as an initial driving signal received by all canonical microcircuit models in the subject module. The proposed structure-building computation is based on feedforward conditioning, an input-driven sequential activation of the minimal canonical microcircuits. An activated microcircuit, belonging to a certain word module, transmits its increased firing rate to those modules that are likely to follow. This signal transmission differentially pre-activates the respective word-representing minimal canonical microcircuits by means of weighted feedforward connections and, thus, creating expectations. In the model, this graded pre-activation corresponds to a shift in the baseline activity of a minimal canonical microcircuit. This brings the system closer to the respective fold bifurcation (see Section 4.2). Therefore, activation of the word *eat* in the verb module pre-activates words in the module of verb-modifiers and in the module of objects, but does not activate words in the module of verbs again (see Fig. 6.2). Subsequent afferent word information then fully activates the respective microcircuit and continues the structure-building process. The verb and object modifying modules allow for competing interpretations of a sentence. Their mutual inhibition in combination with the present level of contextual information ensures that one particular interpretation is supported at a time.

For the parsing of the exemplary sentence *I hit the thief with the club*, I focus on the relevant word-representing microcircuits and disregard connections to uninvolved microcircuits. Hence, the simulated network comprises six interacting minimal canonical microcircuits and their respective connections (Figs. 6.3A and B). Ambiguity, i.e. whether the phrase *with the club* serves as adjective or adverbial phrase, is addressed by separating verb-modifying from object-modifying modules. Both

modules mutually inhibit each other through asymmetrical connection strengths. Further, the present level of contextual information guides the structure-building process. In the simulations, contextual information is modeled as a noisy signal with a constant offset. This inhibitory influence degrades the propagated input to the respective module. Afferent word information is modeled as rectangular stimuli of defined intensity $p_{Stim} = 50 \text{ s}^{-1}$ and duration $t_{dur} = 500 \text{ ms}$ (initial drive: $p_{Stim} = 95 \text{ s}^{-1}$ and $t_{dur} = 840 \text{ ms}$).

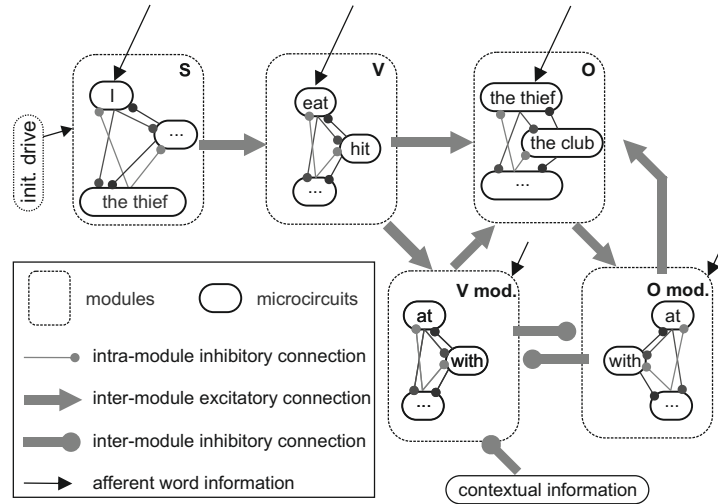


Figure 6.2: Initial syntax parsing model for sentence comprehension. The incremental parsing of syntactic information is represented by a sequential and selective activation of the word-representing minimal canonical microcircuits. Afferent word information selectively excites the minimal canonical microcircuit that represents the recognized phonological word form. The activated microcircuit, for example representing the word *I* in the subject module (S), pre-activates words in the connected verb-module (V). This allows the selective afferent word information to activate the respective microcircuit (e.g., *eat*) via feedforward conditioning. Subsequently, words in both the module of verb-modifiers (V mod.) and in the module of objects (O) are differentially pre-activated by weighted connections. Contextual information was proposed to guide this input-driven structure-building process by modulating the excitability of a targeted microcircuit, such as through inhibition.

6.2.2 Model Analysis

Two interpretations for the same afferent word information are observable. For the first interpretation (Fig. 6.3C) afferent word information activates the minimal canonical microcircuit representing *I* (dark blue) and subsequently *hit* (red). This pre-activates word-representing minimal canonical microcircuits in the ob-

ject module (green and magenta) and in the verb-modifier module (light blue). It then activates *the thief* (green) which pre-activates the object-modifying module (brown). Afferent word information then activates *with* in both the verb-modifying (light blue) and the object-modifying module (brown). This leads to a competitive interaction that eventually resolves the ambiguity. During the competition phase, contextual information (yellow) inhibits the verb-modifying module. Further afferent word information activates *the club* (magenta) and completes this structure-building process for which *with the club* is interpreted as adjective phrase.

For the second interpretation (Fig. 6.3D), no specific contextual information is present. The interpretation proceeds according to the listener's (subjective) pre-disposition. In the model, the individual ratio of mutual inhibition between the modifiers determines this favored interpretation. Consequently, during the competition phase the verb-modifying module (light blue) inhibits the object-modifying (brown) module. Hence, *with the club* is interpreted as an adverbial phrase.

In Section 4.3, I show how the local network balance governs the capability of canonical microcircuit models for the basic operations on which this syntax parsing network rests. Several neurological disorders have been associated with a disturbed network balance at the level of interacting neural populations, such as epilepsy, autism, and schizophrenia [77, 80, 83]. Furthermore, drugs, anesthetics, and other chemicals are known to alter number or efficacy of available neurotransmitter receptors. Valenzuela [128] reports on a perturbation of the network balance in favor of inhibitory influences following alcohol consumption. In the syntax parsing network, this scenario is replicated by increasing the inhibitory synaptic gain slightly from $H_i = 22$ mV to $H_i = 23$ mV (and to $H_i = 24$ mV). This leads to a defective sentence representation (see Figs. 6.3E and F for the two different interpretations). Although the minimal canonical microcircuits receive the respective afferent word information, they are not able to maintain a correct pattern of sustained activity (working memory) so that the syntax parsing process fails.

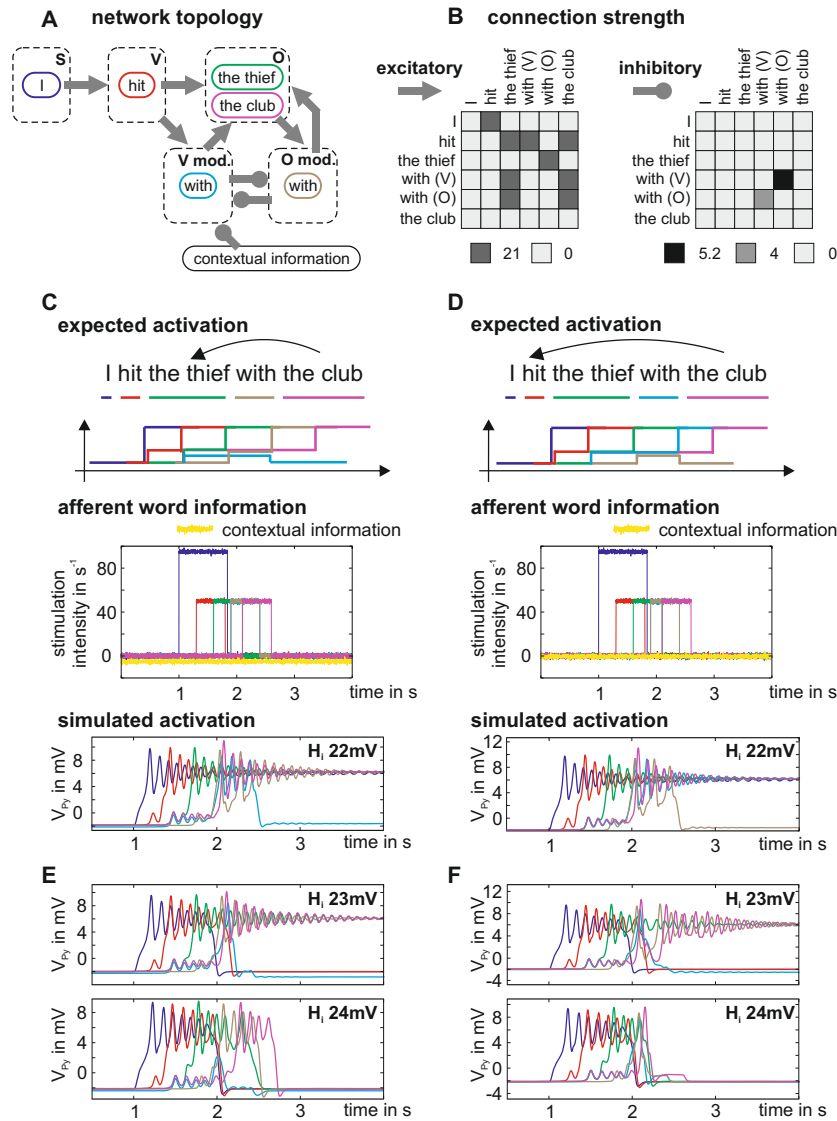


Figure 6.3: Neural representation of a perceived sentence by a distributed network of six interacting canonical microcircuit models. A) The network topology features modules (dashed rectangles) containing the six relevant word-representing minimal canonical microcircuits (solid colored rectangles), which are interconnected through excitatory and inhibitory connections of individual strength (B). The line colors consistently reflect the respective words in all panels. C) Contextual information that inhibits the module of verb-modifiers guides the interpretation. Consequently, the phrase with the club is interpreted as adjective phrase. D) Without contextual information the verb-modifying module becomes activated and with the club is interpreted as adverbial phrase. E)-F) In case of a disturbed local network balance an accurate structure-building fails for both interpretations (E and F) and leads to misinterpretations. In the model this means defective word activation traces (top plots) or even memory loss, that is no lasting activation trace at all (bottom plots).

6.2.3 Model Assessment

In this first syntax parsing model, single words have been represented by minimal canonical microcircuits that are grouped to syntactic modules following the principle of place coding [22]. This simple network model transforms a continuous stream of afferent word information into a sustained neural activity trace. This transformation, the actual syntactic parsing, reflects decoding and storing the syntactic relations. The selectivity in this input-driven sequential activation rests on a required superimposition of already established information (pre-activation through previously activated modules) and novel sensory information (feedforward conditioning). This structure-building process fails if the minimal canonical microcircuits lose the capability for their basic operations, for instance following a disturbed local network balance. The analysis of this first syntax parsing model reveals the following issues:

- **Multiple instantiations of single words:** Due to a single object-module, comprising all known nouns of an individual's vocabulary, this first syntax parsing model is not able to represent a repeated instantiation of the same word as in the sentence *I draw a wall on a wall*. A recognized word activates the respective microcircuit, which becomes insensitive to further stimulation – and is not available for further recruitment.
- **Simple syntax structure:** The modularized organization of the network constrains more complex syntax structures. The selection and flexibility in order of syntactic categories requires the repetition of entire word-grouping modules. However, this implies a redundancy of words and the consequent effort for their maintenance.
- **Self-activation by means of pre-activity:** The superposition of pre-activation and afferent word information (both mediated by feedforward connections) allows an aggregation of pre-activation that may lead to an erroneous self-activation of the canonical microcircuit.

In the following I advance this model using findings on state-dependent processing, that is to say hierarchic conditioning, and functional specialization of minimal canonical microcircuits in hierarchical networks.

6.3 Advanced Syntax Parsing Based on Hierarchical Conditioning

6.3.1 Model Architecture and Mechanism

The advanced syntax parsing model features a similar architecture and makes the same basic assumptions as the first syntax parsing model. In contrast to this initial parsing model, the advanced model repetitively implements the structure-building meta-circuit analyzed in Section 5.5. The advanced parsing model (Fig. 6.4) considers eleven interacting minimal canonical microcircuits that separately represent either syntactic roles, i.e. *syntax nodes*, or the phonological word forms of recognized words, i.e. *word nodes*. Syntax nodes are coupled through lateral connections that target the Py population of connected nodes, similar to feedback connections. Their connectivity pattern reflect syntax rules that have been established during language acquisition. Word nodes are unidirectionally coupled to the respective syntax nodes through feedforward connections. For instance, the noun *thief* connects to both the subject and object nodes (see Fig. 6.4B).

Afferent word information is modeled as rectangular stimuli of defined intensity p_{stim} and duration t_{dur} . Afferent word information activates the word nodes that in turn stimulate the respective syntax nodes via feedforward connections (gray arrows in Fig. 6.4B). Importantly though, only predicted syntax nodes are activatable, that is, they received a lateral facilitative signal before stimulation (dashed arrows), denoting hierarchical conditioning. Notably, stimulations of the word nodes and syntax predictions are assumed to be unspecific. For instance, the noun *thief* stimulates both the subject and the object syntax node. The initial expectation to perceive a sentence is modeled as an initial driving signal to the subject-reflecting syntax node (node 1 in Fig. 6.4B). Contextual information, here exemplarily inhibiting the object-modifier node, assists the semantic disambiguation. Important parameters of the advanced syntax parsing model are summarized in Table 6.1.

Table 6.1: Important parameters of the advanced syntax parsing model.

Forward connection strength	c_f	34
Intensity of acoustic input	p_{stim}	140 s^{-1}
Lateral connection strength	c_{lat}	5
Duration of acoustic input	t_{dur}	500 ms
Contextual information	p_{inh}	3.8 s^{-1}

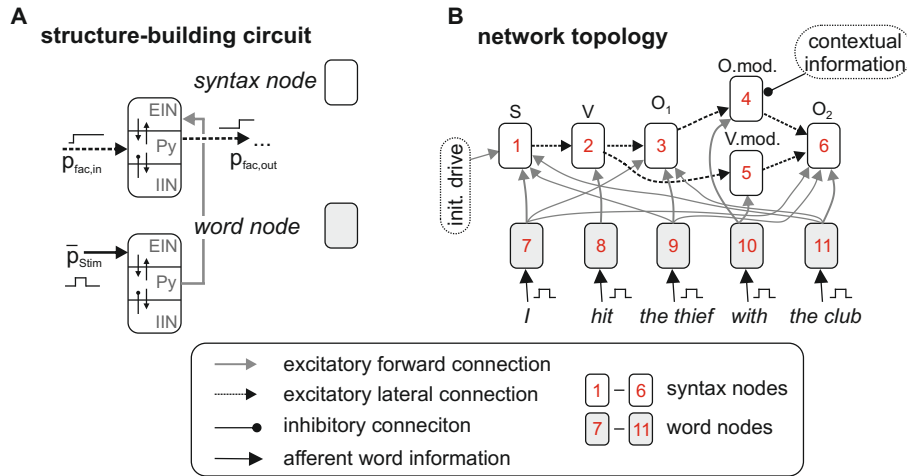


Figure 6.4: Architecture of the advanced syntax parsing model. A) The structure-building meta-circuit (Fig. 5.1C) is used as a building block for the syntax parsing model. The higher-level minimal canonical microcircuit is interpreted as syntax node, representing syntactic roles, and the lower-level as word node, representing the phonological word form. B) In the syntax parsing model, eleven interacting minimal canonical microcircuits represent either syntax nodes (1-6) or word nodes (7-11). Word nodes receive afferent word information and unidirectionally propagate their activity to syntax nodes via feedforward connections. Syntax nodes are interconnected by lateral connections and exchange facilitative signals (syntax predictions) that condition the establishment of sustained activity patterns. Contextual information, here inhibiting the object-modifier node (O.mod.), assists the semantic disambiguation.

6.3.2 Model Analysis

Again, model simulations considered the example sentence *I hit the thief with the club*. During the parsing process, word nodes respond to their consecutive stimulation (top plot in Fig. 6.5A) and selectively activate syntax nodes (bottom plots in Fig. 6.5A). This parsing process provides a sustained activity trace at the end of the parsing process and interprets *with the club* as adverbial phrase. This is expressed by the sustained high activity of the verb modifier in Fig. 6.5A. Different contextual information would lead to the alternative interpretation for which the phrase *with the club* operates as an adjective phrase. The separation of phonological word information and syntactic roles allows multiple instantiations of words independent of their syntactic roles. Hence, parsing the sentence *the thief hit the thief with the club* is possible (Figure 6.5B). This functional specialization to transient activity in word nodes and sustained activity in syntax nodes is a key feature of the structure-

building meta-circuit (Section 5.5). As emphasized in Figure 6.5C, the perception of single words (e.g., *thief*) is transient in the word nodes and excites all connected syntax nodes (e.g., S, O₁, O₂). However, the excitation persistently remains only for syntactically predicted nodes (S, black line) but vanishes for unpredicted nodes that did not receive a facilitative signal (O₁/O₂, red and purple line). These findings considerably increase the face validity of the syntax parsing model.

6.3.3 Model Assessment

In the advanced syntax parsing model, word-representing minimal canonical microcircuits selectively activate higher-order minimal canonical microcircuits that represent syntactic roles. This separation of phonological word information and syntactic roles as well as the consideration of hierarchical conditioning extend the first syntax parsing model in the following points. Firstly, the word nodes respond transiently to perceived afferent word information and independently from the expected syntactic role. This allows the word to be used multiple times in a sentence. Secondly, the sustained activity of syntax nodes, selectively excited by the word nodes, causes predictions among the syntax nodes that guides the parsing according to syntax rules. The separation of minimal redundant word nodes and syntax nodes allows the construction of more elaborate syntax structures up to the point of reflecting phrase structures and symbolic operations [129]; a point that needs to be further investigated in future studies. Thirdly, the accumulation of pre-activation (superimposed inputs between the syntax nodes) leading to an erroneous activation of a syntax node is less likely due to the lateral connections between the syntax nodes. Notably, both afferent word information and syntax predictions are unspecific. For instance, the word *drink* could be an object or a verb and, as an object, could be succeeded by many syntactic categories. Nevertheless, by integrating unspecific word information and unspecific syntactic predictions the syntax parsing network yields a specific syntax prototype (characteristic neural activity pattern).

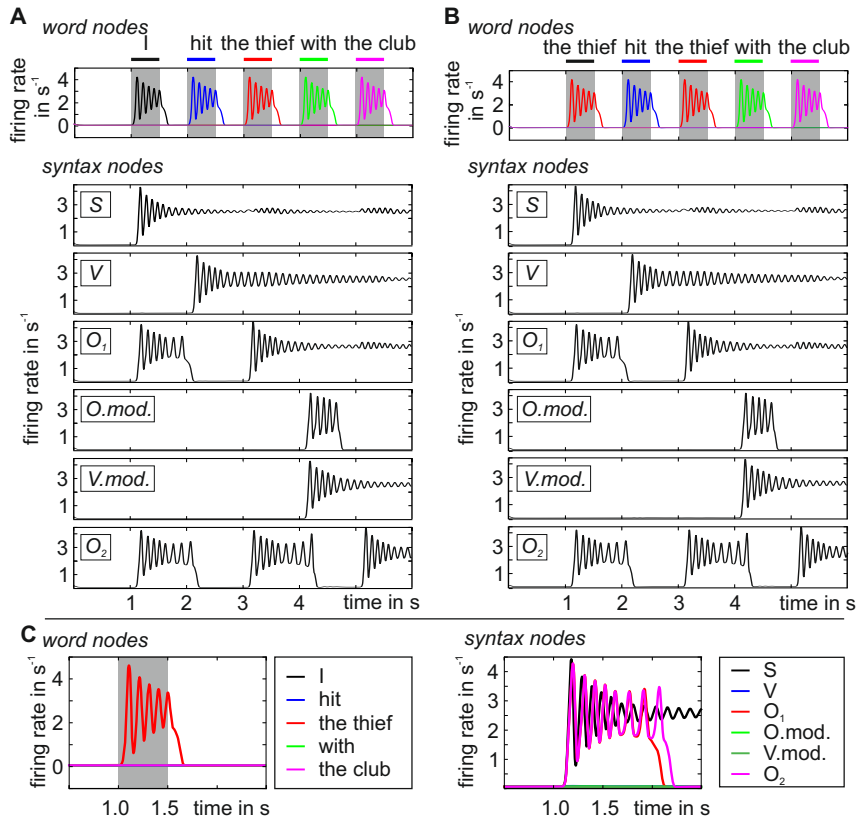


Figure 6.5: Analysis of the advanced syntax parsing model. A) Top plot: Word nodes respond to their consecutive stimulation (gray areas) and selectively activate syntax nodes (bottom plot). B) The selective activation of syntax nodes by word nodes allows the multi-instantiation of single words. The word thief occurs twice but reflects different syntactic categories. C) The perception of single words (e.g., thief) temporarily stimulates the respective word node that excites all connected conceptual nodes (e.g., S, O_1 , O_2). However, only syntactically predicted nodes (here S) remain active, whereas activity in unpredicted nodes (here O_1 , O_2) decays.

6.4 Discussion

In this section I prove the suitability of the framework of interacting minimal canonical microcircuits to investigate computational mechanisms in neurocognitive research. For the particular example of syntax parsing, I showed how syntactic and phonological word information may be neurobiologically represented by a network of minimal canonical microcircuits and how these pieces of information can be bound together via signal propagation within the network. The selective activation of the minimal canonical microcircuits and the defined retention of local infor-

mation rest upon basic operations and fail for a disturbed network balance. The explicit separation of phonological and syntactic word information allows for a flexible construction of syntax structures and solves the problem of multi-instantiation of distinct words. These results are now discussed in more detail.

The notion of an extensive network of similar microcircuits supporting cortical function has been put forward in specialized neurocognitive theories. For example, several proposals provide mechanistic insight into the processing of language [4, 20, 22, 52, 126]. I focused on potential mechanisms of syntax parsing that is proposed to rest on an input-driven sequential activation of coupled minimal canonical microcircuits. The main characteristics of the presented syntax parsing models are: (i) the consideration of afferent word information, (ii) the representation of words and syntactic roles at the neuronal level, and (iii) the directed signal propagation (structure-building) within the network.

I assume that perception and identification of the phonological word form informs but is conceptually distinct from the structure-building computations. When the incoming auditory information is recognized as a word the respective afferent word information (modeled via a rectangular signal of defined duration and intensity) is fed into the minimal canonical microcircuits that are recruited during syntax parsing. Word recognition models that deliver the afferent word information are conceivable to rest on stable-heteroclinic cycles [38], synfire chains [31] or sequence detectors [21]. The successful recognition of a word may temporally raise the firing rate of a neuronal unit. Future studies need to investigate in how far this raised firing rate is rectangular and provides the precise duration and intensity levels that are assumed in this present study. Moreover, future studies need to assess how these stimulation characteristics may be distorted by distributed conduction delays due to variations in axonal calibers [130].

The syntax parsing model exemplifies the notion that minimal canonical microcircuits constitute nodes of associative networks, or operational cell assemblies [4, 20]. In this operational cell assembly minimal canonical microcircuits represent word and syntax nodes. Interaction among the nodes includes basic operations and state-dependent operations (hierarchical or feedforward conditioning) that correspond to associations in cell assemblies. This interaction implements syntax parsing by decoding and storing hierarchical relations of the input using a place code [22]. The word nodes can be considered parts of word webs that encode all aspects of a content word, including semantic information. The selective

activation of word nodes leads to an ignition of the respective word webs and thus, fusion of syntactic and semantic information. It is likely that semantic information also informs syntax parsing. For example semantics may support disambiguation. This hardly understood interaction of semantics and syntax must be addressed in future studies [121].

Syntax parsing denotes the sequential excitation of word nodes and the decoding of their relation. The relation of words in a sentence is encoded through word order (i.e., syntax), morphological information, and qualifying prepositions [22]. I consider the word order as dominant factor (veritable for the English language) and disregard other factors. Syntactic roles (i.e., subject, verb, object etc.) of words are modeled implicitly as groups of word nodes (first syntax parsing model in Section 6.2) or explicitly as separate syntax nodes (advanced syntax parsing model in Section 6.3). Analysis shows that the latter option allows a functional specialization of minimal canonical microcircuits to either signal flow gating or working memory and enables multiple instantiations of single words with different syntactic roles. This separation and a proceeding hierarchical assignment of syntactic roles suggests the construction of detailed (or *hierarchized* [121]) syntax structures up to the point of reflecting phrase structures and symbolic operations [129]. In this context, the present computational framework has shown valuable for mechanistic neurocognitive research. It may serve as a complementary tool for neurolinguistics research approaches that face the limitations of lesion studies [21].

In terms of directed signal propagation (structure-building), both syntax parsing networks rely on a distinction between active (high pyramidal membrane potential, Section 4.2) and non-active (low pyramidal membrane potential) states of the minimal canonical microcircuits. Active minimal canonical microcircuits can (transiently or persistently) excite coupled non-active minimal canonical microcircuits via excitatory feedforward and feedback connections. Feedforward and hierarchical conditioning are means by which the network implements the structure-building. This input-driven functional binding of discrete elements enables the generation of infinite sequences out of a limited number of discrete elements. This principle is referred to as *infinite recursion* [23] and is conceivable for many cognitive operations. The mechanism's flexibility concerning sequence length as well as number and order of elements suits the application in syntax parsing. Sequence detectors and declining dynamics have been proposed as alternative mechanisms for

the serial processing of language [21]. Both mechanisms are realizable by minimal canonical microcircuits^{10,11} and may be subject of future investigation.

Adaptive structure-building in both syntax parsing models relies on the basic operations of minimal canonical microcircuits (analysis in Section 5.5). Clearly, an efficient short-term memory mechanism is needed to implement temporal integration of real-time sequences. The working memory mechanism of the minimal canonical microcircuit provides fast encoding and flexible holding times. The latter is especially relevant in order to account for varying speeds of speech and different sentence lengths. The signal flow gating ensures that words are activated only if they were either pre-activated (first syntax parsing model) or hierarchically predicted (advanced syntax parsing model) by the sentence structure and recognized from the input stream. As these basic operations have been shown to crucially depend on the network balance, altering this parameter leads to failure of the global network operation.

The presented simulations of the syntax parsing models used the three-population model as a minimal canonical microcircuit. In the first syntax parsing model, the coexistence of memory and transfer behavior, which is the hallmark of the three-population model, is not a necessary ingredient. Thus, a two-population model could have been used as well. However, the advantageous architecture of the advanced syntax parsing model allows multiple instantiations of word nodes due to transfer behavior and sustained activity in the syntax nodes due to memory behavior. This competence demands the coexistence of both response behaviors, given that the same canonical microcircuit model architecture is used for word and syntax nodes. Although the advanced syntax parsing model overcomes the limitations of the first syntax parsing model¹², it also has its limits. For instance, at the level of a discourse (multiple sentences related in their context), representations need to be deactivated before a following sentence can be represented. Such a deactivation mechanism is implicitly included in each minimal canonical microcircuit through a brief impulse to the IIN (see Fig. 4.1C). However, the corresponding word webs,

¹⁰ Sequence detector using hierarchical conditioning: circuit A sends feedback signal to circuit C, circuit B sends feedforward signal to circuit C. While sequence $A \rightarrow B$ activates C, it remains inactive for sequence $B \rightarrow A$. An appealing question concerns the supportable temporal (and spatial) distance between A and B.

¹¹ Declining dynamics may be implemented using reverberating activity between two recurrently coupled minimal canonical microcircuits.

¹² That is, multiple instantiations of single words, simple syntax structure, self-activation via of pre-activity.

which the word nodes are arguably a part of, need to decay slowly so that discourse context spanning multiple sentences can be linked through associative processes. Moreover, it has been shown that networks with too many activated nodes (long or complex sentences) tend to become unstable. This destroys the information that is stored in the network state [92]. These points need to be addressed in future studies.

The advanced syntax parsing model integrates unspecific afferent word information and unspecific syntactic predictions in order to yield a specific syntax prototype. The network's topology and connection strength are basis for syntactic predictions¹³ and reflect the rules of syntax, as established during language acquisition. Accordingly, topology and connection strength support only distinct numbers and types of syntax prototypes. Such prototypes, or *templates*, have been suggested by neurolinguistics research [121]. Due to the fixed wiring, the present model architectures represent a rule-based system. By introducing synaptic plasticity and thus self-organization, future work might investigate how this wiring is established during language acquisition [20] and evaluate the parsing model's role in second language acquisition.

“Language processing is a paradigm case of higher cognitive processes” [20] and language processing analysis is challenged by its many peculiarities¹⁴ [131]. The present thesis analyzes mechanisms of sentence processing in compliance with neurocognitive brain theories. This is opposed to the field of computational linguistics, which aims to optimize spoken human-machine interaction using sophisticated technical approaches while sacrificing biological plausibility. In contrast, neuroscientific language research, as presented in this present thesis, strives to identify the biologically plausible mechanisms of language processing and their implementation within the neuronal matter of the brain.

¹³ The predictability of a word in a sentential context has been associated with the posterior STG and syntactic predictions to the posterior IFG [121]. Such predictions are conceivable in the present modeling framework through facilitative feedback signals (hierarchical conditioning).

¹⁴ Including the versatility and flexibility of language in terms of extent (vocabulary, composite words), number and order of elements, syntactic (garden path sentences) and semantic ambiguity (homophones, context-dependency), variability in articulation (accent, dialect), environmental conditions (noise, acoustics, speech superposition), temporal characteristics (speed, elasticity of words, pauses), and semantic relations.

Chapter 7

Discussion, Future Research, and Conclusion

“Since all models are wrong the scientist cannot obtain a ‘correct’ one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity.”

– George E. P. Box, 1976

7.1 Discussion

In this thesis, I developed a biologically plausible generative modeling framework of neural canonical microcircuits that allows to explore and to simulate aspects of neural information processing related to neurocognitive research. This framework includes (i) the identification and formalization of neural canonical microcircuits as neurobiological fundamental processing units, (ii) the definition and evaluation of working memory and signal flow gating as their basic information processing operations, (iii) the description of priming and structure-building as meaningful neural operations that emerge from their interaction, and (iv) the proof of concept of how a processing network supports neurocognitive tasks, as demonstrated for syntax parsing.

Established methods for research on cognitive information processing include functional and behavioral measurements. However, as mentioned in Section 1.1 the interpretation of measured signals provides limited information about the actual mechanisms and operations that underlie cognition. Generative neural models reproduce these mechanisms and provide simulations that relate to measured signals. This thesis emphasizes the exploration of such mechanisms over the exact replication of measurements. Although results of exploratory approaches are not an unconditional truth [132] they improve the understanding of cortical information processing. This thesis shows that simple excitatory and inhibitory signal pathways exhibit basic information processing operations on a low hierarchical level that enable composite adaptive operations and cognitive functions on higher hierarchical levels. Even if many questions and necessary model developments remain on each scale, the multiscale approach of this thesis will be a prototype for future studies.

The present canonical microcircuit model employs neural masses to account for cognitive information processing. Neural mass models represent one category of mean field models and make comprehensive assumptions about the brain's processing, such as the dynamics of synaptic transmission, properties and averaged representation of neurons, topology and properties of connectivity, and the disregard of glia cells. Implementations of such models range from technical oscillators to neurobiologically plausible representations and make important contributions to neuroscientific research [133]. Neural mass models allow a description of the system's state space that extends time domain simulations. That way, fundamental characteristics, such as conditions for basic operations, are uniquely qualifiable. Moreover, such state space characteristics are commonly translatable to other models that may emphasize other aspects of cortical processing. In other words, neural mass models support the transfer of functional theories to neurobiological findings.

Alternatively, spiking neuron models make fewer assumptions about the brain's processing. They model the dynamics of single neurons, consider the transmission of action potentials rather than firing rates, and incorporate measurements on cortical connectivity [134]. The complexity of larger spiking neuron models requires considerable computational resources. Despite their dominance in biological plausibility, spiking neuron models lack the mathematical interpretability of neural mass models that provided the findings of this thesis. However, an experimental validation of these findings still faces conceptual and methodological questions. Spiking neuron models may thus be viewed as a tie between neural

mass modeling and biological reality. Hence, future studies on canonical microcircuits should strive to transfer the results of this thesis to spiking neuron networks to confirm their validity.

Connectionism assumes that higher cognitive functions emerge from the constructive interaction of a high number of fundamental units in distributed neural networks. Associative networks, or operational cell assemblies, conform to the notion of connectionism. I argue that neural population models of canonical microcircuits match the declared functionality of associations in cell assemblies [20] and thus, might constitute nodes of distributed associative networks.

I substantiated this argument by showing (i) how minimal canonical microcircuits support working memory and signal flow gating as basic information processing operations and (ii) how interacting minimal canonical microcircuits enable conditioned associations. These conditioned associations rest on superimposed feedforward information (feedforward conditioning) and integration of facilitative feedback and driving feedforward information (hierarchical conditioning). The local network balance regulates the capability for basic operations. A three-population model, accounting for separate neural populations for afferent and efferent information, was determined as the minimal architecture whose functional repertoire features the basic operations.

In simple meta-circuits, I demonstrated the ability of hierarchically interacting minimal canonical microcircuits to perform adaptive (state-dependent) neural processing operations, namely priming and structure-building. Both operations rest on the insight that feedback information stabilizes working memory behavior that was induced by feedforward stimulation and lowers the perception threshold of a minimal canonical microcircuit, thereby increasing its sensitivity. Connection strength and network balance constrain these neural operations.

Representation of syntax parsing at the neuronal level provides further evidence for the notion that minimal canonical microcircuits represent nodes of associative networks that perform cognitive functions. In the syntax parsing model unspecific phonological word information was related to syntactic information through directed and prediction-based signal propagation among minimal canonical microcircuits using a place code.

7.2 Future Research

Future research needs to address (i) methodological modalities of the model framework itself, (ii) the experimental validation of model predictions, and (iii) further insight into cognitive mechanisms. In particular, the following list includes the points:

Methodology of the modeling framework:

1. Consideration of interaction between excitatory and inhibitory interneurons as well as self-feedback among the excitatory interneurons (Section 4.6).
2. Consideration of simultaneous application of transient feedforward and feedback signals (Section 5.6).
3. Examination of consequences of deformed rectangular stimuli that may result from distributed conduction delays (Section 6.4).
4. Establishment of sequence detectors as an additional meta-circuit and evaluate feasibility of declining dynamics in canonical microcircuit models (Section 6.4).
5. Evaluation of risk and consequences of sudden changes in the global state in networks of canonical microcircuit models for intense interaction (Section 6.4).
6. Examination of modalities to allow a certain degree of self-organization in the model framework, such as through synaptic plasticity (Section 6.4).

Experimental validation of the following model predictions:

7. The retention of local information (store processing events and modify further processing) should be reflected by bistable activity levels that are consistently measured in circumscribed local areas for the same task. A shift in local network balance (drugs, electric stimulation) should disturb this pattern.
8. Analysis showed that constant levels of feedback information stabilize the increased neural activity (working memory behavior) that was induced by feedforward stimulation. Furthermore, feedback information lowers the perception threshold of a minimal canonical microcircuit, thereby increasing its sensitivity.

9. Analysis indicated that feedforward information is not strictly driving and feedback information is not strictly modulatory. Feedback modulation is more robust to variations of afferent stimuli than feedforward modulation.

Further insight into cognitive mechanisms:

10. Employing a hierarchical assignment of syntactic roles, future research should map the representation of hierarchized syntax structures (phrase structures) and symbolic operations from linguistic research to the present neurocomputational framework (Section 6.4).
11. Introduction of self-organization: investigate how syntactic relations are established during language acquisition.
12. Establish application of the identified neural operations and mechanisms in another cognitive system.
13. Apply the modeling framework to electrical brain stimulation and investigate the functional consequences regarding the shift of the local network balance and its relation to behavioral effects of the stimulation (Section 5.6).

7.3 Conclusion

The present results substantiate the concept of stereotypic functions in canonical microcircuits. The local network balance is shown to have a crucial influence to the information processing capacity of neural networks. Moreover, the results disclose the conceptual compatibility between minimal canonical microcircuits and operational cell assemblies. This leads to the idea that minimal canonical microcircuits may serve as biologically plausible nodes in models of hierarchically operating cell assemblies, effectively addressing the structural realization of cell assemblies in the neocortical matter. Thus, the present findings lend support to the connectionist idea that higher brain function arises from networks of relatively similar, though individually tunable, canonical microcircuits. Generative models of canonical microcircuits – that have already been used for EEG, MEG, and fMRI replication – may support and complement systematic experiment-based investigations of cognitive functions. Exemplifying this notion, I found mechanistic evidence that cognitive priming involves the dynamic shift of a perception threshold and that structure-building computations are likely subfunctions in a neural network for syntax parsing.

Appendix A

Appendix

A.1 Deriving a Dimensionless Form of the Temporal Differential Operator of an Alpha Function

In order to avoid rounding errors when handling small or large numbers in the Jacobian matrix and thus, to ensure the operability of DDE-BIFTOOL [100], the system governing differential equations was transformed into a dimensionless form.

According to Equations 3.9 and 3.10 the transformation of an afferent mean firing rate $\phi_{in}(t)$, arriving at an excitatory or inhibitory synapses, into a mean membrane potential $V_c(t)$, $c \in [P, E, I]$, is described by a second-order differential equation as in:

$$\frac{d^2 V_c(t)}{dt^2} + \frac{2}{\tau_{e,i}} \cdot \frac{dV_c(t)}{dt} + \frac{V_c(t)}{\tau_{e,i}^2} = \frac{H_{e,i}}{\tau_{e,i}} \cdot \phi_{in}(t) \quad (\text{A.1})$$

$V_c(t)$ denotes the mean membrane potential of a source population. Recalling Equation 3.13 the incoming firing rate $\phi_{in}(t)$ is usually expressed as a sigmoidal-shaped activation function of the membrane potential $V_c(t)$ as in:

$$\begin{aligned} \phi_{in}(t) &= S(V_c(t)) \\ &= \frac{2e_0}{1 + e^{r \cdot (v_0 - V_c(t))}} \end{aligned} \quad (\text{A.2})$$

With substitution of parameters:

$$\begin{aligned} t &= t_s \cdot \bar{t} \\ V_c &= v_s \cdot \gamma \end{aligned} \tag{A.3}$$

Equation A.1 becomes:

$$\frac{v_s}{t_s^2} \frac{d^2 \gamma}{d\bar{t}^2} + \frac{2v_s}{\tau_{e,i} \cdot t_s} \cdot \frac{d\gamma}{d\bar{t}} + \frac{v_s \cdot \gamma}{\tau_{e,i}^2} = \frac{H_{e,i}}{\tau_{e,i}} \cdot S(v_s \cdot \gamma(t_s \cdot \bar{t})) \tag{A.4}$$

When multiplying with t_s^2/v_s and setting:

$$\begin{aligned} t_s &= \tau_e \\ v_s &= r^{-1} \end{aligned} \tag{A.5}$$

For excitatory synapses one then obtains:

$$\frac{d^2 \gamma}{d\bar{t}^2} + 2 \cdot \frac{d\gamma}{d\bar{t}} + \gamma = H_e \tau_e r \cdot S(r^{-1} \cdot \gamma(\tau_e \cdot \bar{t})) \tag{A.6}$$

and for inhibitory synapses one obtains:

$$\frac{d^2 \gamma}{d\bar{t}^2} + \frac{2 \cdot \tau_e}{\tau_i} \cdot \frac{d\gamma}{d\bar{t}} + \frac{\tau_e^2 \cdot \gamma}{\tau_i^2} = \frac{H_i \tau_e^2 r}{\tau_i} \cdot S(r^{-1} \cdot \gamma(\tau_e \cdot \bar{t})) \tag{A.7}$$

By defining:

$$\eta = \frac{\tau_e}{\tau_i} \tag{A.8}$$

Equation A.7 becomes:

$$\frac{d^2 \gamma}{d\bar{t}^2} + 2 \cdot \eta \cdot \frac{d\gamma}{d\bar{t}} + \eta^2 \cdot \gamma = \eta^2 \cdot H_i \tau_i r \cdot S(r^{-1} \cdot \gamma(\tau_e \cdot \bar{t})) \tag{A.9}$$

Applying this substitution to Equation A.2 yields:

$$\begin{aligned}
S(V_c(t)) &= S(r^{-1} \cdot \gamma(\bar{t})) \\
&= \frac{2e_0}{1 + e^{r \cdot (v_0 - \gamma(\bar{t})/r)}} \\
&= \frac{2e_0}{1 + e^{r \cdot v_0} e^{-\gamma(\bar{t})}} \\
&= \frac{2e_0}{1 + \xi \cdot e^{-\gamma(\bar{t})}}
\end{aligned} \tag{A.10}$$

Expressing the second-order ODEs A.6 and A.9 by two coupled first-order ODE, for the excitatory populations one obtains:

$$\begin{aligned}
\dot{\gamma} &= \delta \\
\dot{\delta} &= H_e \tau_e r \cdot S(\gamma) - 2\delta - \gamma
\end{aligned} \tag{A.11}$$

and for inhibitory populations one obtains:

$$\begin{aligned}
\dot{\gamma} &= \delta \\
\dot{\delta} &= \eta^2 H_i \tau_i r \cdot S(\gamma) - 2\eta\delta - \eta^2\gamma
\end{aligned} \tag{A.12}$$

Accordingly, the generic equation system that was analyzed using DDE-BIFTOOL and evaluated through numerical integration reads:

$$\begin{aligned}
\dot{\gamma}_1 &= \delta_1 \\
\dot{\delta}_1 &= H_e \tau_e r \cdot (N_{EP} \cdot S(\gamma_2 - \gamma_3) + b_1 \cdot p_{ff}) - 2\delta_1 - \gamma_1 \\
\dot{\gamma}_2 &= \delta_2 \\
\dot{\delta}_2 &= H_e \tau_e r \cdot (b_1 \cdot N_{PE} \cdot S(\gamma_1) + (1 - b_1) \cdot N_{PP} \cdot S(\gamma_2 - \gamma_3) + (1 - b_1) \cdot p_{ff} + b_3 \cdot p_{fb}) - 2\delta_2 - \gamma_2 \\
\dot{\gamma}_3 &= \delta_3 \\
\dot{\delta}_3 &= \eta^2 H_i \tau_i r \cdot N_{PI} \cdot S(\gamma_4 - \gamma_5) - 2\eta \delta_3 - \eta^2 \gamma_3 \\
\dot{\gamma}_4 &= \delta_4 \\
\dot{\delta}_4 &= H_e \tau_e r \cdot N_{IP} \cdot S(\gamma_2 - \gamma_3) - 2\delta_4 - \gamma_4 \\
\dot{\gamma}_5 &= \delta_5 \\
\dot{\delta}_5 &= \eta^2 H_i \tau_i r \cdot N_{II} \cdot (1 - b_2) \cdot S(\gamma_4 - \gamma_5) - 2\eta \delta_5 - \eta^2 \gamma_5
\end{aligned} \tag{A.13}$$

A.2 State Space Modulation due to Variation of Synaptic Gains

In this thesis, the canonical microcircuit model's synaptic gains were varied to account for a variation in the local network balance. The following figures complement the information provided in the main text (see Section 4.3 and 4.4). In particular, Figures A.1 and A.2 illustrate bifurcation diagrams of the three-population model for a varying excitatory synaptic gain H_e (Fig. A.1) or a varying inhibitory synaptic gain H_i (Fig. A.2). Similarly, Figures A.3 and A.4 illustrate these variations in the two-population model without inhibition and Figures A.5 and A.6 for the two-population model with inhibition.

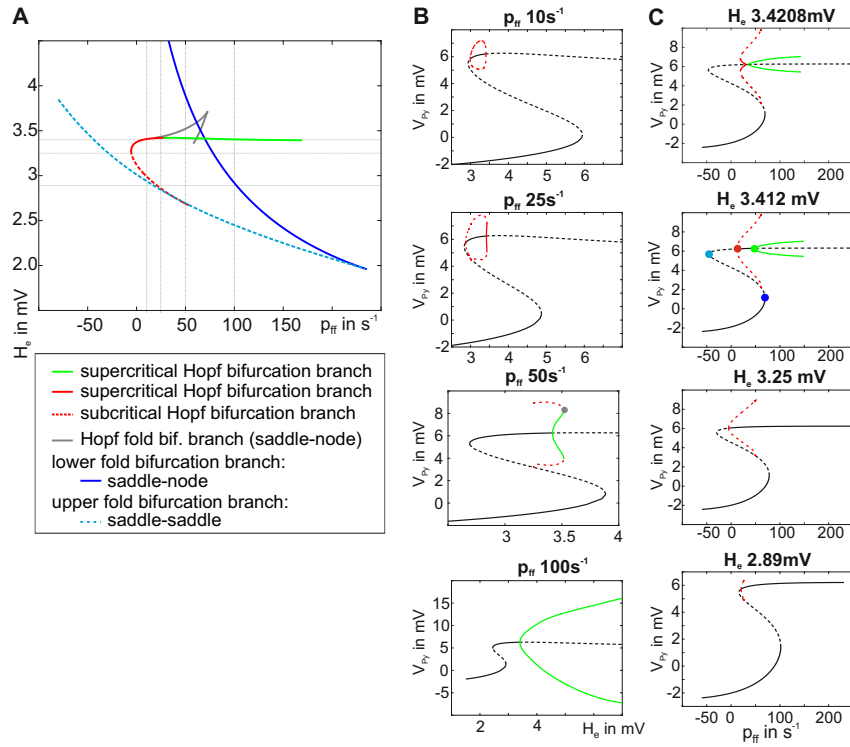


Figure A.1: Variation of the excitatory synaptic gain in the three-population model.

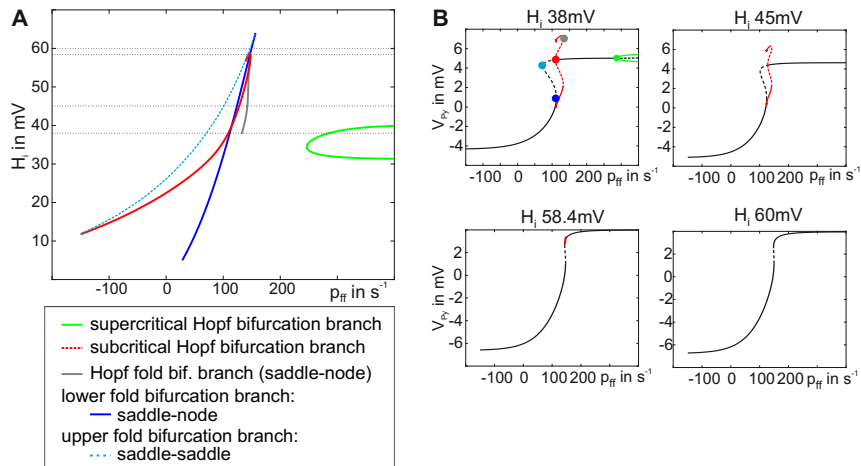


Figure A.2: Variation of the inhibitory synaptic gain in the three-population model.

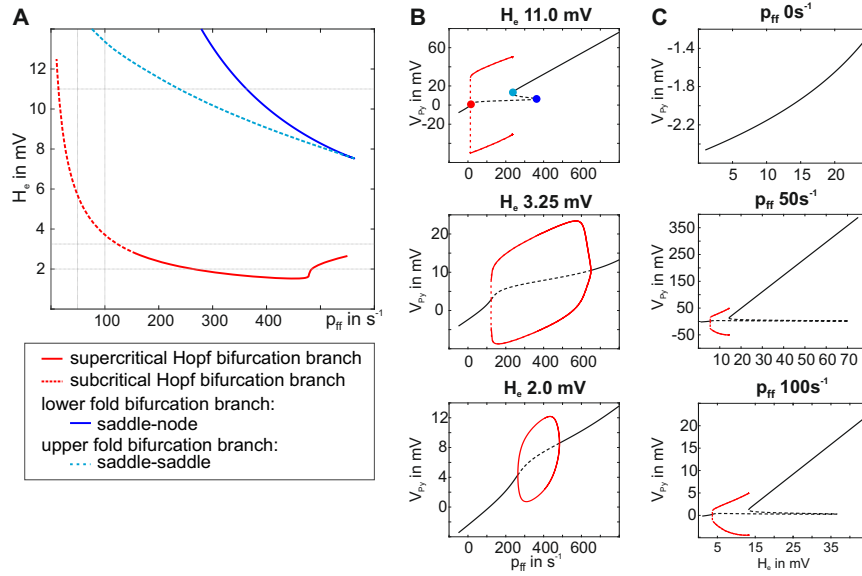


Figure A.3: Variation of the excitatory synaptic gain in the two-population model without disinhibition.

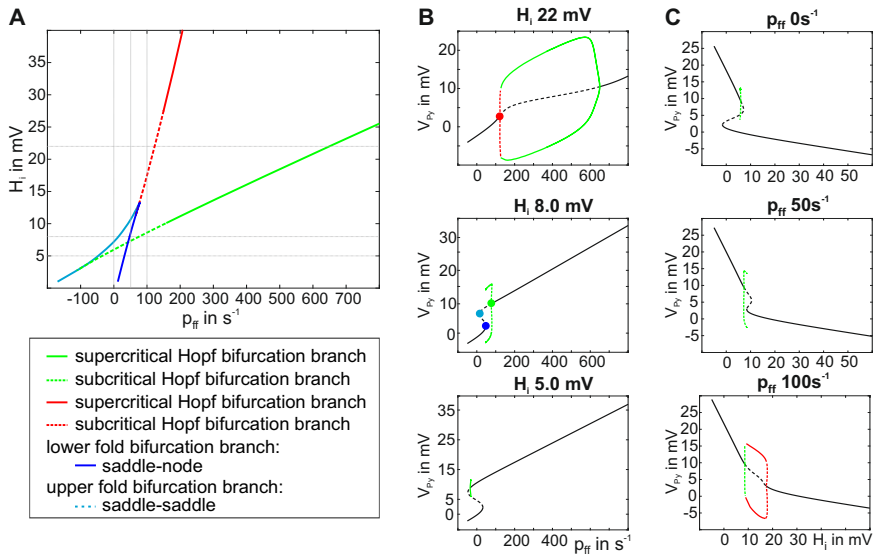


Figure A.4: Variation of the inhibitory synaptic gain in the two-population model without disinhibition.

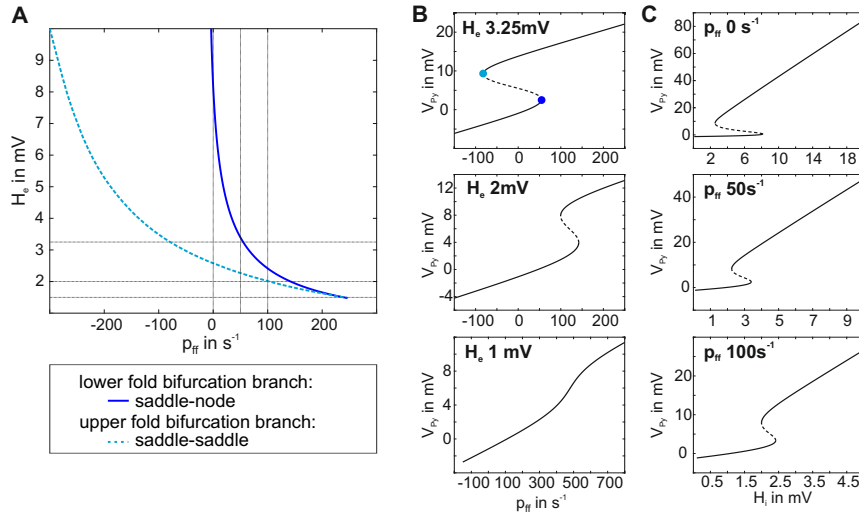


Figure A.5: Variation of the excitatory synaptic gain in the two-population model with disinhibition.

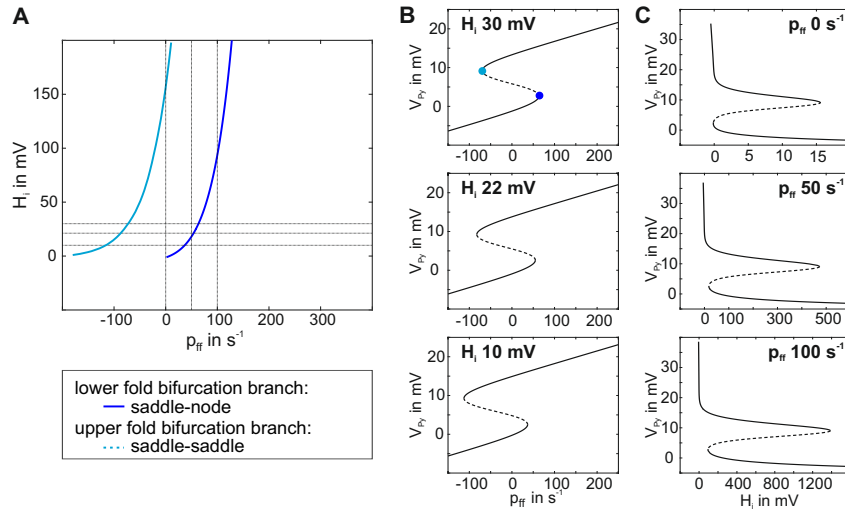


Figure A.6: Variation of the inhibitory synaptic gain in the two-population model with disinhibition.

Appendix B

Glossary

The following list provides brief explanations for the most important terms and concepts used in this thesis. Note that other texts may define them differently.

Basic operations: Denotes the minimal set of fundamental stereotypic functions a canonical microcircuit needs to feature to perform elemental information processing. In this thesis, this minimal set of basic operations comprises signal flow gating and working memory.

Canonical microcircuit: This term refers to theoretical, fundamental processing units in the neocortex that feature a common basic circuitry and stereotypic functions.

Canonical microcircuit model: Computational (formalized) models that represent a canonical microcircuit. Many architectures are conceivable of which three are investigated in this thesis.

Connectionism: Neurocognitive notion that higher cognitive functions emerge from the constructive interaction of a high number of fundamental units in distributed neural networks.

Hierarchical connections: Extrinsic excitatory connections arriving at a canonical microcircuit from lower levels (feedforward connection) and higher levels (feedback connection).

Feedforward conditioning: Mechanism for which cascaded canonical microcircuit models establish conditioned associations by superimposed feedforward information.

Firing threshold: The firing threshold conceptually separates the active and inactive state of a neuronal population (Fig. 3.4A). It was defined relative to the maximum firing rate of 5 s^{-1} , so that about 25% of the maximum firing rate is reached at the threshold of 4 mV.

Hierarchical conditioning: Mechanism for which hierarchically interacting canonical microcircuit models establish conditioned associations by means of simultaneously received feedback and feedforward information.

Intrinsic connections: Excitatory and inhibitory connections among neuronal populations within a canonical microcircuit.

Minimal canonical microcircuit: Canonical microcircuit model which considers interacting excitatory and inhibitory neuronal populations, whose structural topology reflects the homogeneity of neocortical matter, and whose functional repertoire features the basic operations of signal flow gating and working memory.

Neurocognitive brain theory: Denotes concepts explaining the nature of neural interaction that leads to cognitive functions.

Perception threshold: Intensity border that separates nonresponsive and transfer/memory behavior. This threshold reflects the distance between working point and lower fold bifurcation (e.g., Fig. 4.1B).

Phrase structure building: Establishment of a phrase structure (linguistic concept).

Recurrent connections: Cyclic connections (excitatory and inhibitory) within one level of a hierarchy.

Saturation threshold of the sigmoid function: Mean membrane potential of about 14.8 mV for which the sigmoidal function is greater than 99% of the maximum output activity of 5 s^{-1} .

State-dependent processing: Mechanism for which the activity of one canonical microcircuit model governs the stimulation-induced response of another canonical microcircuit model.

Structure-building: Gradual propagation of neuronal activity that establishes complex patterns in time and space.

Syntax parsing: Process of syntactic structure-building as in decoding and temporal storage of syntactic information based on a continuous stream of afferent phonological word information.

Bibliography

- [1] G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput Biol*, 4(8):e1000092, 2008.
- [2] J. Haueisen and T.R. Knösche. Forward modeling and tissue conductivities. In S. Supek and C. J. Aine, editors, *Magnetoencephalography*, pages 107–127. Springer-Verlag, Berlin Heidelberg, 2014.
- [3] T. H. Sander, T. R. Knösche, A. Schlogl, F. Kohl, and colleagues. Recent advances in modeling and analysis of bioelectric and biomagnetic sources. *Biomed Tech (Berl)*, 55(2):65–76, 2010.
- [4] F. Pulvermüller, M. Garagnani, and T. Wennekers. Thinking in circuits: toward neurobiological explanation in cognitive neuroscience. *Biol Cybern*, 108(5):573–593, 2014.
- [5] K. J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, 2003.
- [6] R. L. Lewis. *Cognitive modeling, symbolic*. MIT Press, Cambridge, 1999.
- [7] J. Brunner. Kursorischer Streifzug durch die Geschichte der Neurowissenschaften aus neuroethischer und neurophilosophischer Perspektive. In D. Groß and S. Müller, editors, *Sind die Gedanken frei?: Die Neurowissenschaften in Geschichte und Gegenwart*, pages 2–22. Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, 2007.
- [8] G. O’Brien and J. Opie. How do connectionist networks compute? *Cogn Process*, 7(1):30–41, 2006.

- [9] R. J. Douglas and K. A. Martin. A functional microcircuit for cat visual cortex. *J Physiol*, 440:735–769, 1991.
- [10] R. J. Douglas and K. A. Martin. Mapping the matrix: the ways of neocortex. *Neuron*, 56:226–238, 2007.
- [11] R. J. Douglas, K. A. Martin, and D. Whitteridge. A canonical microcircuit for neocortex. *Neural Comput*, 1:480–488, 1989.
- [12] S. F. Beul and C. C. Hilgetag. Towards a canonicalägranular cortical microcircuit. *Front Neuroanat*, 8(165):1–8, 2014.
- [13] C. C. Hilgetag, M. A. O’Neill, and M. P. Young. Hierarchical organization of macaque and cat cortical sensory systems explored with a novel network processor. *Philos Trans R Soc Lond B Biol Sci*, 355(1393):71–89, 2000.
- [14] P. Haueis. Meeting the brain on its own terms. *Front Hum Neurosci*, 8:815, 2014.
- [15] B. H. Jansen and V. G. Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol Cybern*, 73(4):357–366, 1995.
- [16] F. H. Lopes da Silva, A. Hoeks, H. Smits, and L. H. Zetterberg. Model of brain rhythmic activity. the alpha-rhythm of the thalamus. *Kybernetik*, 15(1):27–37, 1974.
- [17] L. H. Zetterberg, L. Kristiansson, and K. Mossberg. Performance of a model for a local neuron population. *Biol Cybern*, 31(1):15–26, 1978.
- [18] T Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, 1989.
- [19] A. Kristjansson. I know what you did on the last trial—a selective review of research on priming in visual search. *Front Biosci*, 13:1171–1181, 2008.
- [20] T. Wennekers, M. Garagnani, and F. Pulvermüller. Language models based on hebbian cell assemblies. *J Physiol Paris*, 100(1-3):16–30, 2006.
- [21] F. Pulvermüller. A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. *Prog Neurobiol*, 67(2):85–111, 2002.

- [22] E. T. Rolls and G. Deco. Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. *Brain Res*, 1621:316–334, 2015.
- [23] A. Treves. Frontal latching networks: a possible neural basis for infinite recursion. *Cogn Neuropsychol*, 22(3):276–291, 2005.
- [24] P. Dayan and L.F. Abbott. *Theoretical Neuroscience. Computational and Mathematical Modeling of Neural Systems*. The MIT Press, Cambridge, 2001.
- [25] D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1:1–47, 1991.
- [26] J. Lübke and D. Feldmeyer. Excitatory signal flow and connectivity in a cortical column: focus on barrel cortex. *Brain Struct Funct*, 212(1):3–17, 2007.
- [27] V. B. Mountcastle. Modality and topographic properties of single neurons of cat’s somatic sensory cortex. *J Neurophysiol*, 20(4):408–434, 1957.
- [28] O. Sporns, G. Tononi, and R. Kotter. The human connectome: A structural description of the human brain. *PLoS Comput Biol*, 1(4):e42, 2005.
- [29] A. M. Thomson, D. C. West, Y. Wang, and A. P. Bannister. Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2-5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling in vitro. *Cereb Cortex*, 12(9):936–953, 2002.
- [30] Z. Wang, A. T. Sornborger, and L. Tao. Graded, dynamically routable information processing with synfire-gated synfire chains. *PLoS Comput Biol*, 12(6):e1004979, 2016.
- [31] F. Pulvermüller and Y. Shtyrov. Spatiotemporal signatures of large-scale synfire chains for speech processing as revealed by MEG. *Cereb Cortex*, 19(1):79–88, 2009.
- [32] F. Pulvermüller, Y. Shtyrov, and O. Hauk. Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain Lang*, 110(2):81–94, 2009.

- [33] M. Abeles. *Corticonics, Neural Circuits of the Cerebral Cortex*. Cambridge University Press, Cambridge, 1991.
- [34] P. Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci*, 9(10):474–480, 2005.
- [35] K. Friston. Hierarchical models in the brain. *PLoS Comput Biol*, 4(11):e1000211, 2008.
- [36] M. I. Rabinovich, I. Tristan, and P. Varona. Hierarchical nonlinear dynamics of human attention. *Neurosci Biobehav Rev*, 55:18–35, 2015.
- [37] M. I. Rabinovich, P. Varona, A. I. Selverston, and H. D. I. Abarbanel. Dynamical principles in neuroscience. *Rev. Mod. Phys.*, 78(4):1213–1265, 2006.
- [38] I. B. Yildiz, K. von Kriegstein, and S. J. Kiebel. From birdsong to human speech recognition: bayesian inference on a hierarchy of nonlinear dynamical systems. *PLoS Comput Biol*, 9(9):e1003219, 2013.
- [39] V. Braitenberg. Cell assemblies in the cerebral cortex. In S. Levin, editor, *Lecture Notes in Biomathematics*, pages 171–188. Springer-Verlag, Berlin Heidelberg New York, 1978.
- [40] D.O. Hebb. *The organization of behavior: A neuropsychological theory*. Wiley, New York, 1949.
- [41] G. Palm, A. Knoblauch, F. Hauser, and A. Schüz. Cell assemblies in the cerebral cortex. *Biol Cybern*, 108(5):559–572, 2014.
- [42] T. Wennekers, F. Sommer, and A. Aertsen. Cell assemblies. *Theory Biosci*, 122:1–4, 2003.
- [43] D. H. Hubel and T. N. Wiesel. Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *J Comp Neurol*, 158(3):295–305, 1974.
- [44] V. B. Mountcastle. The columnar organization of the neocortex. *Brain*, 120(Pt 4):701–722, 1997.
- [45] T. Binzegger, R. J. Douglas, and K. A. Martin. A quantitative map of the circuit of cat primary visual cortex. *J Neurosci*, 24(39):8441–8453, 2004.

- [46] A. Braitenberg, V.; Schüz. *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer-Verlag, Berlin Heidelberg, 1998.
- [47] K. D. Harris and G. M. Shepherd. The neocortical circuit: themes and variations. *Nat Neurosci*, 18(2):170–181, 2015.
- [48] A. J. Rockel, R. W. Hiorns, and T. P. Powell. The basic uniformity in structure of the neocortex. *Brain*, 103(2):221–244, 1980.
- [49] W.J. Freeman. *Mass action in the nervous system*. Academic Press, New York, 1975.
- [50] R. J. Douglas and K. A. Martin. Neuronal circuits of the neocortex. *Annu Rev Neurosci*, 27:419–451, 2004.
- [51] G. Silberberg, A. Gupta, and H. Markram. Stereotypy in neocortical microcircuits. *Trends Neurosci*, 25(5):227–230, 2002.
- [52] A. D. Friederici and W. Singer. Grounding language processing on basic neurophysiological principles. *Trends Cogn Sci*, 19(6):329–338, 2015.
- [53] K. D. Miller. Canonical computations of cerebral cortex. *Curr Opin Neurobiol*, 37:75–84, 2016.
- [54] A. M. Thomson and A. P. Bannister. Interlaminar connections in the neocortex. *Cereb Cortex*, 13(1):5–14, 2003.
- [55] A. M. Thomson and C. Lamy. Functional maps of neocortical local circuitry. *Front Neurosci*, 1(1):19–42, 2007.
- [56] S. Haeusler, K. Schuch, and W. Maass. Motif distribution, dynamical properties, and computational performance of two data-based cortical microcircuit templates. *J Physiol Paris*, 103(1-2):73–87, 2009.
- [57] W. Maass, P. Joshi, and E. D. Sontag. Computational aspects of feedback in neural circuits. *PLoS Comput Biol*, 3(1):e165, 2007.
- [58] A. M. Bastos, V. Litvak, R. Moran, C. A. Bosman, and colleagues. A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. *Neuroimage*, 108:460–475, 2015.

- [59] T. Kunze, A. D. H. Peterson, J. Haueisen, and T. R. Knösche. A model of individualized canonical microcircuits supporting cognitive operations. *PLoS One*, 12(12):e0188003, 2017.
- [60] D. Mumford. On the computational architecture of the neocortex. ii. the role of cortico-cortical loops. *Biol Cybern*, 66(3):241–251, 1992.
- [61] C. A. Bosman and F. Aboitiz. Functional constraints in the evolution of brain circuits. *Front Neurosci*, 9(303):1–13, 2015.
- [62] M. Carandini and D. J. Heeger. Normalization as a canonical neural computation. *Nat Rev Neurosci*, 13:51–62, 2012.
- [63] R. J. Douglas and K. A. Martin. Recurrent neuronal circuits in the neocortex. *Curr Biol*, 17(13):R496–500, 2007.
- [64] S. Haesler and W. Maass. A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. *Cereb Cortex*, 17(1):149–162, 2007.
- [65] T. R. Vidyasagar. A neuronal model of attentional spotlight: parietal guiding the temporal. *Brain Res Brain Res Rev*, 30(1):66–76, 1999.
- [66] T. Isa and Y. Kobayashi. Switching between cortical and subcortical sensorimotor pathways. *Prog Brain Res*, 143:299–305, 2004.
- [67] S. Johnson, J. Marro, and J. J. Torres. Robust short-term memory without synaptic learning. *PLoS One*, 8(1):e50276, 2013.
- [68] D. T. Liley, P. J. Cadusch, and M. P. Dafilis. A spatially continuous mean field theory of electrocortical activity. *Network*, 13(1):67–113, 2002.
- [69] P. A. Robinson, C. J. Rennie, and D. L. Rowe. Dynamics of large-scale brain activity in normal arousal states and epileptic seizures. *Phys Rev E Stat Nonlin Soft Matter Phys*, 65(4 Pt 1):041924, 2002.
- [70] A. Spiegler, S. J. Kiebel, F. M. Atay, and T. R. Knösche. Bifurcation analysis of neural mass models: Impact of extrinsic inputs and dendritic time constants. *Neuroimage*, 52(3):1041–1058, 2010.
- [71] O. David, S. J. Kiebel, L. M. Harrison, J. Mattout, and colleagues. Dynamic causal modeling of evoked responses in eeg and meg. *Neuroimage*, 30:1255–1272, 2006.

- [72] A. Garnier, A. Vidal, C. Huneau, and H. Benali. A neural mass model with direct and indirect excitatory feedback loops: identification of bifurcations and temporal dynamics. *Neural Comput*, 27(2):329–64, 2015.
- [73] F. Wendling, F. Bartolomei, J. J. Bellanger, and P. Chauvel. Epileptic fast activity can be explained by a model of impaired gabaergic dendritic inhibition. *Eur J Neurosci*, 15(9):1499–1508, 2002.
- [74] D. Malagarriga, A. E. Villa, J. Garcia-Ojalvo, and A. J. Pons. Mesoscopic segregation of excitation and inhibition in a brain network model. *PLoS Comput Biol*, 11(2):e1004007, 2015.
- [75] C. van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.
- [76] Y. Shu, A. Hasenstaub, and D. A. McCormick. Turning on and off recurrent balanced cortical activity. *Nature*, 423(6937):288–293, 2003.
- [77] N. Dehghani, A. Peyrache, B. Telenczuk, M. Le Van Quyen, and colleagues. Dynamic balance of excitation and inhibition in human and monkey neocortex. *Sci Rep*, 6:23176, 2016.
- [78] 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. *J Neurosci*, 26(17):4535–4545, 2006.
- [79] J. Ziburkus, J. R. Cressman, and S. J. Schiff. Seizures as imbalanced up states: excitatory and inhibitory conductances during seizure-like events. *J Neurophysiol*, 109(5):1296–1306, 2013.
- [80] T. Bourgeron. A synaptic trek to autism. *Curr Opin Neurobiol*, 19:231–234, 2009.
- [81] N. Gogolla, J. J. Leblanc, K. B. Quast, T. C. Sudhof, and colleagues. Common circuit defect of excitatory-inhibitory balance in mouse models of autism. *J Neurodev Disord*, 1(2):172–181, 2009.
- [82] S. Vattikuti and C. C. Chow. A computational model for cerebral cortical dysfunction in autism spectrum disorders. *Biol Psychiatry*, 67(7):672–678, 2010.

- [83] T. Sigurdsson. Neural circuit dysfunction in schizophrenia: Insights from animal models. *Neuroscience*, 321:42–65, 2016.
- [84] O. Yizhar, L. E. Fenno, M. Prigge, F. Schneider, and colleagues. Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature*, 477(7363):171–178, 2011.
- [85] M. S. Rowan, S. A. Neymotin, and W. W. Lytton. Electrostimulation to reduce synaptic scaling driven progression of alzheimer’s disease. *Front Comput Neurosci*, 8:39, 2014.
- [86] A. M. Bastos, W. M. Usrey, R. A. Adams, G. R. Mangun, and colleagues. Canonical microcircuits for predictive coding. *Neuron*, 76:695–711, 2012.
- [87] S. Ardid, X. J. Wang, and A. Compte. An integrated microcircuit model of attentional processing in the neocortex. *J Neurosci*, 27(32):8486–8495, 2007.
- [88] J. Heinzle, K. Hepp, and K. A. Martin. A microcircuit model of the frontal eye fields. *J Neurosci*, 27(35):9341–9353, 2007.
- [89] A. D. Friederici. Towards a neural basis of auditory sentence processing. *Trends Cogn Sci*, 6(2):78–84, 2002.
- [90] H. S. Meyer, V. C. Wimmer, M. Oberlaender, C. P. de Kock, and colleagues. Number and laminar distribution of neurons in a thalamocortical projection column of rat vibrissal cortex. *Cereb Cortex*, 20(10):2277–2286, 2010.
- [91] M. Goodfellow, K. Schindler, and G. Baier. Self-organised transients in a neural mass model of epileptogenic tissue dynamics. *Neuroimage*, 59(3):2644–2660, 2012.
- [92] T. Kunze, A. Hunold, J. Haueisen, V. Jirsa, and A. Spiegler. Transcranial direct current stimulation changes resting state functional connectivity: A large-scale brain network modeling study. *Neuroimage*, 140:174–187, 2016.
- [93] I. Merlet, G. Birot, R. Salvador, B. Molaee-Ardekani, and colleagues. From oscillatory transcranial current stimulation to scalp eeg changes: a biophysical and physiological modeling study. *PLoS One*, 8(2):e57330, 2013.
- [94] F. Grimbert and O. Faugeras. Bifurcation analysis of jansen’s neural mass model. *Neural Comput*, 18(12):3052–3068, 2006.

- [95] J. Touboul, F. Wendling, P. Chauvel, and O. Faugeras. Neural mass activity, bifurcations, and epilepsy. *Neural Comput*, 23(12):3232–3286, 2011.
- [96] N. Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J Comput Neurosci*, 8:183–208, 2000.
- [97] H. Meffin, A. N. Burkitt, and D. B. Grayden. An analytical model for the large, fluctuating synaptic conductance state typical of neocortical neurons in vivo. *J Comput Neurosci*, 16(2):159–175, 2004.
- [98] S.H. Strogatz. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering*. Westview Press, Boulder, CO, 2014.
- [99] E. M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, Cambridge, 2010.
- [100] K. Engelborghs. Numerical bifurcation analysis of delay differential equations using DDE-BIFTOOL. *ACM Trans Math Softw*, 28(1):1–21, 2002.
- [101] J. M. Beggs. The criticality hypothesis: how local cortical networks might optimize information processing. *Philos Trans A Math Phys Eng Sci*, 366:329–343, 2008.
- [102] J. Hesse and T. Gross. Self-organized criticality as a fundamental property of neural systems. *Front Syst Neurosci*, 8:166, 2014.
- [103] T. C. Potjans and M. Diesmann. The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. *Cereb Cortex*, 24(3):785–806, 2014.
- [104] N. Cain, R. Iyer, C. Koch, and S. Mihalas. The computational properties of a simplified cortical column model. *PLoS Comput Biol*, 12(9):e1005045, 2016.
- [105] D. Durstewitz, J. K. Seamans, and T. J. Sejnowski. Neurocomputational models of working memory. *Nat Neurosci*, 3:1184–1191, 2000.
- [106] S. Royer, M. Martina, and D. Pare. Bistable behavior of inhibitory neurons controlling impulse traffic through the amygdala: role of a slowly deinactivating K⁺ current. *J Neurosci*, 20(24):9034–9039, 2000.

-
- [107] C. J. Stam and E. C. van Straaten. The organization of physiological brain networks. *Clin Neurophysiol*, 123(6):1067–1087, 2012.
- [108] G. Liu. Local structural balance and functional interaction of excitatory and inhibitory synapses in hippocampal dendrites. *Nat Neurosci*, 7(4):373–379, 2004.
- [109] E. Marder and J. M. Goaillard. Variability, compensation and homeostasis in neuron and network function. *Nat Rev Neurosci*, 7(7):563–574, 2006.
- [110] D. L. Schacter and R. L. Buckner. Priming and the brain. *Neuron*, 20(2):185–195, 1998.
- [111] E. Tulving and D. L. Schacter. Priming and human memory systems. *Science*, 247(4940):301–306, 1990.
- [112] R. P. Rao and D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*, 2(1):79–87, 1999.
- [113] K. Friston. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*, 360(1456):815–836, 2005.
- [114] S. Shipp. Neural elements for predictive coding. *Front Psychol*, 7:1792, 2016.
- [115] L. Spillmann, B. Dresch-Langley, and C. H. Tseng. Beyond the classical receptive field: The effect of contextual stimuli. *J Vis*, 15(9):7, 2015.
- [116] E. N. Covic and S. M. Sherman. Synaptic properties of connections between the primary and secondary auditory cortices in mice. *Cereb Cortex*, 21:2425–2441, 2011.
- [117] M. Y. Cheng, M. Aswendt, and G. K. Steinberg. Optogenetic approaches to target specific neural circuits in post-stroke recovery. *Neurotherapeutics*, 13:325–340, 2016.
- [118] A. Fertonani and C. Miniussi. Transcranial electrical stimulation: What we know and do not know about mechanisms. *Neuroscientist*, 23(2):109–123, 2016.
- [119] J. M. Fuster. Network memory. *Trends Neurosci*, 20(10):451–459, 1997.

-
- [120] A. D. Friederici. The brain basis of language processing: from structure to function. *Physiol Rev*, 91:1357–1392, 2011.
- [121] A. D. Friederici. The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn Sci*, 16(5):262–268, 2012.
- [122] A. D. Friederici and S. M. Gierhan. The language network. *Curr Opin Neurobiol*, 23:250–254, 2013.
- [123] J. Brennan, Y. Nir, U. Hasson, R. Malach, and colleagues. Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain Lang*, 120:163–173, 2012.
- [124] S. Hickok, G.; Small. *Neurobiology of Language*. Academic Press, London, San Diego, Waltham, Oxford, 2016.
- [125] F. Pulvermüller. Words in the brain’s language. *Behav Brain Sci*, 22(2):253–279, 1999.
- [126] H. Markert, A. Knoblauch, and G. Palm. Modelling of syntactical processing in the cortex. *Biosystems*, 89(1-3):300–315, 2007.
- [127] D. Sammler, S. A. Kotz, K. Eckstein, D. V. Ott, and A. D. Friederici. Prosody meets syntax: the role of the corpus callosum. *Brain*, 133(9):2643–2655, 2010.
- [128] C. F. Valenzuela. Alcohol and neurotransmitter interactions. *Alcohol Health Res World*, 21(2):144–148, 1997.
- [129] N. Chomsky. *The Minimalist Program*. The MIT Press, Cambridge, 1995.
- [130] J. A. Perge, J. E. Niven, E. Mugnaini, V. Balasubramanian, and P. Sterling. Why do axons differ in caliber? *J Neurosci*, 32(2):626–638, 2012.
- [131] J. Hirschberg and C. D. Manning. Advances in natural language processing. *Science*, 349(6245):261–266, 2015.
- [132] G. Lohmann, K. Erfurth, K. Müller, and R. Turner. Critical comments on dynamic causal modelling. *Neuroimage*, 59:2322–2329, 2012.
- [133] V.K. Jirsa, W.C. Stacey, P.P. Quilichini, A.I. Ivanov, and C. Bernard. On the nature of seizure dynamics. *Brain*, 137(Pt 8):2210–2230, 2014.
- [134] H. Markram. The blue brain project. *Nat Rev Neurosci*, 7(2):153–160, 2006.

List of Figures

2.1	Canonical microcircuit models	13
3.1	Local feedback topologies of the canonical microcircuit model . .	21
3.2	Regrouping two excitatory populations to a single population . . .	22
3.3	Generalized architecture of the canonical microcircuit model . . .	28
3.4	Stimulation principle and categorization of the dynamic response behavior	31
4.1	Response behaviors of the three-population model for feedforward stimulation	35
4.2	Dynamics of the distinct response behaviors in a projection of the state space	37
4.3	Phase-dependency between the system's intrinsic oscillation and stimulus switch off time	38
4.4	Dynamic function map for the three-population model with indi- rect excitatory local feedback architecture	40
4.5	Two-parameter bifurcation plot of the three-population model . . .	42
4.6	Gradual transformation of the indirect excitatory local feedback to a direct excitatory local feedback	44
4.7	Dynamic function map for the direct excitatory local feedback ar- chitecture	45
4.8	Bifurcation plot of the two-population model with direct excitatory local feedback	46
4.9	Increasing recurrent inhibitory self-feedback in the two-population model	48
4.10	Dynamic function map for the two-population model with disinhi- bition	50

4.11 Two-parameter bifurcation plot for the two-population model with disinhibition	51
4.12 Characteristic fingerprints for transient stimulation of Py	52
5.1 Simultaneous stimulation of the canonical microcircuit model with feedforward and feedback input	60
5.1 Prototypical meta-circuits of cooperating canonical microcircuit models	63
5.2 Principle of the priming mechanism	64
5.3 Analysis of the priming mechanism	66
5.4 Impact of stimulus salience, connectivity gains, and network balance to the priming mechanism	69
5.5 Characterization of stimuli that shift the perception threshold	70
5.6 Influence of the priming meta-circuits' topology	71
5.7 Facilitative feedback signals enable the memorization of stimulation events	72
6.1 Schematic of the cortical language circuit	79
6.2 Initial syntax parsing model for sentence comprehension	83
6.3 Neural representation of a perceived sentence by a distributed network of six interacting canonical microcircuit models	85
6.4 Architecture of the advanced syntax parsing model	88
6.5 Analysis of the advanced syntax parsing model	90
A.1 Variation of the excitatory synaptic gain in the three-population model.	104
A.2 Variation of the inhibitory synaptic gain in the three-population model.	104
A.3 Variation of the excitatory synaptic gain in the two-population model without disinhibition.	105
A.4 Variation of the inhibitory synaptic gain in the two-population model without disinhibition.	105
A.5 Variation of the excitatory synaptic gain in the two-population model with disinhibition.	106
A.6 Variation of the inhibitory synaptic gain in the two-population model with disinhibition.	106

List of Tables

3.1	Default parameterization of the canonical microcircuit model. . .	27
5.1	Parameter values for the assessment of the priming mechanism. . .	67
6.1	Important parameters of the advanced syntax parsing model. . . .	87